# Perception-based QoE Criteria for XR Systems

Jukka Häkkinen
Department of Psychology
University of Helsinki

Evan Center
Center for Ubiquitous Computing
Faculty of Information Technology and Electrical Engineering
University of Oulu

19.1.2021

**1 Introduction** **3**

**2 Display requirements** **5**
2.1 Display types 5
2.2 Field-of-view 6
2.2.1 Definition 6
2.2.2 Factors affecting the uFoV 9
2.2.2.1 Position in eye motion box 9
2.2.2.2 Optical distortion 10
2.2.2.3 Eye rotation 10
2.2.2.4 Binocular and monocular FoV 11
2.2.2.5 Increasing the FoV by reducing binocular overlap 12
2.1.3 FoV size recommendations 15
2.1.3.1 Immersive experiences use case 15
2.1.3.2 Opaque monocular display 17
2.1.3.3 See-through display in stationary use 19
2.1.3.4 See-through display in mobile use use 19
2.3 Resolution 19
2.3.1 Modulation transfer function 20
2.4 Luminance and contrast 20
2.4.2. Low light conditions 21
2.4.2.1 Reduced dark adaptation 21
2.4.2.2 Dazzling glare 22
2.4.3 Interocular luminance differences 22
2.5 Framerate 22
2.6 Color breakup 23

**3 Virtual display** **24**
3.1 Vergence plane 24
3.2 Interocular distance 25
3.2.1 Divergence 26
3.2.2 Other changes in vergence demand 27
3.2.3 Reduction of uFOV 27
3.2.4 Distorted depth perception 27
3.2.5 IPD requirements 27
3.2.6 Changing the IOD 28
3.3 Focal plane 28
3.4 Properties of stereoscopic contents 29

**4 Optical system** **31**
4.1 Distortions 31
4.1.1 Pincushion distortion 31
4.1.2 Chromatic aberration 31
4.3 Eye relief 32
4.4 Eye motion box 32
4.5 Misalignments 33

4.5.1 Vertical 33
4.5.2 Rotational differences 33
4.5.3 Magnification differences 33

**5 Latency 34**

**6 Weight and center of gravity 34**

**7 Tips for further research 35**
7.1 Testing 35
7.2 Experimental Design 35
7.2.2 Outcome measures 37
7.2.3 Sampling 37
7.2.4 Grouping 38
7.3 Interpreting results and more notes on experiment planning 39
7.3.1 Null hypothesis testing vs. estimation 39
7.3.2 Statistical power 40

**8 References 42**

# 1 Introduction

**Quality of experience (QoE).** Simply put, how good is the experience in various dimensions? This subjective outcome from the perspective of the user is our primary concern when considering XR device perceptual requirements. Brunnström et al. (2013) define quality of experience as “the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user’s personality and current state.” This definition captures the multidimensional nature of quality of experience as an interaction between technology, user, and context. Is the display quality, image quality, video quality, audio quality and interaction experience satisfactory compared to other devices that the user sees in the environment?

These factors are also subject to time and the quality criteria change constantly as technology develops. For example, a television that had superb quality 30 years ago would be completely unacceptable for users today. In other words, the quality criteria are defined by the technological environment of the user.

Theoretically there are maximum (or optimal) parameter values for factors making up quality of experience after which the quality experience saturates (or deteriorates), but as we are not yet there in any dimension, we need to define the current parameter space with subjective testing. This also means that the quality experience cannot be directly predicted from technical properties; instead, there likely exists an optimal set of parameter values for each use case and user.

While the majority of this guide will focus on making recommendations for different aspects of device hardware and software, this last point highlights the need to take into account the use case and user, and as such we will make note of such considerations as they arise throughout the document. Every context has its own unique set of demands and it is up to XR developers to discover the set of solutions that best meet the demands for that context and consumer base. We might find ourselves naturally inclined to aspire towards the most cutting edge, high performance solution, and sometimes this is indeed the optimal solution, but remember that a path favoring intuitiveness and simplicity is often better for a given problem.

Likewise, remember that humans are characterized by a vast array of individual differences. XR research has only begun to scratch the surface in terms of understanding how individual differences interact with the various components discussed throughout this document, but they are nonetheless critical. Take for example that women on average have smaller head circumferences than men, and similarly, smaller inter-pupillary distances. Combine this with the fact that women are largely underrepresented in XR research, both as researchers and participants (Peck et al, 2020), and the result is that we could be failing to adequately capture the quality experience of about half of the world’s population. Paying attention to use cases and individual differences will help to ensure the highest possible quality of experience for all users.

Perceptual requirements can be categorized into several categories according to the experience:

- **Positive experiences.** This category includes immersive experiences such as spatial and social presence as well as increases in positive emotions. In a broad sense, Botella et al (2012) describe positive interactions with technology as those that are hedonic (bringing pleasure), eudaimonic (bringing self-growth), and/or social (bringing a sense of connectedness). XR devices have immense potential to bring positive experiences in each of these categories, whether that comes in the form of playing a game, learning a new skill, or connecting with a distant loved one. Positive experiences are of course rewarding to users and will promote continued interaction with the technology.

- **Adverse symptoms.** Unfortunately there is also a well documented downside to XR experiences which entails their capacity to induce sickness, described in the literature as cybersickness, simulator sickness, or virtual reality sickness. We will use cybersickness as a catch-all term for XR-related sickness throughout the rest of this document. Typically cybersickness is clustered into three symptom categories: nausea, oculomotor and disorientation symptoms (Kennedy et al, 1993). Oculomotor symptoms can also be called eye strain or asthenopia. These adverse symptoms, in addition to being inherently undesirable, also break immersion and discourage continued use of the technology.

- **Experiences related to visual processing.** XR devices introduce artificial signals to our natural visual systems that can interact in a variety of ways and produce a variety of perceptual experiences, some harmonious and others discordant. When this interaction is harmonious users are able to effectively make use of the introduced signals. In the case of VR, users are effectively tricked into accepting the simulation as reality, often described as having a sense of “presence” in the virtual environment. This outcome is in contrast to when the interaction is discordant, where the results can include experiences such as blurred vision or double vision, unstable percepts, or as can be the case in VR, the “cardboard effect” where simulation of object depth fails.

- **Experiences related to performance.** Regardless of the specific XR application, there is always some type of goal to be achieved whether that is finding a product in a warehouse or something as simple as having a fluid conversation with another person. The design choices for XR hardware and software can greatly affect performance en route to achieving that goal. Good design choices can improve performance, in best case scenarios facilitating experiences of flow or “effortless attention” where users are given the resources needed that complement their own skills in a way that allows for optimal completion of the task (Csikszentmihalyi, 1990). Bad design choices conversely can impede performance, in worst case scenarios leading to experiences of high cognitive load, exhaustion, and/or frustration. In addition to the primary experience of performance in the moment, there are also secondary so called “metacognitive” experiences that make up the user’s self-evaluation of their own ease of completing a task.

**Measuring experience.** Each of the types of experiences described above can be evaluated through objective and subjective measures. Objective measures include behavioral measurements, like detection performance in a visual task or recall performance in memory task, and physiological measurements, such as changes in heart rate or gastrointestinal activity as detected by sensors. The word “objective” here is not meant to convey that such measurements are infallible, as they still require proper interpretation within a context; rather, it comes from the idea that this type of measurement is considered more consistent and less biased by individual opinion than subjective measurements. Subjective measures include ratings on questionnaires, like mean opinion scores (MoS) or the simulator sickness questionnaire (SSQ) where participants rate their experiences on scales, or open ended responses, where participants describe their experiences which are subsequently coded by experimenters for evaluation. Subjective measures are considered more prone to biases, but may offer a window into experience that is impossible to capture via objective measures in many scenarios. Both types of measures are valuable tools in an experimenter’s toolbox and both are included in the data that has informed this document.

**Limitations.** Much work remains in terms of defining the XR parameter space. The field of XR research is relatively young and much of the research completed to date falls short of the sample sizes and methodological rigor required to arrive at truly satisfactory solutions to XR problems. There is no such thing as a perfect study or perfect way to measure experience quality, and as such we are constantly striving to develop better studies and measurements. Further complicating things, the sheer number of variables that can differ in regards to the hardware and software specifications of different XR setups used in different studies makes for a challenging problem in terms of precisely estimating the individual contribution of any one variable. While this is true for most of the parameter space, some variables are better understood than others, and in cases where the effects of a variable are particularly murky or contested, we make note of this limitation. In this guide we offer our best recommendations based on the current state of the XR literature.

# 2 Display requirements

## 2.1 Display types

XR displays can be categorized according to their properties.

Number of image sources and the way they are conveyed to eyes

1. **Binocular device** that has two image sources which are separately shown to the left and right eye. This can be used to show stereoscopic 3D images.
2. **Biocular device** that has one image source that is shown to both eyes.
3. **Monocular device** that has one image source shown to one eye.

Opaqueness

1. **Opaque XR** blocks the view to the outside world and the user views only the display. Often the term VR device implicates an opaque device. Sometimes the term non-see-through device is used.
2. **Optical see-through device** that has a transparent display so the user sees the display overlapping the outside environment.
3. **Video see-through device** that combines the video image of the outside environment and the display.

HMD reference point

1. **Device referenced XR device:** display contents move according to the head movements.
2. **Ground referenced XR device:** contents remain stationary when the user moves head.

## 2.2 Field-of-view

### 2.2.1 Definition

Field-of-view (FoV) is the angular extent subtended by the virtual image of a XR system (Greivenkamp 2004). The total FoV cannot usually be used, as edge areas can be distorted by optics or occluded by device parts. The area that an user can use is called the useful FoV (uFoV; Fig.1; Fig.2). The uFoV may vary according to the user, as the distorted or occluded area may depend on the position of the eye in front of the device optics. An important unresolved issue is the exact definition as well as subjective and objective measurement of the uFoV, where the field could benefit from the implementation of universal standards.

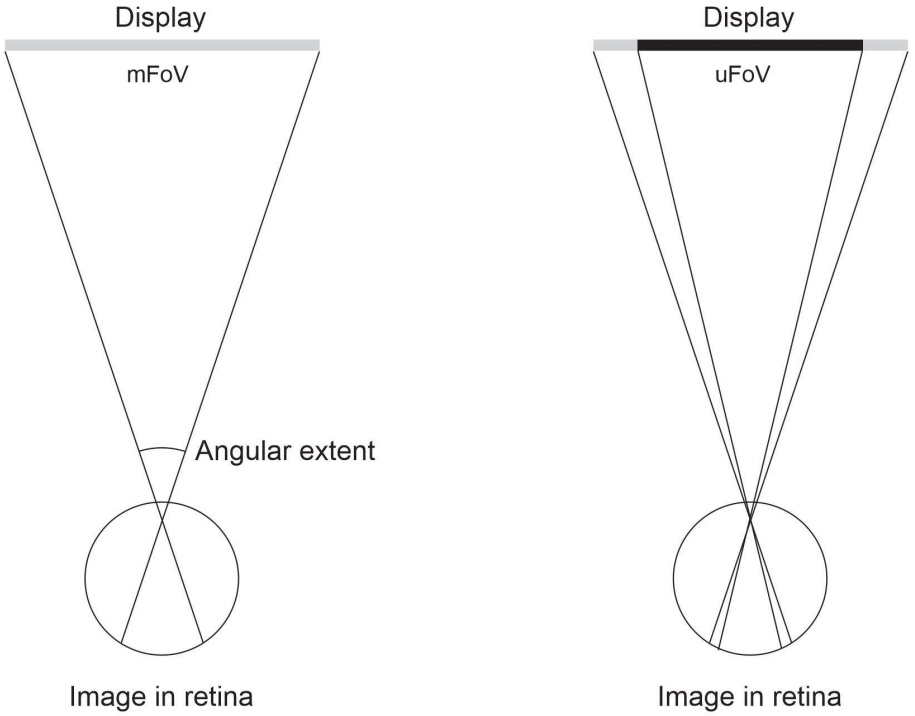



*Figure 1. On the left the thick grey line indicates the display and maximum possible FoV (mFoV). On the right the useful FoV (uFoV) is indicated by black color.*

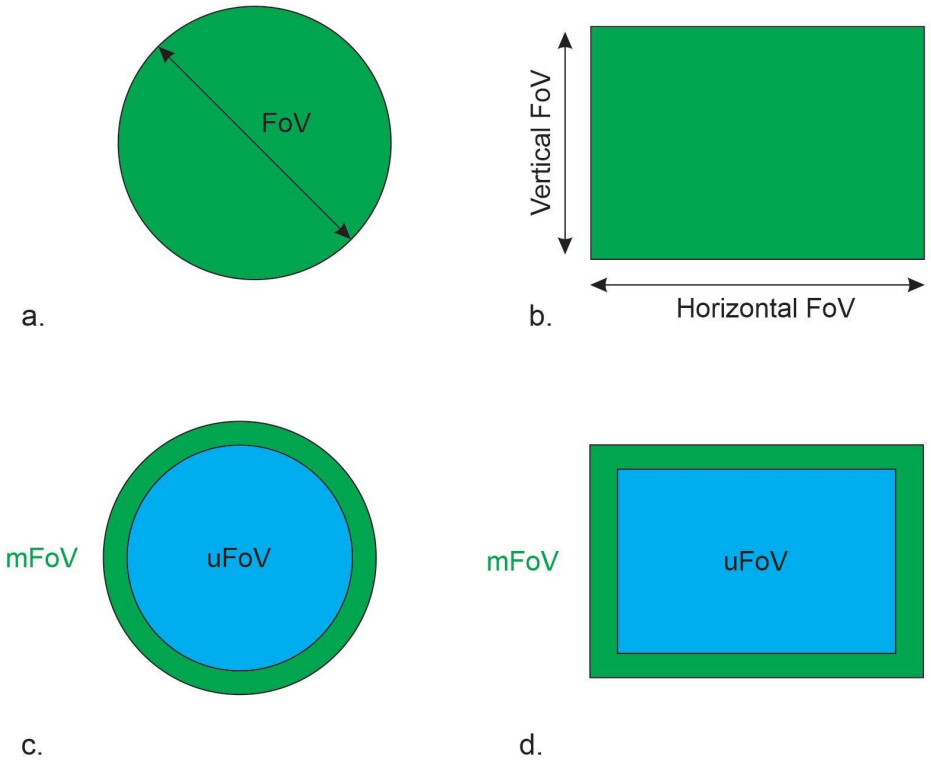



*Figure 2. FoV as seen by the user. a. Circular FoV. b. Rectangular FoV. c. Circular maximum FoV (mFoV) and useful FoV (uFov). d. Rectangular mFoV and uFoV.*

If the display does not cover the visual field completely, there are additional variables that need to be considered. Firstly, in an optical see-through display the mFoV may consist of a transparent area which is larger than the area of the active display. To differentiate these two transparent FoV (tFoV) is added (Fig 3.).

Secondly, in a display that leaves part of the environment visible, the device occludes part of the visual field. In figure 3. This is indicated by the black area. The occluding area can surround the display completely or occlude only some edges.

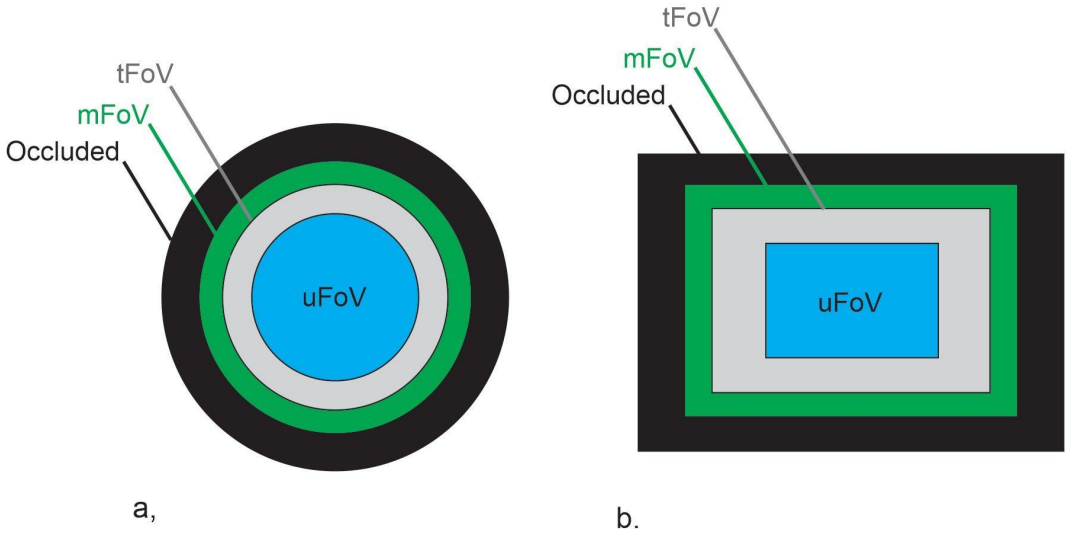



*Figure 3. a. Circular useful FoV (uFoV), transparent FoV (tFoV), maximum FoV (mFoV). The black area indicates environment occluded by the device. b. Rectangular versions of uFoV, tFoV, mFoV and occluded area.*

### 2.2.2 Factors affecting the uFoV

#### 2.2.2.1 Position in eye motion box

The eye motion box is the volume in which the eye can move in front of the display and get a satisfactory perception of the display contents. Also terms exit pupil and qualified viewing space (QVS) are used (Järvenpää & Pölönen, 2009; Järvenpää et al, 2010; Järvenpää & Pölönen 2012; Järvenpää & Salmimaa, 2016).

If the eye is not completely in the eye motion box, part of the display is not visible. These situations arise when, for example:

- The eye is horizontally and/or vertically away from the optimum viewing point of the display
- The eye is further away from the optimum distance from the exit pupil. This can happen when there is a non-optimal fit of the XR device
- There is a mismatch between interpupillary distance (IPD) of the user and the interocular distance (IOD) in a system in which the IOD cannot be changed
- There is a mismatch between interpupillary distance (IPD) of the user and the interocular distance (IOD), because the user has not been able to set a correct IOD in a system where the IOD can be changed

These problems can occur if the user is not able to position the device correctly because

- The device does not fit the head of the user well
- The user does not wear the device correctly
- The device slips during use

With non-optimal position in the eye motion box, the perceptual experience can be

- Partial occlusion of the display
- Dimming, blurring, form, or color distortion of display areas near the edges

#### 2.2.2.2 Optical distortion

Even with optimal position inside the eye motion box, the properties of the XR system optics might create differences in image fidelity in central and peripheral visual fields.

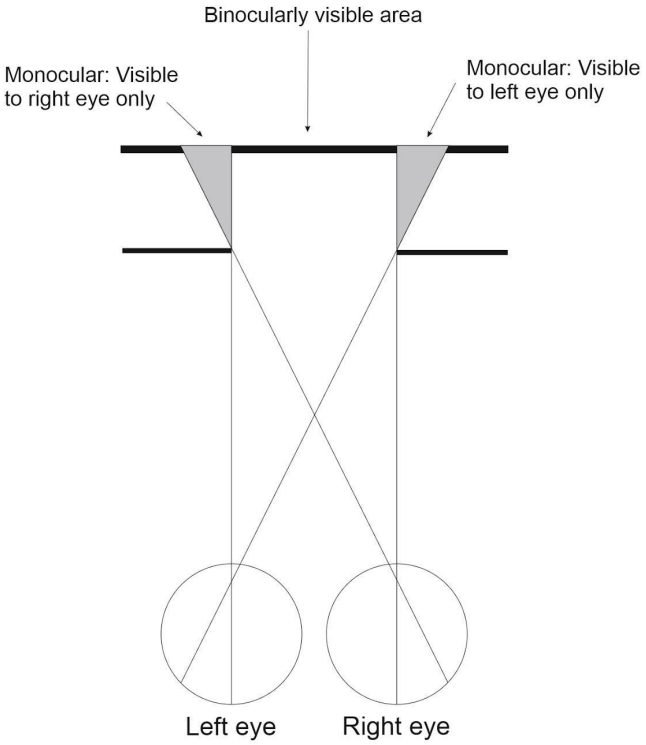
#### 2.2.2.3 Eye rotation

If the person looks straight ahead, the visible area is larger. If the person looks at location away from the center of the display, images near the edges may disappear. This effect happens because the rotation center of the eye is 10 millimeters behind the entrance pupil of the eye.

#### 2.2.2.4 Binocular and monocular FoV

In a binocular immersive headset the fused view contains binocular and monocular zones as shown in Fig.4a. The monocular areas are integrated to binocular areas, as they have a valid occlusion interpretation in the real world(Fig 4b; Shimojo & Nakayama, 1990). Thus, the binocular FoV is always larger than the monocular FoV.


a.




b.

*Figure 4.*

### 2.2.2.5 Increasing the FoV by reducing binocular overlap

When small FoV has been a problem in augmented reality displays, reducing the binocular overlap has been seen as a possible solution (Klymenko et al, 1994; Melzer, 2008 ). In figure 4a only the left and right sides of the left and right displays are binocularly fused. When parts of the displays remain unfused, the total fused FoV is larger.

In practise this might be difficult, as the binocular fusion system is very flexible and always tries to maximize the correlation of the left and right view. Thus, the partial overlap cannot be achieved just by moving the left and right display horizontally, as the binocular system will try to fuse them (Kooi, 1993). The result is either complete binocular fusion of the displays (Figure 4b) or inability to fuse the displays, which means that both displays are seen. The latter condition has multiple downsides. Firstly, not fusing the displays means that stereoscopic 3D cannot be presented. Secondly, both displays induce binocular rivalry, which reduces the visibility of information and annoys and strains the user. Thirdly, when two unfused displays are visible, the displays constantly move vertically and horizontally, as the latent horizontal and vertical phoria becomes visible with unfused stimuli.

The partial overlap can be achieved by placing fusion locks to the displays (Figure 4c). The four black dots serve as a stimulus for binocular fusion (Klymenko, 1994). When the binocular system fuses the dots, partial overlap can be achieved.

However, with partial overlap there are other problems. The displays are always surrounded by the opaque device parts (Figure 5a). When the displays are partially fused the monocular areas overlap with the opaque areas.

The location of the fusion lock determines the resulting fused display configuration. If the fusion locks are on the nasal part of the left and right visual fields (Fig. 5a), the monocular area that is on the left side of the central binocular area is visible to the left eye and the monocular area on the right side of the central binocular area is visible to the right eye. This is called the convergent configuration.

In Fig.5b is the reverse configuration. When the fusion lock is in the temporal side of the visual field, the monocular area on the left is visible to the right eye and the monocular area on the right is visible to the left eye.

In both conditions the overlap of dark opaque areas with bright displays creates binocular rivalry. In the literature this is called luning (Melzer, 1998; Klymenko et al 2000), but in the modern context the term binocular rivalry should be sufficient.

Binocular rivalry causes problems:

1. The user perceives uncontrollable changes between different views of the left and right eye (Patterson et al, 2007). In the non-overlapping FoV case rivalry is probably milder, as the bright display area dominates the rivalry. Furthermore, as the dark overlapping area is homogenous, it is perceived only as occasional dimming of the display area. However, binocular rivalry may still reduce task performance as the user may miss information when the display area is momentarily suppressed.
2. The need to pay extra attention to the rivalrous areas may increase cognitive load.
3. Binocular rivalry is uncontrollable and thus irritating.

To reduce the binocular rivalry, there are design recommendations for partially overlapping displays:

1. The size of the overlapping area should be at least 40 degrees (Patterson et al, 2014)
2. A convergent overlap design should be used. In this design the monocular region has a valid real-life interpretation: it can be perceived as being occluded by the display edge (Klymenko, 1994). This has been shown to reduce binocular rivalry (Shimojo & Nakayama, 1990).

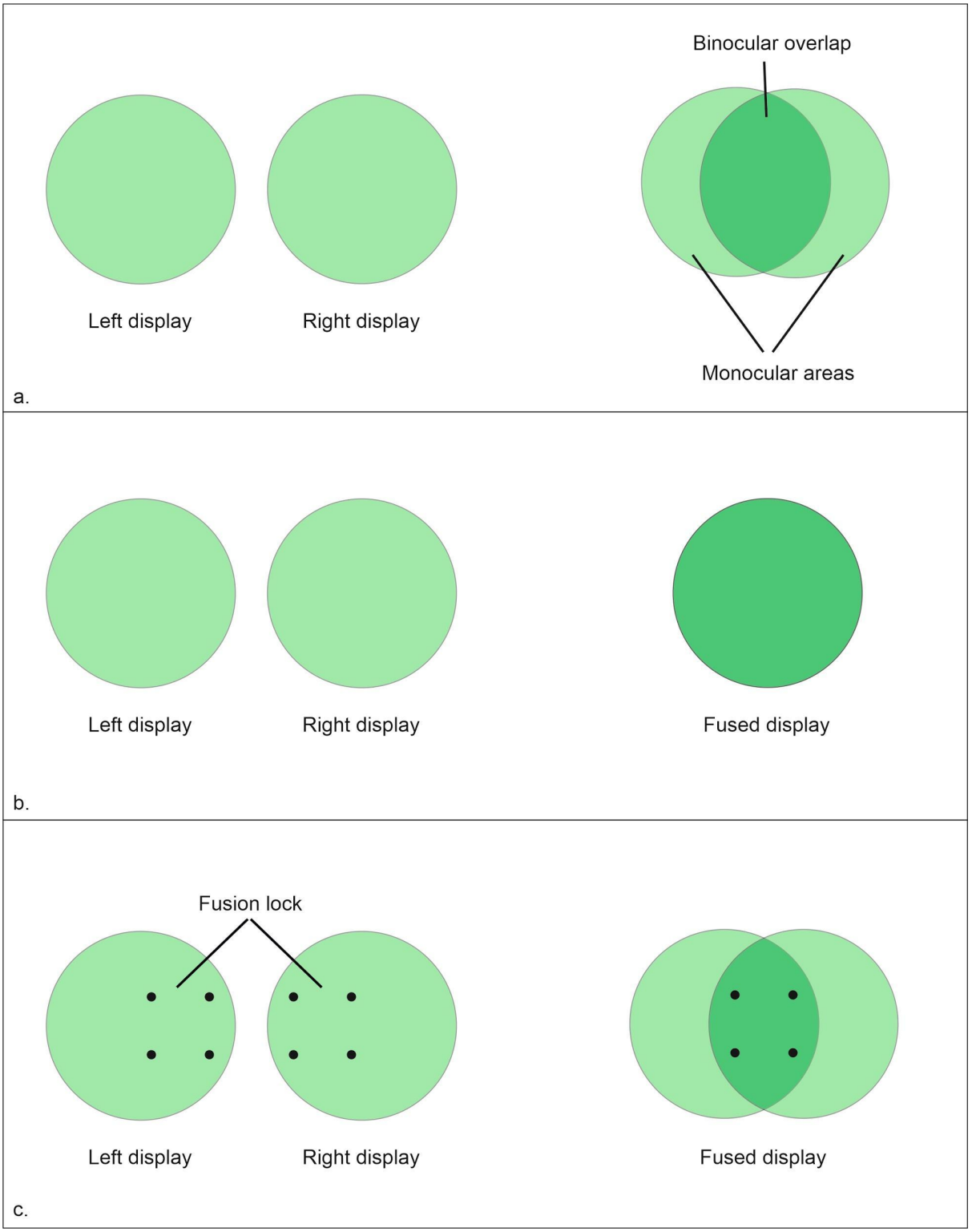



*Figure 5. a. FoV can be expanded by partially fusing the display. This creates a central binocularly overlapping area and left and right monocular areas. b. In practise the displays are a strong stimulus for binocular fusion and the displays are completely fused. c. Partial overlap can be achieved by using fusion locks. These are the four black dots that are fused by the binocular vision.*

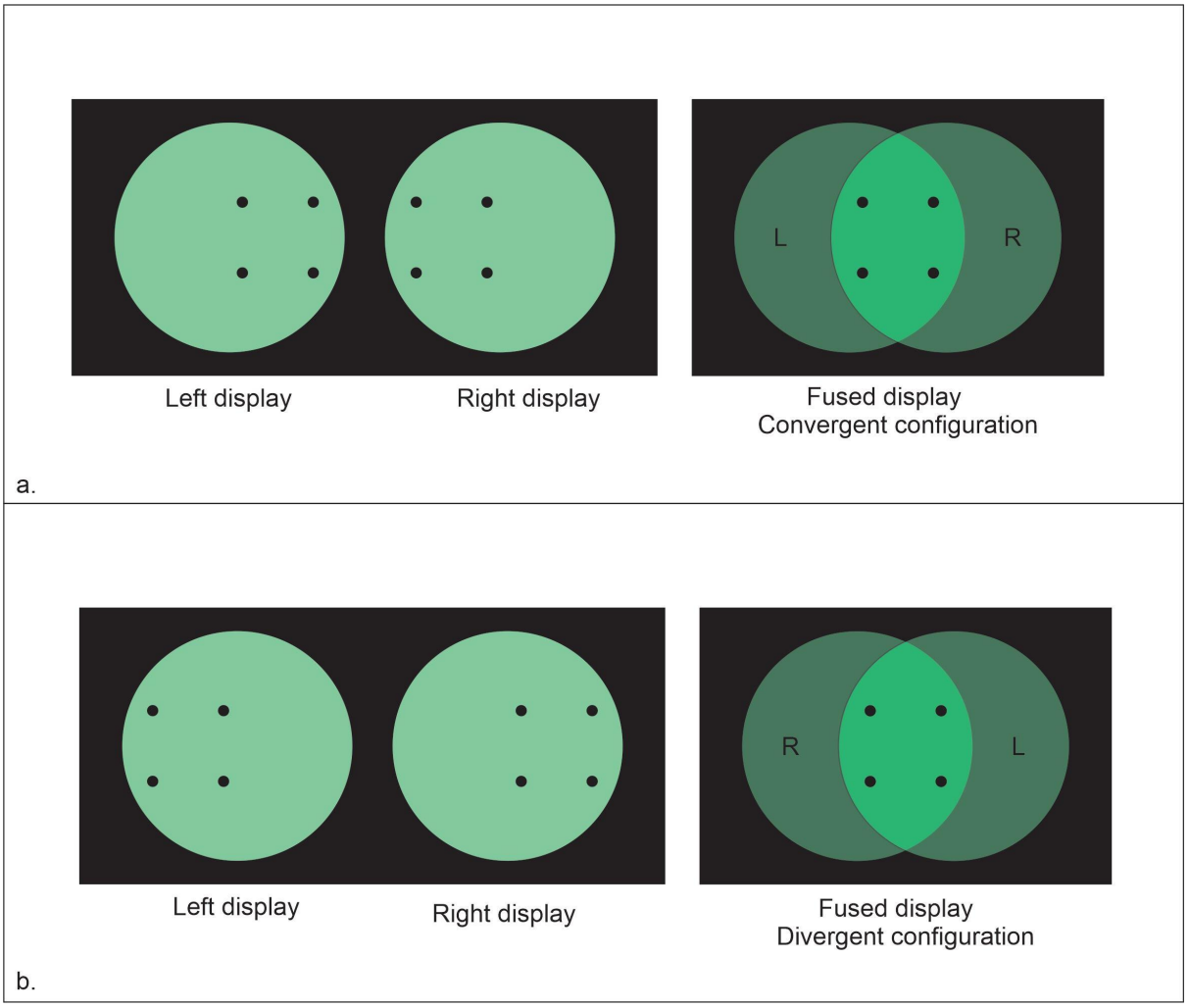



*Figure 6. a. The bright display areas are always surrounded by the device, which is usually an opaque dark area in the visual field. When bright displays are partially fused, the central fused area (brighter green) is surrounded by the monocular areas (darker green) that overlap with the opaque areas. The location of the fusion lock determines the resulting fused display configuration. In this case the resulting configuration is called convergent. The monocular area that is on the left side of the central binocular area is visible to the left eye and the monocular area on the right side of the central binocular area is visible to the right eye. b. In the divergent condition the location of the fusion lock determines that the monocular area on the left is visible to the right eye and the monocular area on the right is visible to the left eye.*

### 2.1.3 FoV size recommendations

#### 2.1.3.1 Immersive experiences use case

**Typical display type:** Opaque display, simple magnifier optics

**Tasks:** Games, movies, architecture, design, simulation, modeling, therapy

Requirements: Larger FoV increases sickness but increases also immersive experiences, like presence (Weech et al, 2019), making the situation nuanced. Although the relation of the

FoV and experience has been widely studied, definitive conclusions cannot yet be made. It is known that increasing FoV increases sickness in some scenarios and presence in others, though usually not both within the same users and settings. Beyond these observations, there is currently no general consensus in the literature on a more specific characterization of the CS-presence relationship, or even a known maximum FoV where the likelihood of inducing sickness plateaus. Small sample sizes and failure to control for mediating factors are likely to contribute to contrasting findings in this area. Controlling for such mediating factors is indeed difficult. XR technology has developed at a rapid pace, and a multitude of updates to hardware and software have come about which can interact with the FoV, making comparison of results obtained years apart complicated. Even comparison of results obtained within the same timeframe can be problematic when the XR content used to test FoV effects is fundamentally different across different studies.

A systematic investigation using a large sample, controlling for hardware and software specs, and manipulating content across conditions would help to resolve withstanding issues. While further research is needed, we can give some preliminary recommendations based on the types of factors that are currently believed to mediate the relationship between FoV and cybersickness:

- Type of content: XR experiences that demand a deep level of presence should opt for a wider FoV than those where presence is less relevant

- Target user: XR experiences intended for certain populations that are more likely to resist cybersickness, e.g. pilots or experienced gamers, could opt for a wider FoV than for those intended for general audiences

- Level of vection: XR experiences with high likelihood of inducing vection should avoid using a wide FoV whereas those with low vection may tolerate a wide FoV

- Sensory modality: XR experiences that deliver high fidelity signals to multiple sensory modalities are less likely to induce cybersickness and thus might tolerate a wider FoV than those that deliver high fidelity information only within the visual modality; if operating exclusively within the visual modality, higher resolutions are more likely to induce cybersickness (Chang et al, 2020) and thus might benefit from adopting a smaller FoV

#### 2.1.3.2 Opaque monocular display

**Typical display:** Opaque monocular display, many possible optical solutions

**Tasks:** Maintenance, warehouse, factory

**FoV-related factors that affect the quality of experience:**

- **FoV size and the amount of outside world occluded.** Too large of a FoV should be avoided because it increases the size of the device, which occludes a larger part of the outside world. The possible problems caused by occlusion are:

    - **Disruption of stereopsis.** Accurate three-dimensional perception at near distances is based on the comparison of the views of the left and right eye. When part of the visual field is visible to one eye only, accurate three-dimensional perception is disrupted.

    - **Binocular rivalry**. If the left and right eye see different views, binocular rivalry can occur. Perceptually this means uncontrollable alternation of the left and right views. This is annoying, leads to visual and cognitive strain, as well as may prevent the user from seeing critical signals either from the outside world or from the display.

      Rules for rivalry:

        a. Visual field with higher resolution dominates
        b. Brighter visual field dominates

      Suggestions to alleviate binocular rivalry in opaque monocular displays:

        1. Present monocular information for very brief durations
            - Perception of stereoscopic fusion and depth occurs very rapidly (~10-20 ms) while onset of binocular rivalry takes longer (~200 ms), so presentations sufficiently long to communicate information yet still less than 200 ms could be employed (Patterson et al, 2007). The minimum sufficient duration will likely vary depending on information content.

        2. If longer durations are needed, periodically manipulate contrast of monocular information to restore dominance to the opaque monocular display
            - A contrast increase corresponding to 0.5 log units (or equivalently an increase in d' of 1.5) could compensate for the loss of sensitivity in the non-dominant percept (Patterson et al, 2007)

        3. Alternatively to (2), present a "probe" within the critical content region of the monocular display
            - A probe, e.g. a small bright dot, flashed onto a suppressed region can speed reversal time (Blake et al, 1990; Fox 1991)

and a probe flashed onto an object in the suppressed region (as opposed to somewhere else in the background within view of the non-dominant eye) can further speed reversal time (Metzger & Beck, 2020)

Note: More research is required in this area as these suggestions have yet to undergo empirical testing in live applications.

- **Rotation limits of the eye and the FOV.** If the monocular display is device-referenced, where head movements cannot be used to bring areas of interest to central vision, the limits of the eye rotation should be considered. Because maintaining a non-central gaze position strains the eyes, the visual system operating in a natural environment uses coordinated movements of the head and eyes when focusing on areas of interest. Saccades move the eyes quickly to the target, after which the head turns toward the target. As the head moves, the eyes make slow compensatory movements so that the fovea remains on the target. After about 1 second, the head is turned toward the target and the eyes are close to the primary straight-ahead position of gaze. This strategy of coordinated head and eye movements cannot be used with a device-referenced system. This limits the useful display area; comfortable saccade length from the primary central position of the gaze is 24°. The gaze can physically reach the +-40° area around the central vision, but this strains the eyes.

- **Display location in the visual field.** The position of a small monocular display is determined by task requirements and eye strain. Positioning the display in front of the eye is least straining for the muscles that control the eye movements, but critical information in the outside world might be occluded. Positioning the display away from the central visual field increases the visibility of the outside word, but increases the possibility of eye strain. Occasional glancing of a non-central display is most straining if the display is at top of the visual field or at the temporal side of the visual field. Bottom parts of the visual field are most comfortable as the human eyes have adapted to occasional downwards glancing.

#### 2.1.3.3 See-through display in stationary use

**Typical display type:** Ground referenced optical or video see-through display (e.g. Hololens)

**Tasks:**

- Construction, surgery, technical support, architecture design review, structural models visualization, interior design review https://www.intellectsoft.net/blog/microsoft-hololens-usage-in-construction/

**FoV-related factors that affect the quality of experience:**

- The FoV should be large enough so that the user does not need to turn his head all the time to be able to see everything. The small active FoV of an augmented reality device is like a light beam that only lights a small area at a time. Small FoV causes difficulties in image, map, and event understanding as well as in keeping moving objects in the visual field (Dolezal, 1982). It also affects visual search of targets (Wells et al, 1989; Piantanida, 1992; Arthur, 2000), impairs layout and object location memory (Alfano & Michel, 1990), eye-hand coordination (Alfano & Michel, 1990; Dolezal, 1982), depth perception (Psotka, 1998; Nolan et al, 2012) and situational awareness.

#### 2.1.3.4 See-through display in mobile use use

This is similar to the stationary case, but mobility increases the need to have large tFoV to through which the outside world is visible, as the user needs sufficient situational awareness of the environment. Too small tFoV causes difficulties in walking (Alfano & Michel, 1990) and obstacle avoidance (Toet et al, 2008ab; Jansen et al, 2011).

## 2.3 Resolution

Display resolution affects both experience and performance. Higher resolution produces higher image quality, which leads to higher immersion. Higher resolution also increases readability, legibility, and detection accuracy, which leads to decreased visual and cognitive strain.

Display resolution is inversely related to FOV. In other words, increasing FOV with a specific display decreases the resolution.

Based on the sampling frequency of the eye, 1 arc minute pixel size, i.e., 60 pixels per degree of visual angle (or 30 cycles per degree), is often considered the limit after which the pixels are no longer visible and the resolution is sufficient.

However, users can differentiate values beyond 30 cpd. Masaoka et al (2013) demonstrate that there is an increase in the realness ratings until 155 cpd. This means that the displays near 310 ppd would be nearly indistinguishable from real images. Similarly, Park et al (2019) suggest that there is experience benefit with very high pixel densities. The results indicate the perception cannot be explained by the retinal sampling. Other factors, such as hyperacuity, affect the perceptual experience.

Recent efforts have attempted to take advantage of resolution discrepancies across the retina by employing what is called “foveated rendering” (e.g. Patney et al, 2016). Visual acuity is known to be best at the fovea and degrade as items fall into the periphery. This gradation is owed to the relative densities of rods and cones across the retina and the differentiation into magnocellular and parvocellular pathways that carry visual information to the brain. Foveated rendering uses eye-tracking to monitor where the eyes are foveating and renders high resolution video to these parts of the visual field while rendering lower

resolution to the rest of the visual field, thus maximizing visual quality where it matters most and sacrificing quality where it is less needed.

### 2.3.1 Modulation transfer function

The perceived image quality does not depend only on resolution, as the optics and other components of the XR system also affect the perception. This can be characterized with modulation transfer function (MTF), which indicates how a modulation of contrast differences between two adjacent image features with various spatial frequencies affects the ability to see small contrast differences.

## 2.4 Luminance and contrast

The luminance of the display should be high as the visual acuity and contrast sensitivity are better with higher luminances (de Valois et al, 1974; Bierings et al, 2019).

However, the luminance should not be too high, as it can reduce visual performance, cause visual discomfort and with extremely high values, even eye damage (above 300.000 cd/$m^2$).

Localized high intensity light can cause scotomatic glare (photostress/flash blindness), which is perceived as an afterimage that fades after approximately 30 seconds (Glaser et al, 1977). Scotomatic glare is caused by the temporary excessive bleaching of retinal photopigments. The afterimage can disrupt the visual performance if it overlaps critical information.

### 2.4.1 Daylight conditions

With optical see-through displays the relation of environment luminance and XR display luminance is crucial. The display should have a luminance that provides good visibility against environment illumination. If the environment luminance is too high compared to the display luminance, the contrast in the display becomes too low and performance suffers. With very high luminance the display becomes completely washed out and cannot be used.

With combiner-based XR displays, the contrast (C) is the ratio of the virtual display luminance and the environment luminance that reaches the eye:

$C = L_{Virtual\ display} / L_{environment}$

where

$L_{Virtual\ display} = r_c\, t_o\, L_c + L_{environment}$

$r_c$ = reflectance of the combiner
$t_o$ = transmittance of optics
$L_c$ = Image source luminance

$L_{environment} = t_v\, t_c\, L_a$
$t_v$ = transmittance of the visor

$t_c$ = transmittance of the combiner
$L_a$ = ambient luminance

In high environment illuminance 30% visor transmittance has been recommended.

High environment luminance may also cause discomfort glare, where the user is able to use the optical see-through display, but high environment luminance causes discomfort and reduces task performance (Stringham et al, 2003).

A possible countermeasure is utilization of filters that reduce the intensity of illumination that reaches the eyes. However, filters that are effective in daylight conditions may reduce visual performance in dim environments.

### 2.4.2. Low light conditions

In low light conditions too bright an XR display may be problematic.

#### 2.4.2.1 Reduced dark adaptation

Too high display luminance diminishes the dark adaptation of the eye and thus disrupts visual performance in the environment. With an optical see-through display the performance reduction occurs for the environment seen through the device and with an immersive display the performance reduction occurs after the user removes the XR device. Reaching the maximum visual performance level in the dark takes 30 minutes.

#### 2.4.2.2 Dazzling glare

If the user has been in the low light conditions long enough to be dark adapted, suddenly looking at a bright XR display can cause a discomforting sense of excessive brightness and momentary visual disability, which is called dazzling or saturation glare (Vos, 2003). Dazzling glare resembles discomfort glare, but the experienced discomfort is higher. The momentary visual disability is higher than in disability glare.

Dazzling glare often causes behaviors that aim to reduce the amount of light reaching the eye such as squinting, wincing or directing eyes away from the light (Stringham et al, 2003).

The neural mechanisms of the experienced discomfort with dazzling glare are not clear, but they are assumed to be related to rapid constriction of iris under bright light and vasodilation of the trigeminal nerve (Stringham et al, 2003).

### 2.4.3 Interocular luminance differences

Maximum allowed luminance difference in bi- or binocular display is 10%.

## 2.5 Framerate

The framerate of a display determines the apparent smoothness of the motion of the display's content. Too low a framerate is visible as flicker, which many individuals can observe by viewing an old CRT monitor through peripheral vision. The threshold value for the perception of flicker, known as the critical flicker fusion frequency (CFF), is determined by the Ferry-Porter law which defines the CFF as a function of the log luminance of the stimulus:

$$CFF = k(\log L - \log L_0)$$

where $L$ is the luminous intensity of the stimulus, $L_0$ is the threshold intensity, and *k* has a typical value of about 12 Hz/decade (Ferry, 1892). The CFF can range from less than 10 Hz in low luminance conditions to greater than 70 Hz in high luminance conditions (Tyler & Hamer, 1990). It would be convenient if luminance were the only factor to consider, but as the example of the CRT monitor above foreshadows, the CFF is not uniform across the visual field and it is also subject to individual differences. Some factors that affect the CFF include (Kalloniatis & Luu, 2007):

- luminance (brighter is more likely to flicker)
- spectral composition (lower wavelengths more likely to flicker at lower luminances)
- retinal position (periphery more likely to flicker than fovea)
- stimulus/display size (larger items more likely to flicker)

High amounts of flicker can cause nausea in some users and the close proximity of the display to the eye in HMDs only exacerbates the potential to detect flicker, thus a framerate of 90 fps or greater is recommended (Champel et al, 2017). Even in the absence of adverse effects, users still significantly preferred a 90 fps display to displays of 60 fps and lower in a high motion 360° video experience (Hofmeyer et al, 2019).

## 2.6 Color breakup

The colors in displays are typically created with adjacent subpixels that produce low, middle and long wavelength light. However, color can also be produced temporally with colored pixels flashing in time sequence, which produces a set of subframes that the visual system integrates temporally into a single perception. These sequential color displays are alluring given that they require only one light source per unit space whereas simultaneous methods require triads of light sources, meaning that sequential displays can in theory achieve resolutions three times the PPD of an equivalent simultaneous display method.

In a time-sequential display there may appear a color breakup or rainbow effect in which an object breaks up to the color components.The effect can happen in various conditions:

- The object moves fast
- User makes a large eye movement over the display
- The display is vibrated by external force (e.g. during walking)

In fast object motion and display vibration cases the effect is caused by subframes being in physically different locations. The same applies to the eye movement case. During fast eye movement the location of each subframe is different and thus the colored subframes are physically in separate locations on the retina. Interestingly, this should not be visible due to saccadic suppression that occurs during the eye movement.

The visibility of subframes has been explained by processes that produce perceptual space constancy. These are processes that maintain the stability of the world even though the retinal image moves with each eye movement. The perceived location of an object is determined by the summation of retinal location of the object and the internal information about eye position. Thus, when motion signal from retina and compensatory signal about eye position are summed, the sum equals zero and the retinal movement is not perceived as movement.

It has been reported that the compensation does not work perfectly around the time of the saccade. When a briefly flashed stimulus is presented before, during and after a saccade, the perceived position of the flash is systematically mislocalized (Bockisch & Miller, 1999; Boucher et al, 2001; Dassonville et al, 1999). If the flash was presented just before the saccade, it was perceived mislocalized toward the direction of the saccade. Flash presented after the saccade were mislocalized to the opposite direction.

Watanabe et al (2005) showed that a flash presented during the saccade can be perceived and its apparent location depended on the time difference between the start of the saccade and the time of the flash. When a temporal method is used for color production, the subframes are colored flashes that appear at different time periods, and thus their perceived location may be distorted. The mislocalization is time-dependent, so each of the subframes is mislocalized differently and consequently the colored subframes are seen separately.

There are results indicating the led flicker can be perceived in low light (<1lux) conditions during saccades until 1.98 KHz frequency (Roberts & Wilkins, 2013). This is indicative of the sensitivity of the visual system, but it is unknown whether the experimental conditions used by Roberts & Wilkins allow accurate prediction of results in HMD context.

Recent efforts that have had some success in overcoming the problem of color breakup have included using high subframe display rates of 720 Hz or greater (Abeeluck et al, 2018), small spatial offsets of subframes (Johnson et al, 2014), and in a light field display application, recombination of subpixels with small ray-tracing errors from different elemental images (Qin et al, 2019).

# 3 Virtual display

## 3.1 Vergence distance

Vergence distance is the distance to which the eyes converge to foveate the same point in space. The foveated point is called the fixation point and the axis from the fovea to fixation point is the primary visual axis of the eye. The angle α between the primary axes is the vergence angle.

The angle α can be calculated from the interpupillary distance and distance to the object:

$$\alpha = 2 \arctan(IPD/2x)$$

In which IPD is interpupillary distance and x is the distance in meters from the viewer to the object. The vergence required to binocularly foveate a stimulus is the vergence demand of the stimulus. The vergence demand changes as a function of object distance.

There are various ways to define the vergence demand. It is important to understand the various metrics as different metrics are used in different papers, which might create confusion (Shibata et al, 2011).

One possibility is prism diopter ($^{\Delta}$), which describes the rotations of the eye's gaze from parallel gaze:

$$P = IPD/D$$

P = prism diopter
IPD = Interpupillary distance
D = distance in meters

For example, $1^{\Delta}$ displaces one eye's gaze direction horizontally by 1 cm at a distance of one meter.

This metric is used by ophthalmologists, as it is convenient with base-in and base-out prisms used to change the convergence demand. A disadvantage is that the metric depends on the observer's interpupillary.

An alternative metric is meter angle (MA) that eliminates the dependency on Interocular distance:

$$MA = 1/D$$

D = distance from the midpoint of the interpupillary axis to the stimulus in meters

Meter angle is convenient as it is equivalent to diopters as used to represent the focal distance (i.e. the level of focus).

The main difference between these measures is that the meter angle is a measure of stimulus distance while prism diopter is a measure of the angle an eye must rotate to converge at a particular distance. The conversion from prism diopters (P) to meter angles is

MA = P / IPD

The meter angle is mathematically equivalent to the diopter used to characterize refraction. Thus, the term diopter can be used to characterize both the required vergence and required accommodation level (Shibata et al, 2011).

Selecting proper vergence distance is important, as inappropriate distance may cause eye strain or double images.

**Recommendations**

- The vergence place should be at the same plane as the focal plane
- The vergence plane should not be too near. Absolute minimum is 15 cm.
- The vergence demand of the system set at infinity may be problematic, as this increases the probability of cases with negative vergence, which is not comfortable.

## 3.2 Interocular distance

A typical way to achieve a specific convergence demand is to adjust the distance between virtual displays to create a proper convergence demand. In this context it is relevant to consider the interocular distance (IOD) of the device, which is the distance between the centers of exit pupils of optical systems in a biocular or binocular XR device.

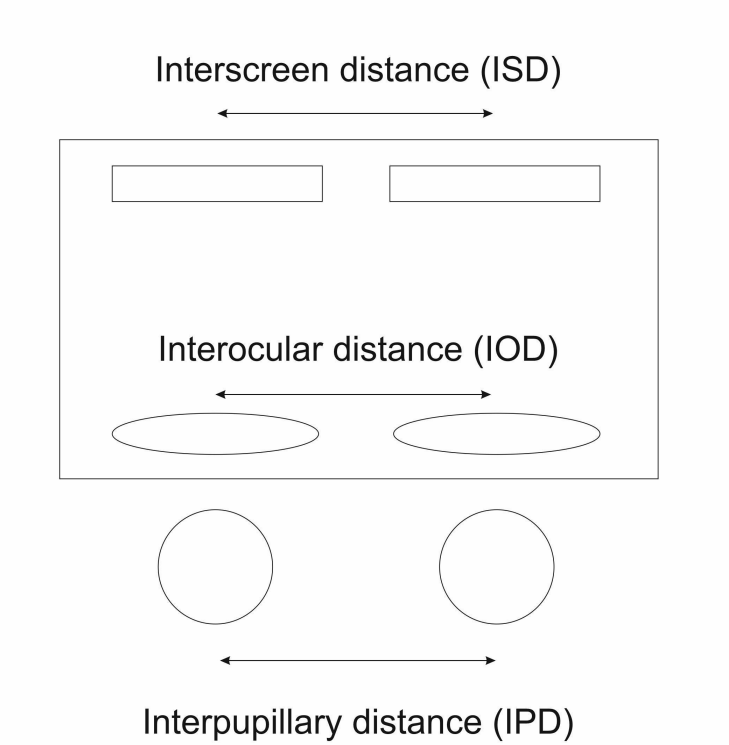



*Figure 7.*

The IOD and ISD of the system should match the human interpupillary distance, which is defined as the distance between pupils when visual axes of the eyes point straight ahead. If there is convergence, the IPD is reduced. For example, with a convergence angle of 8 degrees, the IPD of 63.2 mm is reduced to 61.8 mm.

The reduction R can be computed with the following equation:

$R = 20 \sin \theta/2$

where $\theta$ = convergence angle

If there is a mismatch between the IOD and IPD, various problems may follow. The possible problems are covered in the next sections.

### 3.2.1 Divergence

If the IOD is larger than the user interpupillary distance (IPD) there might be eyestrain, headache or other visual system problems, as this requires divergence of the visual axes, which is not well tolerated by the visual system. Though setups where the IOD is smaller than the IPD can also cause discomfort and disrupt performance (McIntire et al, 2018).

### 3.2.2 Other changes in vergence demand

Additional problems caused by IPD-IOD are caused because the person views the lens system on an off-center axis, which leads to distortion as convex lenses can be regarded as

two prisms, which change the intended vergence demand of the system. The literature recognizes the effects of prism adaptation from a non-VR point of view (Sheard, 1934; Lambooij et al, 2009).

The prismatic effect can be calculated with Prentice's rule:

Prismatic effect in prism diopters = Lens power [D] x decentration [cm]

### 3.2.3 Reduction of uFOV

IPD-IOD mismatch may reduce the size of uFOV as parts of the FOV are occluded or distorted for users with discrepant IPDs. This can be more problematic with systems that have a small eye motion box.

### 3.2.4 Distorted depth perception

Another possible effect of IOD-IPD mismatch is distorted depth perception, where IPD>IOD causes objects to be seen too far away and IPD<IOD too near (Yamanoue et al, 2006). However, these effects are probably so small that they do not have relevance in consumer applications. In tasks requiring accuracy there may be relevance. These could be related to medical, military or civilian aviation use cases.

### 3.2.5 IPD requirements

The XR system should match the IPD of the intended user group, so the design should be based on population values shown in the table below (Dodgson, 2004). The adult population mean of IPD is 63 mm, but the variation is between 45 - 78 mm.

With children the values are lower. If children need to be accommodated, the minimum is 40 mm. If babies need to be accommodated, the minimum is 30 mm.

| Percentage of population included | Range | Comments |
|---|---|---|
| 96% | 55.3 - 72.7 mm | 96 % of male/female population |
| 98% | 54.5- 73.8 mm | 98 % of male/female population |
| 99.9% | 50.0 – 78.0 mm | 99.9 % of male/female population |

| | | |
|---|---|---|
| 99.99% | 45.0 – 78.0 mm | 99.99 % of male/female population. Rare cases included |

*Table 2.*

### 3.2.6 Changing the IOD

Many current XR systems have the possibility of changing the IOD. This reduces the problems with IOD-IPD mismatch. On the other hand, it is not clear whether the IOD selected by the user is appropriate and whether users are willing to spend time optimizing the IOD.

Asking users to measure their own IODs using explicit measurement methods has led to inaccurate and unreliable results. One study asked 52 participants to measure their own IODs by following online instructions that detailed a method using a ruler and mirror, and then compared these user derived measurements to those taken of the same users by a trained examiner using a pupillometer. The researchers found poor agreement between the user and professional measurements, as users overestimated their own IODs by about 0.5 mm on average, with a 95% confidence of more than +/- 5 mm (McMahon et al, 2012). Measurements taken by friends and mobile apps within the same study were even less accurate than self measurements.

Rather than relying on users to know or explicitly measure their own IODs, commercial headsets often use visual display cues intended to help users to implicitly approximate the IOD to the proper setting, though to our knowledge, there are no studies that can attest to the efficacy of these methods. There is clearly a need for further research.

A novel approach to IOD calibration is one where user error is altogether bypassed by implementing hardware and software to automatically detect and set the IOD (e.g. Son et al, 2019). Son et al note that they achieved accuracy within 1 mm in their implementation but that the motors required for their headset's IOD adjustment caused it to be undesirably heavy.

## 3.3 Focal plane

The refractive power of the lens of the human eye can be changed by the ciliary muscles surrounding it. This process is called accommodation. With further objects the lens flattens and refractive power decreases. With near objects the curvature of the lens increases, increasing refractive power. If the ciliary muscles relax and the lens is at its flattest, the focus is at infinity and the eye is unaccommodated.

Accommodation is expressed in diopters, which are calculated as inverse of the object distance from the eye in meters. For example, an object at 0.25 meter distance requires accommodation of 1/0.25m = 4 diopters. Completely accurate accommodation to the target

is not needed as the eye has a depth of focus in which objects near the accommodated distance appear sharp (Wang & Ciuffreda, 2006). Accommodation does not stay constant over time, there is continuous variation, which can be up to 0.5 diopters.

With empty field and/or dark field the accommodation settles to 1.7 diopters (focus distance of 59cm) on average (Leibowitz & Owens, 1975). There is individual variation in this.

From the XR device point of view the properties of the accommodation system determine the focal distances that should be used. The most important guideline is the avoidance of too short focal distances. If the focal distance is too near, the user may see the image completely blurred, or may have difficulties in maintaining constant focus, or may start to experience eye strain. The strain comes from the fact that the nearer the focal plane, the more the muscles have to work to keep the focus. Thus, near focal distances strain the accommodation system of the eye more.

For young adults the nearest distance they can accommodate is about 25 cm. This is the near point of accommodation. When expressed in diopters the point is called the amplitude of accommodation.

Amplitude of accommodation decreases as a function of age. In practice this is experienced as difficulty in focusing to near distances.

The amplitude can be calculated by 18.5 - (0.30 * age in years).

**Recommendation:**

It is important to notice that for a long term use the accommodation amplitude is not an appropriate guideline as users cannot maintain accommodation that is more than one third of their accommodation amplitude for long durations. Thus, a sufficiently long focal distance is advisable in XR devices. Focal distance of 2 meters covers many AR use cases.

## 3.4 Properties of stereoscopic contents

Stereoscopic disparity or binocular parallax refers to the differences in the views of the left and right eye. The visual system computes the depth perception from horizontal (or lateral) disparity. To understand disparity, the basic terms of visual space need to be defined. When the left and right eye point to the same location in visual space, this is called the fixation point. All the points that are in the same depth level with the fixation point are at the fixation plane. Perceptually the plane is slightly curved toward the observer in peripheral vision, but the magnitude of curvature is very small.

If a person fixates to point F and an object O is further away or nearer from the fixation point, there are alternative names for the location of the point. The naming convention is different in vision science and the movie industry.

In the vision science the term uncrossed disparity is used to describe far depth but in the movie industry it is called positive parallax.

In the vision science the term crossed disparity is used to describe near depth but in the movie industry it is called negative parallax.

From the XR device point of view the most important issue is the avoidance of too large disparities. If disparity is too large, there are a several perceptual consequences:

- Eye strain
- Blurred vision
- Double vision (diplopia)

There are individual differences in threshold values for these effects. Furthermore, context sensitive factors, such as length of use and existing eye strain affect the thresholds.

To avoid the negative effects a rule of one degree has been used. According to the rule a maximum disparity of one degree of visual angle is sufficient for comfortable viewing. In addition to this, Shibata et al (2015) offer additional guidelines:

- Near disparity should not exceed 3-4% of screen width
- Far disparity should not exceed 1-2% of screen width.

The maximum disparity is affected by the following parameters::

- The shorter the presentation time is, the less the maximum possible disparity should be. This rule applies to presentation times below 250-500 ms.
- If objects are fused when they have small disparity and then the disparity is increased gradually, the maximum possible disparity can reach much higher than normal. This phenomenon, called hysteresis, can increase the double vision thresholds significantly (Fender & Julesz, 1967; Hyson et al, 1984; Piantinada, 1986; Diner & Fender, 1987). The effect is asymmetric: When the limiting value is reached and the image is perceived as diplopic, it is necessary to go back to low disparity values so that the image can be fused again. Although the hysteresis effect allows higher depth tolerance, it does not prevent eye strain.
- If the object is not in the central vision, i.e., the eye is not pointing to the object, larger depth values are tolerated. The double vision threshold increases 0.13 degrees per degree towards peripheral vision (Hampton and Kertesz 1983). The larger tolerance is based on the cortical magnification factor, which describes the transformation by which the visual space, as imaged on the retina, is mapped onto the visual cortex (Rovamo & Virsu, 1979).
- The diplopia threshold depends on the highest visible spatial frequency components visible in the image. The more high frequency components present in the image, the lower the diplopia threshold. In other words, blurred images can have larger disparities than sharp images. With very blurred images the diplopia threshold can increase from 1° to 90° (Schor, 1984; Kulikowski, 1978). This has practical implications. In the real world the fixated object is clear, but due to the depth-of-field of the human eye nearby objects are blurred. This increases the diplopia thresholds in everyday scenes.

- Diplopia threshold increases if the object is larger. This is due to two phenomena. Firstly, with larger objects one or both object contours are always further in peripheral vision. Secondly, larger objects have low spatial frequency components, which increase the threshold.
- If there is only a single object in an otherwise uniform scene presented with a stereoscopic device, the object gets more easily diplopic. The effect occurs more easily if the person viewing the stimulus has phoria. This can be prevented by adding additional structures in the scene. These are called fusion locks.
- If there are several disparate objects in the screen and they are viewed with central vision, then the disparity limit for maintained fusion of two points decreases as the angular separation between the points decreases. The maximum possible disparity can be calculated from the disparity gradient between the objects (Burt and Julesz 1980; Burt and Julesz 1980):

  Disparity gradient = Disparity difference between objects/ Distance between the objects

  If the disparity gradient value exceeds 1.0, double vision occurs. Notice that this result applies to situations where both objects are sharp, i.e., they are both presented in stereoscopic display. In real life situations the other object is usually at least slightly blurred due to the depth of field of the human eye.

# 4 Optical system

## 4.1 Distortions

### 4.1.1 Pincushion distortion

Pincushion distortions occur when magnification increases with distance from the optical axis, resulting in bowed-in edges of the generated display. While the distortion can be noticeable and annoying to those actively looking for it, Kuhl et al (2009) found no detrimental effect of pincushion distortion on distance estimation and no subjective awareness of even an exaggerated amount of pincushion distortion (beyond that typically found in uncorrected HMDs) for naive observers in a modest twelve participant sample.

### 4.1.2 Chromatic aberration

Chromatic aberration occurs when different wavelengths of light passing through a lens are refracted at different angles and fail to converge at their focus point. Whereas color breakup can be thought of as a distortion of color in the temporal domain, chromatic aberration is a related distortion in the spatial domain. Chromatic aberration is also known as “color fringing.” In the extreme case, it is akin to creating a rainbow via a prism. In commercial lenses, the artifact manifests as the appearance of ghost-like greenish and purplish distortions around objects. Many currently available commercial HMDs demonstrate

evidence of chromatic aberration, most notably towards display peripheries (Beams et al, 2019). Chromatic aberration in HMDs should be addressed to the extent that creates deviations from our normal visual experience, and approaches exist for correcting it (e.g. Zhan et al, 2019; Wang et al, 2016; Boult & Wolberg, 1992). Interestingly, at least one study found benefits in leveraging the natural chromatic aberration produced in the human eye by rendering some amount of chromatic aberration to the display in order to increase realism (Cholewiak et al, 2017).

## 4.3 Eye relief

The eye relief is the distance between the outer surface of the nearest optical element or supporting structure of the HMD to either cornea of the eye or to the exit pupil of the eye. The exit pupil of the eye is located 3 mm behind the front of the cornea.

The eye relief should be as large as possible. However, a large eye relief might make the system larger and reduce the size of the perceived FOV.

Too small an eye relief can generate discomfort when the eyelashes touch the optics. Eye-glasses have an eye-relief of 12mm, so this is probably the minimum value.

Sufficiently large eye relief is important to eyeglass wearers. Eye glasses should fit comfortably underneath the HMD optics with no physical contact. At least 25 mm is a good value.

If small eye relief is required, the alternative is to provide refractive correction for each user.

## 4.4 Eye motion box

The eye motion box is the volume of space where the observer’s pupil must be positioned in order for the observer to see the entire FOV of the display.  The optimum location of the eye is sometimes called the design eye point. If the eye is not in the eye motion box, uFoV decreases as parts of the display may be occluded or distorted. If the eye motion box is small, the device should have a design that allows easy wearing of the device and that keeps the device in the correct position during the use. A combination of a small eye motion box, difficult wearability and/or a fit that allows the device to slip in the head can be very frustrating to the user. Furthermore, in tasks where quick reactions and good situational awareness is needed, this could lead to missing of task-critical information.

The size of the volume is determined by the lens diameter, the lens focal length, and the size of the display:

E = eye motion box size

D = lens diameter

S = display size

F = lens focal length

L = eye relief

E = D – (L S / F)

The calculations vary depending on the optical properties of the technology used.

## 4.5 Misalignments

In a binocular HMD there might be small differences between the two display positions. The differences can be in size (magnification difference), orientation (rotation difference) or location (vertical or horizontal misalignment).

These might cause sickness and vision problems (diplopia, blurred vision).

With optical axes misaligned, or with differences in magnification or rotation, all corresponding points in the two images are misaligned both vertically and horizontally. With axis misalignment, the amount of misalignment is uniform over the field. With either rotation or magnification differences, angular misalignment increases with angular distance from the field center.

Here, the adverse symptoms and vision problems are caused by the strain of the muscles that control eye movements. These muscles can compensate for misalignment to certain extent, but after threshold values various QoE problems appear.

### 4.5.1 Vertical misalignment

Vertical misalignment causes sickness very easily. Maximum misalignment is between 7-15 arc minutes.

### 4.5.2 Rotational differences

The difference can be calculated based on the maximum possible vertical misalignment. With rotation difference, R, maximum vertical misalignment, V, in the field of view (FOV) is on a horizontal line through field centers at the field's edge. Maximum vertical misalignment (V) in minutes of arc for a rotation R in minutes with a total FOV of F degrees is:

$$V = R \sin (F/2)$$

### 4.5.3 Magnification differences

With a magnification difference (aniseikonia) of d percent, the largest V in the FOV is at the edge of the FOV on a vertical line through the field center. For a total FOV of F degrees, V in minutes is:

$$V = 0.3 Fd$$

# 5 Latency

Latency is the small delay that occurs when a virtual reality system reacts to human movements.

Earlier studies show that adaptation to virtual environment is reduced with tracking lag delays over 60 ms and clearly impaired with delays over 120 ms. The feeling of virtual presence starts to break down after a delay of 200 ms. Similarly, Fieldman & al. (1992) conclude that lags greater than 100 ms in the rendering of hand motion can cause users to restrict themselves to slow, careful movements while discrepancies between head motion and rendering can cause sickness symptoms. However, the older long delay results are no longer relevant in modern VR/AR systems, as the technology has developed significantly.

A modern way to describe the performance of a VR system is to use the term motion-to-photon latency, or MTP. It means the time it takes from user motion to corresponding change in the display.

A high motion-to-photon latency causes problems: motion sickness, nausea, reduced presence experience and reduced performance.

All the components of the system affect the MTP. For example, the following parts are important:

- Pixel switching time: The time needed to update all the pixels in the display. This can be improved by display technology, for example by switching from LCD to OLED
- Refresh rate
- CPU, GPU and game engine techniques
- Head tracking

An MTP latency of less than 20 ms is regarded as good (Champel et al, 2017). Above this threshold, participants start to become aware of the latency (Yang et al, 2019; Adelstein et al, 2003; Ellis et al, 2004). MTP latencies between 35 and 45 ms begin to trend toward negatively impacting quality of experience, and by a latency of 55 ms, significant disruptions to comfort and immersion, as well as significant increases in cybersickness, are observed (Brunnstrom et al, 2018).

# 6 Weight and center of gravity

The HMD should be as light as possible. As HMD weight increases and/or center of gravity becomes more imbalanced, users experience greater torque in head rotations and rate their experiences as less comfortable, with rotations requiring the head to look up impacted most strongly (Chihara & Seo, 2018). There tends to be a natural tradeoff between adopting the most advanced hardware specs and keeping the HMD lightweight. While surprisingly little research has been performed in this area, it stands to reason that this balance should be attended, as users will be unlikely to spend much time with even the most dazzling displays if the experience is accompanied by soreness and strain.

Other variables

- fit to head
- nose pads
- cable management

# 7 Tips for further research

As discussed briefly in the introductory section, the field of XR research is still relatively young and there remain many unknowns. The majority of the topics discussed in this document deserve further investigation in their current state and this sentiment only grows as XR technology and peoples' expectations regarding the technology continue to evolve. This section offers suggestions that can help developers conducting their own research obtain high quality, reliable answers to these evolving questions.

## 7.1 Testing

How will you evaluate quality of experience? Many options exist, from observational methods like analyzing existing user data or conducting interviews, to empirical methods like conducting an experiment, and each provides value depending on the context. Observational methods are great for discovering trends and generating hypotheses, whereas only experiments can truly evaluate the evidence for hypotheses and isolate root causes for different outcomes. There are many avoidable, yet often neglected, pitfalls that arise from conducting and interpreting experiments, so it is here that we will primarily focus for this section.

## 7.2 Experimental Design

If you have done your observational research and arrive at a hypothesis worthy of testing with an experiment, there are some key factors to consider to ensure a sound experimental design, discussed in the following sections.

### 7.2.1 Independent variables, controls, & confounds

What will be your independent variable? This is the element that you will change between your experiment's conditions which you expect will ultimately have an effect on your outcome measure. Do we need to increase the refresh rate from 60 Hz to 90 Hz in order to achieve a smooth motion experience? An experiment with refresh rate as your independent variable can help to address this question. In the experimental group (or condition), the independent variable will be set to a value expected to cause a change to the outcome (90 Hz in this example), whereas in the control group (or condition) the independent variable will either be missing or, more often, set to a value expected to cause no change to the outcome (60 Hz in this example). This control group serves as your baseline. Say your experimental group shows some effect. What does it mean? Without comparison to a control group, our baseline, we have no way of properly understanding how to interpret the effect.

Finding a proper control can be trickier than it might seem at face value. Imagine in another example we are implementing software in AR glasses that tracks eye movements and displays corresponding visual cues as a means to train users to increase their attention spans. We recruit two groups of participants that both take an exam that measures attention span on day 1, then the experimental group uses the glasses (our independent variable) every day for 4 weeks. What should our control group do? Let us assume for now that we have the control group do nothing, just going about their lives as usual. At the end of 4 weeks we get both groups back to take the follow up exam and find that the experimental group has increased their attention span slightly more than the control group. While we might be tempted to celebrate the success of our training, can we really be sure that the software was responsible for the effect? What if we obtained a kind of "placebo effect" where the mere act of physically wearing glasses happened to increase attention span and it had nothing to do with the software's cues? This is known as a "confound," and we can remedy this confound by implementing an *active* control group, where the controls also wear the same glasses but do not receive the visual cues, or even better, receive visual cues which are unrelated to their eye movements. Matching our control group as closely as possible to our experimental group in all ways other than in the independent variable gives us our most ideal baseline for properly interpreting an effect.

It can be tempting to try to see how multiple variables affect your outcome at once, however it is absolutely critical that only the variable of interest is changed between conditions while all other variables are held constant. For example, if the field of view is suspected to have an impact on quality of experience, design your experiment so that the content, task, display resolution, etc. are identical between conditions and only the field of view changes between conditions to ensure you are evaluating the effect of field of view without influence from competing variables. These competing variables are also potential confounds to your experiment's design which can prevent you from properly interpreting the effect of your field

of view manipulation. Confounds can be surprisingly easy to miss, thus careful planning is required.

Sometimes even when focusing on only one independent variable, confounds can still creep into the experiment if it is not conducted in a controlled environment. A noisy office or lobby does not make for a good testing environment. Find a quiet room free from distractions. The advantage of doing an experiment comes from this ability to isolate the effect of your independent variable and control for confounds arising from other factors, be they from other aspects of the software or hardware, or environmental distractions.

In observational methods, the ability to control for confounds is highly limited and their potential influence must always be kept in mind.

### 7.2.2 Outcome measures

As described in the introductory section, there are subjective and objective measures which can each provide valuable windows into a user's quality of experience. Some situations may demand one type of measure over the other, but often it is possible to collect both simultaneously. This situation is desired for its ability for each data type to complement the weaknesses of the other; the subjective measures are prone to biases but provide insight into a person's subjective experience, whereas objective measures are less prone to biases but also provide less insight into subjective experiences. Combining both measures can provide a dataset that is greater than the sum of its parts due to each type of measure's ability to complement and temper the other.

When choosing a measure, make sure to check to see what types of measures others have used to probe this type of outcome in the past. There is usually no need to invent a new questionnaire when others have already been established for your task. There are a couple of important reasons for preferring established measures. Firstly, not all measures are created equal. For example, it is important that your questionnaire is both valid, i.e. it differentiates between responses in a way that accurately maps onto the concept you are attempting to measure, and reliable, i.e. it gives back the same or very similar responses for the same people over time, all other things being equal. Most established measures have been vetted for validity and reliability whereas these factors will have not yet been established for a novel measure. Secondly, using established measures allows you to compare your results to those of others, and others can more easily interpret your results. This benefit not only acts as a sanity check for your data but also allows you see how they are situated in a broader context and facilitates accumulation of collective knowledge on a topic. Note that established measures often come with clear guidelines for administration and interpretation. These should be followed. Breaking from these guidelines largely erases the benefits of using an established measure.

### 7.2.3 Sampling

To paraphrase Yao et al (2014), no one knows your product better than you, and it is exactly for this reason that you are your own worst participant. You have adapted to your product and your experience clouds your ability to recognize all of its potential delights and annoyances. This fact necessitates user testing in some capacity. Naive users will more

readily perceive details to which you have become accustomed and will thus serve as better barometers for detecting pain points than those involved in the development of the product.

When the intended consumer base is a general audience, the larger and more diverse a sample you can obtain, the better. Differences among demographics and individuals manifest in ways that can greatly limit your product's ability to be enjoyed by a wider audience. Recall for example population differences in head sizes and IPDs. A biased sample is another potential source of confounds. Acquiring a large and truly *random* sample can help prevent sampling bias. A minimum of 20 participants per group is recommended, though many questions will require larger samples (see section 7.3.2 on statistical power for a more nuanced discussion on sample sizes). When required to work with a truly small sample ($n < 20$), selecting for representative demographics rather than randomly sampling might be preferred, and results should be interpreted more in terms of qualitative suggestions rather than quantitative evidence.

Sometimes your intended audience is a specific one rather than a wide one. It is still important in this case to acquire a representative sample composed of individuals from this specific population for quantitative analyses and to leverage their opinions in qualitative analyses. Do not assume that you understand the perspective of that specific audience; engage with them during development and discover their perspective!

### 7.2.4 Grouping

Will you use a between-subjects or within-subjects design? There are benefits and drawbacks to both.

In a between-subjects design, two separate groups of people experience two separate conditions, usually divided into an experimental group and a control group. Between-subject designs are advantageous when exposure to one condition can influence responses within another condition or when the passage of time is thought to influence responses. They are sometimes the only logical design for assessing certain outcomes. A common example comes from medicine. Participants are often divided into a placebo group, a traditional drug group, and an experimental drug group. Administering all three drug types to the separate groups simultaneously controls for the passage of time, and each group only receives one treatment which rules out interactions among the different drugs that could arise from sequential administration within the same group in a within-subjects design. The major drawback of a between-subjects design is the possibility for sampling bias to impact results, especially in small samples. It is very important that participants are randomly assigned into experimental and control groups in between-subject designs for this reason. Non-random assignment can lead to a form of sampling bias where the features selected for grouping cause the observed effects rather than the independent variable. Because the different groups contain different individuals, it is possible (though increasingly unlikely as sample sizes increase) that sampling bias contributes to effects even when using random assignment.

In a within-subjects design (also known as a repeated measures design), the same group of people undergo both an experimental condition and a control condition. A benefit of this design is that it sidesteps the issue of random assignment because the same individuals experience both experimental and control conditions. However, it is still important that the

sample is randomly selected from the population for which the product is intended so that the effects may generalize to this population. An example of a within-subjects design might have participants try three different movement speeds through a virtual environment to see which is most comfortable. One important consideration for within-subjects design is the notion of counterbalancing. A drawback of a within-subjects design, due to the same participants experiencing all conditions, is that it cannot guarantee that a participant's experience of one condition does not have an effect on how that participant judges a subsequent condition. Counterbalancing helps to address this issue by changing the order that each participant experiences a condition, such that participant 1 might experience the motion conditions in the slow-medium-fast order, participant 2 might get medium-fast-slow, and so on until every possible sequence of condition orders has been experienced by an equal number of participants in the sample. Through counterbalancing we may neutralize order effects when the final estimates are calculated.

Finally, it is also worth noting that mixed designs are possible whereby aspects of both between- and within-subjects designs are implemented. Think back to our example about attention span training with AR glasses. Our sample was split into a control group and a treatment group, like a between-groups design, and each took a pre and post test, like a within-subjects design. Mixed designs can be powerful but note that they still carry the potential pitfalls of both between- and within-subjects designs.

## 7.3 Interpreting results and more notes on experiment planning

Now you have run your carefully thought out experiment and have obtained your data. How do you interpret the results?

### 7.3.1 Null hypothesis testing vs. estimation

From our data we can use statistics to help guide our decisions. We get both descriptive statistics, like means and standard deviations that help us to summarize our data, as well as inferential statistics, like p-values and confidence intervals, which can help us decide what to make of our data. For some context we will present a rough and oversimplified discussion of the p-value. In an ideal world, the p-value tells you how likely your observed sample's distribution is to have come from a population representing a null distribution where the null hypothesis is true and your experimental manipulation has no effect. If this value comes back as .05 or less, we infer that our sample did not come from this null distribution and that our manipulation did have an effect. This is only an inference however, and this threshold of .05 represents our false positive rate, meaning that 5% of the time we will get a p-value that is less than .05 and infer that our sample did not come from the null distribution when in fact it did.

Academic papers put the p-value on a pedestal, with projects dying or thriving based on their ability to pass this arbitrary threshold of .05. Despite its seemingly holy status for academic publishing, the p-value in fact provides very little information about an effect. The little value that a p-value does carry is only as worthy as the (population) effect size, sample size, and experimental design that produced it. It turns out that poor research practices can push any effect below the .05 threshold (Simmons et al, 2011; Ioannides, 2005), and even without implementing any poor research practices, vanishingly small effects become "significant"

with large enough samples, contributing to persistent misinterpretations of the p-value (Cohen, 1994).

So what's so significant about a p-value? For developers, we urge that the answer is not much. Effect sizes and the confidence intervals around those effect sizes are what are really important, because these statistics give useful information about the magnitude of the impact that our manipulation has on our outcome, and an idea of the level of precision that we have obtained for our estimate of this impact. As an example let us imagine we have run an experiment testing the effect of field of view on presence and discover that a larger FoV results in increased presence, $p = .03$. We jump for joy at the knowledge we have gained; the smaller FoV has to go into the final product! Looking further into it, the effect size seems decent: $d = .5$, a medium sized effect. But let us now check the confidence interval on that estimate: $CI = .05 - .95$. That's a huge range! We have discovered that the field of view probably affects presence, but our estimate of how much it affects presence ranges from absolutely game changing to practically unnoticeable. What are we to do with this information?

This idea of focusing on estimation rather than rejecting a null hypothesis (as is the function of a p-value) helps us to make better informed decisions. It is mildly helpful to know for example that increasing the field of view by some amount likely increases presence, but it is much more helpful to know that increasing the field of view by X amount leads to a Y increase in presence with Z level of confidence in this estimate. Obtaining this estimate of confidence can then more fruitfully guide cost-benefit analyses.

### 7.3.2 Statistical power

This idea of the level of confidence we have in our estimates relates to the concept of statistical power. Power is our ability to detect or estimate an effect. Power is primarily determined by the size of your sample and the size of the effect. To illustrate, imagine that you are searching for an object in the dark. The brighter your flashlight, the easier time you will have finding the object. Similarly, the larger the object, the easier it will be to find. You can think of the brightness of your flashlight as representing your sample size, and the size of the object as representing the effect size. You can confidently find and make statements about an object in the dark when the object (sample) is large and you have a bright flashlight (large sample), but will be less confident with other combinations of smaller objects and weaker flashlights. Finding a very small object might require a very bright flashlight. Alternatively, when the object is very large, you may not need to invest in such a bright flashlight. We often have little control over the size of the effect we are looking for, but the size of our sample is more within our control. Brighter flashlights are more resource intensive, but they are often worth having the power to see what is truly going on rather than relegating ourselves to fumbling around in the dark, hazarding guesses based on weak evidence.

In practice, this translates to a recommendation of doing what is called a "power analysis" to make sure your flashlight is bright enough before you venture out into the dark. A given effect size with a given desired level of precision will require a certain minimum number of participants to obtain that desired level of precision, and as long as you have an idea of that effect size and desired level of precision before running the experiment, you can calculate

this minimum number of required participants and set yourself up for success. Free online software like G*Power (Faul et al, 2007) makes this process straightforward.

The more challenging part of the process often comes in estimating the effect size beforehand in order to perform the power analysis. You may not have a great guess about the effect size you are looking for. Fortunately, there are a handful of viable solutions to address this problem.

One common solution is to look for previous studies that tested similar types of treatments and use their effect size as your estimate. While convenient, note that any deviation between the previous experiment's protocol and your own could alter the effect. Another solution is to run a pilot experiment, a condensed version of the experiment with a smaller number of participants, in order to get a rough estimate of the effect size. This method is more resource intensive but protects against experimental differences. A third solution is to decide on the smallest effect size or widest amount of precision that would be interesting or useful to you, and use this information to perform your power analysis. You might not be interested in looking for barely noticeable changes to presence ratings, but a more sizable change could be worth trying to get a better grip on. This method is a very practical one but might require some intuition about what size of an effect really makes a worthwhile difference. A fourth solution is to run the experiment in batches of 20 plus participants, checking the confidence interval around the effect size after each batch of 20 plus, until you obtain a level of precision that is satisfactory or decide the experiment is not worth continuing. This method does not require a predetermined estimate of the effect size (avoiding power analysis entirely) and allows for some level of resource mitigation, but should never be used when p-values are the outcome measure. Under a null hypothesis testing approach this method would be considered "p-hacking" and would lead to an increased rate of false positives. Under an estimation approach however this is less of a concern because your goal is to narrow the precision of the estimate rather than to pass below an arbitrary threshold.

The analogy of looking for objects in the dark should make clear that lots of data will be required to answer questions about more subtle effects. Keep in mind that these effects may manifest in important ways across a large consumer base despite their subtlety. Without a sufficient amount of data to complement the size of the effect you are studying, you will remain in the dark and lack the power to determine whether or not a factor is worth changing. We should also note that part of what makes up an effect size is the variability of the outcome measure, with more variable outcomes translating to smaller effects, so naturally having many observations (i.e. a larger sample) can help to overcome variability and give greater precision despite high variance. Power analysis is a tool that can help you appropriately size your sample.

Of course we realize there are deadlines and a quick rough answer is sometimes preferred to a long precise one, but there are many situations where investing the resources up front will net larger gains down the line. If the result of running a small sample is that you are still unable to accurately measure the effect, you are better off running a larger sample or no sample at all. It is a waste of your resources to fumble around in the dark mushroom hunting with a tea candle. Worse still, you might end up misidentifying the mushroom.

# 8 References

Abeeluck, A. K., Iverson, A., Goetz, H., & Passon, E. (2018). High-Performance Displays for Wearable and HUD Applications. SID Symposium Digest of Technical Papers, 49(1), 768–771. https://doi.org/10.1002/sdtp.12358

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003). Head tracking latency in virtual environments: Psychophysics and a model. PsycEXTRA Dataset, 2083–2087. https://doi.org/10.1037/e576872012-001

Alfano, P. L., & Michel, G. F. (1990). Restricting the Field of View: Perceptual and Performance Effects. Perceptual and Motor Skills, 70(1), 35–45. https://doi.org/10.2466/pms.1990.70.1.35

Arthur, K. W. (2000). Effects of field of view on performance with head-mounted displays. University of North Carolina.

Beams, R., Kim, A. S., & Badano, A. (2019). Transverse chromatic aberration in virtual reality head-mounted displays. Optics Express, 27(18), 24877. https://doi.org/10.1364/oe.27.024877

Bierings, R. A. J. M., Overkempe, T., van Berkel, C. M., Kuiper, M., & Jansonius, N. M. (2019). Spatial contrast sensitivity from star- to sunlight in healthy subjects and patients with glaucoma. Vision Research, 158(March 2018), 31–39. https://doi.org/10.1016/j.visres.2019.01.011

Blake, R., Westendorf, D., & Fox, R. (1990). Temporal perturbations of binocular rivalry. Perception & Psychophysics, 48(6), 593–602. https://doi.org/10.3758/bf03211605

Bockisch, C. J., & Miller, J. M. (1999). Different motor systems use similar damped extraretinal eye position information. Vision Research, 39(5), 1025–1038.

Botella, C., Riva, G., Gaggioli, A., Wiederhold, B.K., Alcañiz, M., & Baños, RM. (2012). The present and future of positive technologies. Cyberpsychol Behav Soc Netw, 15(2), 78–84. http://doi:10.1089/cyber.2011.0140

Boucher, L., Groh, J. M., & Hughes, H. C. (2001). Afferent delays and the mislocalization of perisaccadic stimuli. Vision Research, 41(20), 2631–2644.

Boult, T., & Wolberg, G. (1992). Correcting chromatic aberrations using image warping. Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.1992.223201

Brunnström, K., Beker, S.A., de Moor, K., Dooms, A., Egger, S., et al.. (2013) Qualinet white paper on definitions of quality of experience.

Brunnström, K., Sjöström, M., Imran, M., Pettersson, M., & Johanson, M. (2018). Quality of Experience for a Virtual Reality simulator. Electronic Imaging, 2018(14), 1-9. https://doi.org/10.2352/issn.2470-1173.2018.14.hvei-526

Burt, P., & Julesz, B. (1980). A disparity gradient limit for stereo fusion. Science, 208, 615–617.

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. Perception, 9(6), 671-682.

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. International Journal of Human–Computer Interaction, 36(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Champel, M.-L., Doré, R., & Mollet, N. (2017). Key Factors for a High-Quality VR Experience. 2017 SPIE Optical Engineering and Applications, 10396. https://doi.org/10.1117/12.2274336

Cholewiak, S. A., Love, G. D., Srinivasan, P. P., Ng, R., & Banks, M. S. (2017). Chromablur: Rendering chromatic eye aberration improves accommodation and realism. ACM Transactions on Graphics, 36(6), 1–12. https://doi.org/10.1145/3130800.3130815

Cohen, J. (1994). The earth is round (p < .05). American Psychologist, 49(12), 997–1003. https://doi.org/10.1037/0003-066x.49.12.997

Csikszentmihalyi, M. (1990). Flow: The Psychology Of Optimal Experience. Harper & Row.

de Valois, R. L., Morgan, H., & Ma Snodderly, D. (1974). Psychophysical studies of monkey vision - III. Spatial luminance contrast sensitivity tests of macaque and human observers. Vision Research, 14, 75–81.

Dassonville, P., Schlag, J., & Schlag-Rey, M. (1995). The use of egocentric and exocentric location cues in saccadic programming. Vision Research, 35(15), 2191–2199.

Diner, D. B., & Fender, D. H. (1987). Hysteresis in human binocular fusion: Temporalward and nasalward ranges. Journal of the Optical Society of America A, 4(9), 1814–1819.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In A. J. Woods, J. O. Merritt, S. A. Benton, & M. T. Bolas (Eds.), Stereoscopic Displays and Virtual Reality Systems XI, Proc.SPIE Vol.5291 (pp. 36–46). IS&T/SPIE.

Dolezal, H. (1982) Living in a world transformed: Perceptual and performatory adaptation to visual distortion.

Ellis, S. R., Mania, K., Adelstein, B. D., & Hill, M. I. (2004). Generalizability of latency detection in a variety of virtual environments. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2632–2636. https://doi.org/10.1037/e577162012-006

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Ferry, E.S. (1892). Persistence in vision. American Journal of Science, 44, 192-207.

Fender, D., & Julesz, B. (1967). Extension of Panum's fusional area in binocularly stabilized vision. Journal of the Optical Society of America, 57(6), 819-830.

Fox, R. (1991). Binocular rivalry. D.M. Regan (Ed.), Binocular vision and psychophysics, MacMillan Press, New York.

Glaser, J. S., Savino, P. J., Sumers, K. D., McDonald, S. A., & Knighton, R. W. (1977). The photostress recovery test in the clinical assessment of visual function. American Journal of Ophthalmology, 83(2), 255–260. https://doi.org/10.1016/0002-9394(77)90624-9

Greivenkamp, J.E. (2004). Field Guide to Geometrical Optics, SPIE Press, Bellingham, WA, USA. (See: https://spie.org/publications/fg01_p27_field_of_view)

Hampton, D. R., & Kertesz, A. E. (1983). The extent of Panum's area and the human cortical magnification factor. Perception, 12(2), 161-165.

Hofmeyer, F., Fremerey, S., Cohrs, T., & Raake, A. (2019). Impacts of internal HMD playback processing on subjective quality perception. Electronic Imaging, 2019(12). https://doi.org/10.2352/issn.2470-1173.2019.12.hvei-219

Hyson, M. T., Julesz, B., & Fender, D. H. (1983). Eye movements and neural remapping during fusion of misaligned random-dot stereo- grams. Journal of the Optical Society of America, 73(12), 1665–1673.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. Chance, 18(4), 40–47. https://doi.org/10.1080/09332480.2005.10722754

Jansen, S. E. M., Toet, A., & Werkhoven, P. J. (2011). Obstacle crossing with lower visual field restriction: Shifts in strategy. Journal of Motor Behavior, 43(1), 55–62.

Järvenpää, T., & Pölönen, M. (2009). Advances in Near-To-Eye Display Optical Characterization. SID Symposium Digest of Technical Papers, 40(1), 507. https://doi.org/10.1889/1.3256827

Järvenpää, T., Salmimaa, M., & Levola, T. (2010). 23.4: Qualified viewing spaces for near-to-eye and autostereoscopic displays. 48th Annual SID Symposium, Seminar, and Exhibition 2010, Display Week 2010, 1, 335–338. https://doi.org/10.1889/1.3500449

Järvenpää, T., & Pölönen, M. (2010). Optical characterization and ergonomical factors of near-to-eye displays. Journal of the Society for Information Display, 18(4), 285. https://doi.org/10.1889/jsid18.4.285

Järvenpää, T., & Salmimaa, M. (2016). Optical measurements of different near-eye display types. Proceedings of the Society for Information Display 2016 International Symposium, Seminar and Exhibition, 1056–1059.

Johnson, P. V., Kim, J., & Banks, M. S. (2014). The visibility of color breakup and a means to reduce it. Journal of Vision, 14(14), 10–10. https://doi.org/10.1167/14.14.10

Kalloniatis, M., & Luu, C. (2007). Temporal Resolution, Webvision: The Organization of the Retina and Visual System. http://webvision.med.utah.edu

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology, 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Klymenko, V., Verona, R. W., Beasley, H. H., & Martin, J. S. (1994). Convergent and divergent viewing affect luning, visual thresholds, and field-of-view fragmentation in partial binocular overlap

helmet-mounted displays. Helmet-and Head-Mounted Displays and Symbology Design Requirements, Proc.SPIE, 2218, 82–96.

Klymenko, V., Harding, T. H., Beasley, H. H., & Martin, J. S. (2000). Investigation of helmet-mounted display configuration influences on target acquisition. In R. J. Lewandiowski, L. A. Haworth, & H. J. Girolamo (Eds.), Helmet- and Head-Mounted Displays V, Proc.SPIE Vol.4021 (Issues 316–334). SPIE.

Kuhl, S. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). HMD calibration and its effects on distance judgments. Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization - APGV '08. https://doi.org/10.1145/1394281.1394284

Kulikowski, J. J. (1978). Limit of single vision in stereopsis depends on contour sharpness. Nature, 275(5676), 126–127.

Kytö, M., Hakala, J., Oittinen, P., & Häkkinen, J. (2012). Effect of camera separation on the viewing experience of stereoscopic photographs. Journal of Electronic Imaging, 21(1), 011011. https://doi.org/10.1117/1.jei.21.1.011011

Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. Journal of Imaging Science and Technology, 53(3), 030201. https://doi.org/10.2352/j.imagingsci.technol.2009.53.3.030201

Leibowitz, H.W. & Owens, D.A. (1975) Anomalous myopia and the intermediate dark focus of accommodation. Science 189, 646-648.

Lin, J. J.-W., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. Proceedings IEEE Virtual Reality 2002, 164–171. https://doi.org/10.1109/VR.2002.996519

Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. Journal of Vision, 8(11), 19–19. https://doi.org/10.1167/8.11.19

Masaoka, K., Nishida, Y., Sugawara, M., Nakasu, E., & Nojiri, Y. (2013). Sensation of realness from high-resolution images of real objects. IEEE Transactions on Broadcasting, 59(1), 72–83. https://doi.org/10.1109/TBC.2012.2232491

Mcintire, J. P., Havig, P., Harrington, L. K., Wright, S. T., Watamaniuk, S. N. J., & Heft, E. (2018). Microstereopsis is good, but orthostereopsis is better: precision alignment task performance and viewer discomfort with a stereoscopic 3D display. Three-Dimensional Imaging, Visualization, and Display 2018. https://doi.org/10.1117/12.2297378

Mcmahon, T. T., Irving, E. L., & Lee, C. (2012). Accuracy and Repeatability of Self-Measurement of Interpupillary Distance. Optometry and Vision Science, 89(6), 901–907. https://doi.org/10.1097/opx.0b013e318257f37b

Melzer, J. (1998). Overcoming the field of view : Resolution invariant in head mounted displays. In R. J. Lewandowski, L. A. Haworth, & H. J. Girolamo (Eds.), Helmet- and Head-Mounted Displays III, Proceedings of SPIE Vol. 3362. SPIE - The International Society for Optical Engineering. https://doi.org/10.1117/12.317441

Metzger, B. A., & Beck, D. M. (2020). Probing the mechanisms of probe-mediated binocular rivalry. Vision Research, 173, 21–28. https://doi.org/10.1016/j.visres.2020.04.011

Nolan, A., Delshad, R., & Sedgwick, H. A. (2012). Compression of Perceived Depth as a Function of Viewing Conditions. Optometry and Vision Science, 89(12), 1757–1767.

Park, D., Kim, Y. J., & Park, Y. K. (2019). Hyperrealism in full ultra high-definition 8K display. Digest of Technical Papers - SID International Symposium, 50(Book 2), 1138–1141. https://doi.org/10.1002/sdtp.13130

Patney, A., Kim, J., Salvi, M., Kaplanyan, A., Wyman, C., Benty, N., … Luebke, D. (2016). Perceptually-based foveated virtual reality. ACM SIGGRAPH 2016 Emerging Technologies on - SIGGRAPH '16. https://doi.org/10.1145/2929464.2929472

Patterson, R., Winterbottom, M., Pierce, B., & Fox, R. (2007). Binocular rivalry and head-worn displays. Human Factors, 49(6), 1083–1096. https://doi.org/10.1518/001872007X249947

Peck, T. C., Sockol, L. E., & Hancock, S. M. (2020). Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research. IEEE Transactions on Visualization and Computer Graphics, 26(5), 1945–1954. https://doi.org/10.1109/tvcg.2020.2973498

Piantanida, T. P. (1986). Stereo hysteresis revisited. Vision Research, 26(3), 431-437.

Piantanida, T. P., Boman, D., Larimer, J., Gille, J., & Reed., C. (1996). Studies of the Field-Of-View/Resolution Tradeoff in Virtual-Reality Systems. Proceedings of the SPIE - The International Society for Optical Engineering, 1666(1992), 448–456.

Psotka, J., Lewis, S. A., & King, D. (1998). Effects of Field of View on Judgments of Self-Location: Distortions in Distance Estimations Even When the Image Geometry Exactly Fits the Field of View. Presence: Teleoperators & Virtual Environments, 7(4), 352–369.

Qin, Z., Chou, P.-Y., Wu, J.-Y., Huang, C.-T., & Huang, Y.-P. (2019). Resolution-enhanced light field displays by recombining subpixels across elemental images. Optics Letters, 44(10), 2438. https://doi.org/10.1364/ol.44.002438

Roberts, J. E., & Wilkins, A. J. (2013). Flicker can be perceived during saccades at frequencies in excess of 1 kHz. Lighting Research and Technology, 45(1), 124–132.

Rovamo J. & Virsu, V. (1979) An estimation and application of the human cortical magnification factor. Experimental Brain Research 37, 495-510.

Schor, C., Wood, I., & Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. Vision Research, 24(7), 661–665.

Sheard, C. (1934). The Prescription Of Prisms. Optometry and Vision Science, 11(10), 364–378. https://doi.org/10.1097/00006324-193410000-00001

Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, *11*(8), 11–11.

Shimojo, S., & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry. Vision Research, 30(1), 69–80. https://doi.org/10.1016/0042-6989(90)90128-8

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. PsycEXTRA Dataset. https://doi.org/10.1037/e519702015-014

Son, J.-Y., Lee, H., Kim, J., Lee, B.-R., Son, W.-H., & Venkel, T. (2019). A HMD for users with any interocular distance. Proceedings of the International Display Workshops, 995-998. https://doi.org/10.36463/idw.2019.3d8_3dsa8-2

Stringham, J. M., Fuld, K., & Wenzel, A. J. (2003). Action spectrum for photophobia. Journal of the Optical Society of America A, 20(10), 1852. https://doi.org/10.1364/josaa.20.001852

Toet, A., Jansen, S. E. M., & Delleman, N. J. (2008a). Effects of field-of-view restriction on maneuvering in a 3-D environment. Ergonomics, 51(3), 385–394.

Toet, A., van der Hoeven, M., Kahrimanovic, M., & Delleman, N. J. (2008b). Effects of field of view on human locomotion. Head- and Helmet-Mounted Displays XIII: Design and Applications, SPIE-6955, 69550H–1–69550H–11.

Tyler, C. W., & Hamer, R. D. (1990). Analysis of visual modulation sensitivity IV Validity of the Ferry–Porter law. Journal of the Optical Society of America A, 7(4), 743. https://doi.org/10.1364/josaa.7.000743

Vos, J. J. (2003). Reflections on glare. Lighting Research and Technology, 35, 163–175.

Wang, B., & Ciuffreda, K. J. (2006). Depth-of-focus of the human eye: theory and clinical implications. Survey of Ophthalmology, 51(1), 75–85.

Wang, P., Mohammad, N., & Menon, R. (2016). Chromatic-aberration-corrected diffractive lenses for ultra-broadband focusing. Scientific Reports, 6(1). https://doi.org/10.1038/srep21545

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. Frontiers in Psychology, 10(FEB), 1–19. https://doi.org/10.3389/fpsyg.2019.00158

Watanabe, J., Noritake, A., Maeda, T., Tachi, S., & Nishida, S. (2005). Perisaccadic perception of continuous flickers. Vision Research, 45(4), 413–430.

Wells, M. J., Venturino, M., & Osgood, R. K. (1989). The effect of field-of-view size on performance at a simple simulated air-to-air mission. In Helmet-Mounted Displays, Proc. SPIE (Vol. 1116, pp. 126–137).

Yamanoue, H., Okui, M., & Okano, F. (2006). Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. IEEE Transactions on Circuits and Systems for Video Technology, 16(6), 744–752. https://doi.org/10.1109/tcsvt.2006.875213

Yang, M., Zhang, J., & Yu, L. (2019). Perceptual Tolerance to Motion-To-Photon Latency with Head Movement in Virtual Reality. 2019 Picture Coding Symposium (PCS). https://doi.org/10.1109/pcs48520.2019.8954518

Yao, R., Heath, T., Davies, A., Forsyth, T., Mitchell, N., & Hoberman, P. (2014). Oculus VR Best Practices Guide. Oculus VR.

Zhan, T., Zou, J., Xiong, J., Liu, X., Chen, H., Yang, J., … Wu, S. T. (2019). Practical Chromatic Aberration Correction in Virtual Reality Displays Enabled by Cost-Effective Ultra-Broadband Liquid

Crystal Polymer Lenses. Advanced Optical Materials, 8(2), 1901360. https://doi.org/10.1002/adom.201901360