Identification of visual functional thresholds for immersion assessment in virtual reality

Hébert-Lavoie, Maxence[1], Doyon-Poulin, Philippe[1], and Ozell, Benoît[2]

[1]Département de mathématiques et génie industriel, Polytechnique Montréal
[2]Département de génie informatique et génie logiciel, Polytechnique Montréal

maxence.hebert-lavoie@polymtl.ca

Abstract

We consider that to objectively measure immersion, one needs to assess how each sensory quality is reproduced in a virtual environment. In this perspective, we introduce the concept of functional threshold which corresponds to the value at which a sensory quality can be degraded without being noticed by the user of a virtual environment. We suggest that the perceived realism of a virtual experience can potentially be evoked for sensory qualities values ranging from the perceptual threshold to the functional threshold. Thus, the identification of functional thresholds values allows us to constrain immersion. To lay the foundation for the identification of functional thresholds, we applied a modified version of the method of limits. We measured the value at which 30 participants were able to identify the degradation of their field of view (FOV), visual acuity, and contrast sensitivity while executing a multi-directional selection test. This enabled us to identify functional perceptual thresholds of 96.6 degrees for FOV, 12.2 arcmin for visual acuity, and 25.6% for contrast sensitivity.

1 Introduction

Immersive virtual reality simulates human sensory and perceptual processes through mechanical and computational means. Its use suggests multiple applications, offering opportunities to transform actual training and design methods (Oberhauser and Dreyer, 2017; Safi et al., 2019). To be valid and reliable, such usage requires a virtual environment to be equivalent to its real-world counterpart. Considered a measure of the realism of a virtual experience (Slater, 2009; Slater et al., 2010; Ommerli, 2020; Milleville-Pennel and Charron, 2015; Deniaud et al., 2014; Skarbez et al., 2017), the concept of presence could be used to evaluate such correspondence (Milleville-Pennel and Charron, 2015).

Presence is a quale, it is subjective and its interpretation is variable for each individual (Slater, 2009). It is composed of two principal aspects; coherence and immersion (Skarbez et al., 2020). Coherence refers to the ability of the environment to act as the user expects (Skarbez et al., 2017). Immersion refers to the extensiveness of the simulation of sensory modalities by a system (Slater, 2009; Skarbez et al., 2017).

Currently, the assessment of presence is done mostly by indirect methods such as subjective questionnaires (Witmer and Singer, 1998; Usoh et al., 2000; Milleville-Pennel and Charron, 2015), but also with physiological and behavioral measures (Riley et al., 2004; Milleville-Pennel and Charron, 2015). These methods are criticized for their lack of sensitivity, diagnosticity, and their validity as a measure of the construct (Gardner and Martin, 2007; Slater, 2009). The subjective character of presence renders its direct measurement complex (Riley et al., 2004; Slater, 2009). However, because of its objective nature (Slater, 2009), evaluating immersion by quantifying each sensory modality could represent a way to circumvent a part of this complexity.

To do so we propose an approach similar to Perroud et al. (2019b), but in a more holistic way; to assess how all sensory modalities are reproduced in a virtual environment. This requires measuring the sensory qualities of each sensory modality, i.e. the distinctive attributes that enable a sensory modality to carry out its function (Wolfe et al., 2006). In that regard, we've identified the sensory qualities of vision. These are visual accommodation, eye movements, stereopsis, perception of movement, color perception, FOV, visual acuity, and light and contrast sensitivity. This paper intends to lay the foundation for the determination of a range of sensory quality values, bounded by perceptual and functional thresholds, in which immersion can occur.

1.1 Thresholds of perception

Perceptual thresholds are the limiting sensory quality values for which humans can perceive, e.g. for visual acuity, it may be the smallest object that can be discriminated by the human eye. Within the framework of immersion, perceptual thresholds can be considered as the values to be reached by a system to maximize presence. However, at the moment, creating an immersive system able to reach these values represents a technical challenge that has yet to be solved.

Still, concepts such as useful FOV (UFOV) (Wolfe et al., 2017) imply that humans do not always perform at the limit of their perceptual capacities. This makes it relevant to evaluate to what extent a sensory quality can be degraded without being noticeable, i.e. identifying functional thresholds. By using functional thresholds, it would be possible to constrain immersion functionally, between a maximal bound, defined by the perceptual thresholds, and a minimal bound, defined by the functional thresholds.

Functional thresholds have not been established for any sensory qualities. Therefore, the objective of this study is to determine three functional thresholds related to vision. In that regard, we measured the value for which a degradation was perceived for FOV, visual acuity, and contrast sensitivity in a virtual environment while executing a task where the attention is not entirely devoted to the discrimination of these phenomena. This will allow us to propose a framework for the study of immersion.

Thus, we first report on studies linked to the FOV, visual acuity, and light and contrast sensitivity. Then, we detail the methodology applied to evaluate their respective functional threshold in an immersive virtual reality. Thereafter, we present the measured functional thresholds, discuss their meaning in the context of the study and their foreseeable implications.

2 Related work

This section presents the state of knowledge related to the three visual qualities that are targeted by this study. For each quality, we describe the notions that are useful to understand their role toward the visual system, determine their perceptual threshold, and highlight relevant information for the identification of functional thresholds.

2.1 Field of view

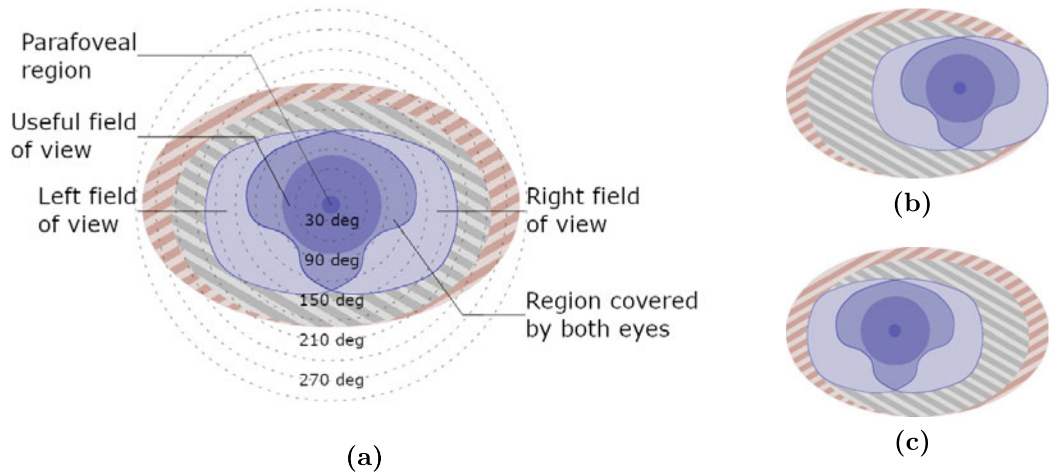
The visual experience is a reconstruction of the world based on two distinct images from the retina of each eye. The portion of space that the eyes can capture when motionless is called the field of view (Wolfe et al., 2006).

Figure 1 shows a representation of different sections of the FOV. Binocularly, when the eyes are fixed on a point, the human visual system covers a horizontal FOV of approximately 180 to 200 degrees (Perroud et al., 2019b; Blissing, 2020) of which 100 to 120 degrees are covered by both eyes (Wolfe et al., 2006; Blissing, 2020; Riecke et al., 2006), and a vertical FOV of 120 to 140 degrees (Elbamby et al., 2018; Wolfe et al., 2006; Riecke et al., 2006). In head-mounted displays (HMD), it is important to consider the fact that the eyes can move in the presented visual field. When the head is fixed, the eyes, by moving, can cover an area of 290 degrees horizontally and 190 degrees vertically (Riecke et al., 2006). These values could serve as a perceptual threshold for FOV. Also, it is important to note that it is possible to sustain the eye gaze in only 40 to 50 degrees in the central region of the FOV which is called the fixed foveated region (Kress, 2019).

In comparison, the horizontal FOV of most current generation HMDs is between 90 and

Figure 1

Schematic representation of different sections of the FOV.



Note. in (a) the blank blue sections represent the portion of space that is captured by the eyes when fixed on a point, the stripped regions indicate the parts of the FOV that can be covered by moving the eyes; (b) shows the eyes looking at the right limit of the FOV; (c) shows the eyes looking at the left limit of the fixed foveated region.

110 degrees. We know that such FOV size makes any task where the observer is detecting objects in peripheral vision difficult to perform (Blissing, 2020). Research that evaluated the FOV size in a virtual environment tends to demonstrate that higher presence is evoked by values closer to the total FOV size (Lin et al., 2002; Slater et al., 2010; Duh et al., 2002). This relation is also reflected in performance measure (Ragan et al., 2015). However, since the portion of the FOV used by an observer depends on the task performed (Kemeny, 1999) these relations do not allow us to pinpoint a FOV size that would be sufficient for general use.

Visual performance is not equal across the entire FOV. Increasing eccentricity from the fovea affects motion perception (Blissing, 2020; Lidestam et al., 2019; Hoffman et al., 2018), visual acuity (Ogawa and Shidoji, 2011; Siekawa et al., 2019; Thibos et al., 1996), light and contrast sensitivity (Thibos et al., 1996; Patney et al., 2016), and color perception (Patney et al., 2016; Hoffman et al., 2018; Duchowski et al., 2009; Hansen et al., 2009). These differences in visual qualities performance across the FOV may be responsible for the fact that different functions are attributed to different portions of the visual field. For instance, the UFOV describes the function of the central part of the visual field. It is defined as the surface around the point of fixation inside which information can be perceived and processed during a visual task (Langlois, 2013). This definition reveals a similarity with our conception of a functional threshold of FOV. The UFOV size varies depending on the executed task (Ringer et al., 2014). Its range is often estimated to cover a diameter of 30 to 40 degrees from the central gaze point (Wolfe et al., 2017). Yet, methods assessing the extent of the UFOV tend to neglect the contribution of peripheral information that can be acquired (Wolfe et al., 2017), and it could cover a diameter of up to 80 degrees (Itoh et al., 2009). Thus, the state of knowledge on UFOV does not enable us to clearly establish a functional threshold of FOV.

In summary, the data available does not allow us to establish a functional threshold of FOV. For tasks that necessitate peripheral vision information acquisition, the FOV size offered by the current generation of HMD is theoretically too small and makes it harder to perform the task. However, according to the definition of UFOV, we could have expected such size to be sufficient. Hence, more research is needed to establish a functional threshold.

2.2 Visual acuity

Visual acuity refers to the finest level of detail that can be detected or identified (Bennett et al., 2019). It is physiologically determined by the spacing between photoreceptors on the retina (Wolfe et al., 2006). In a virtual environment, the notion of resolution acuity is generally used when measuring visual acuity. It is defined as the smallest angular separation between two neighboring objects that can be detected (Wolfe et al., 2006). Visual acuity depends on the environment's visual conditions. Reaching the perceptual threshold of visual acuity requires a stable image displayed within 0,5 degrees of the center of the fovea, maximum contrast, and photopic conditions (Iskander et al., 2018). Under these conditions, the human eye has a resolution acuity between 0,5 and 2 arcmin with an average of 1 arcmin. (Kemeny, 1999; Perroud et al., 2019b; Kim et al., 2019; Livingston, 2006; Elbamby et al., 2018).

The actual state of knowledge does not enable us to pinpoint a functional threshold for visual acuity. We know that the angular resolution of most current HMD is 4 to 6 arcmin. This is due to a pixel density too low to reproduce maximal visual acuity (Maxwell et al., 2018). Theoretically, this resolution implies a lack of information that makes objects difficult to discern at distances where they would be clearly identified in real life (Kemeny, 1999; Blissing, 2020). The relation between resolution and perceived realism of a virtual environment is reflected only to some extent in the measure of presence (Felton and Jackson, 2021). An increased resolution was shown to significantly increase the reported presence (Duh et al., 2002; Hvass et al., 2017). However, these results were obtained with a great difference between the high and low-resolution conditions, and it appears that a smaller difference between conditions does not yield significant results (Dinh et al., 1999). This might indicate that a functional threshold can be reached with the actual technology. The inability to see a difference between two resolution conditions could be due to the fact that humans see most of the world through their peripheral vision that cannot render images as sharply as the foveal region (Lee, 2004). Visual acuity is not constant throughout the retina and decreases with increasing eccentricity. For a fixed stimulus, the resolution acuity at 8 degrees of eccentricity of the fovea is 2.8 arcmin, dropping at 7.0 arcmin at 13 degrees (Ogawa and Shidoji, 2011). At an eccentricity of 25 degrees, it is 12 arcmin (Siekawa et al., 2019), and at 30 degrees a spacing of up to 15 arcmin cannot be discriminated (Hoffman et al., 2018).

Briefly, we were unable to establish a functional threshold for visual acuity. However, previous research indicates that current HMD technology resolution acuity is sufficient to reach a visual acuity functional threshold. We would expect such value to correspond to the visual acuity in peripheral vision.

2.3 Light and contrast sensitivity

Light sensitivity corresponds to the number of photons absorbed by the cones and rods lining the retina. Due to neural encoding, the human eye is sensitive to differences in the intensity of regions of the receptive field rather than the average intensity (Wolfe et al., 2006). Light sensitivity is a time-varying phenomenon that depends on previous lighting conditions (Yan et al., 2018). To take into account the time-varying factor and address this visual quality in a way that enables the identification of a context-independent perceptual threshold, the concept of contrast sensitivity is employed. Thus, we will use the terms "contrast sensitivity" when referring to this phenomenon onwards.

Contrast sensitivity is equivalent to the smallest difference in luminance required to identify a target (Bennett et al., 2019). It is usually expressed as the ratio between the low and high luminance of an image. Contrast sensitivity depends on spatial frequency (Bennett et al., 2019), illumination conditions (Barten, 1989), and FOV size (Perroud et al., 2019a). There is a robust contrast threshold of 1% which is independent of the visual conditions (Pelli and Bex, 2013). Under optimal conditions, however, the minimum contrast that can be perceived in the fovea, the perceptual threshold, is 0.2 % (Thibos et al., 1996). For an HMD with a FOV of 110 degrees and a resolution equivalent to 4.6 arcmin this value increases to 0.45% (Barten, 1989).

Contrast sensitivity in the periphery is decreased compared to the foveal contrast sensitivity (Thibos et al., 1996). For an aliased stimulus, it is reported to range from 19% to 23% at 30 degrees of eccentricity (Thibos et al., 1996).

Yet, we are unable to identify a functional threshold. We know that the contrast sensitivity of observers using an HMD is poorer than in a real environment (Sproule et al., 2019). Due to a limited range of displayed luminance (Ledda et al., 2004), chromatic aberrations, and insufficient resolution, it has been concluded that VR technology is not able to reproduce standard contrast sensitivity testing (Vivas-Mateos et al., 2020). Still, we can reasonably expect the contrast range offered in current HMDs to be sufficient to let its user feel present in a virtual environment.

Also, we were unable to identify studies that targeted the effect of varying levels of contrast on presence or performance in a virtual environment. One study evaluated the effect of contrast on the perceived realism of desktop computer images, identifying a threshold for negative contrast offset of $0.6678(V(x, y) - 0.5) + 0.5$, where $V(x, y)$ is the value channel of the HSV color model, i.e. the pixel luminance (Dolhasz et al., 2016). However, since their results were reported as a contrast offset, which is image-dependent, it can't be used as a direct indication to establish a functional threshold.

Again, the body of knowledge was insufficient to identify a functional threshold value for contrast sensitivity. As for visual acuity, we would expect to find a functional threshold of contrast sensitivity that corresponds to values in the peripheral region since humans see most of the world through this area (Lee, 2004).

2.4 Research objective

Following this review of related work, we can see that previous studies focused on establishing the function of the different visual qualities, and identifying perceptual thresholds. We also understand that no functional thresholds have been established for the three visual qualities targeted by this study. For each, some information is available but it is not sufficient to pinpoint them. Hence, the objective of this paper is to measure the functional thresholds for FOV, visual acuity, and contrast sensitivity in a virtual environment with an HMD.

3 Method

The methodology outlined below aims to identify thresholds at which a user of a virtual environment can perceive a degradation of its FOV size, visual acuity, and contrast sensitivity. The value at which a degradation is perceived is considered a functional threshold of the degraded quality for this specific environment. Generally, the methodology consisted of separately degrading these three visual qualities while the participant was performing a multi-directional selection test. Such a task was used so the participant's attention was not directed exclusively toward the identification of the degradation of the visual quality. The experiment was carried out in virtual reality using a Unity implementation of a multi-directional selection test.

3.1 Participants

The participants recruited were volunteers over 18 years of age, had no uncorrected vision problems, and did not have a known prevalence of being subject to cybersickness. The research project was reviewed by Polytechnique Montréal's research ethics board and met the current research ethics standards on ethical conduct for research involving humans. 30 participants (13 females, 17 males) aged between 22 and 63 years old ($M = 30.33$) took part in the experiment. All participants executed the experiment with their dominant hand, 26 participants were right-handed and 4 left-handed. During the experiment, 11 participants wore glasses, 3 contact lenses and 16 did not need any device to correct their vision. Participants estimated their total experience time with virtual reality HMDs and had a median experience of 1 hour ($min = 0$, $M = 5.8$, $max = 100$).

3.2 Materials

3.2.1 HTC Vive Pro Eye

The participants used the HTC Vive Pro Eye HMD to complete the task. The HTC Vive Pro Eye is a virtual reality headset that allows the immersion of a participant in a virtual environment. It has a dual OLED display with a diagonal of 3.5 inches allowing a 110 degrees FOV, has a refresh rate of 90 Hz, and a resolution of 1440 x 1600 pixels per eye, for a maximum resolution acuity of 4.6 arcmin according to Maxwell et al. (2018). The headset generates a luminance between 0.04 and 130 cd/m^2 and occupies the AdobeRGB colorspace (Clausen et al., 2019). It is equipped with an audio system that was removed for the experiment. The controllers used in the experiment are those supplied with the device. The experiment's computer used an NVIDIA GeForce RTX 2080 SUPER GPU and an Intel Core i9-9900K CPU with 64GB RAM.

3.2.2 Virtual Environment

Figure 2

The virtual environment from three different points of view.

(a) *Participant's initial view* **(b)** *West and North wall* **(c)** *East and South wall*

The virtual environment was created with Unity version 2019.3. As shown in figure 2, it consisted of a 6m x 4m x 6m room with a column on the northern wall. The floor had a concrete texture, the northern and western walls had a white paint texture, and the eastern and southern walls, as well as the ceiling, had a newly installed gypsum texture. A network of dark grey pipes extending to the northern wall and rectangular light fixtures covered the ceiling. On the northern side were plants and boxes of various colors. In front of the western wall were an industrial sink, a watering can, and yellow metal lockers. The eastern wall had a dark gray door and the southern wall behind the participant was empty. The purpose of presenting many objects in the virtual environment was to diversify the range of colors, contrast, and level of visual detail across the participant's FOV.

3.3 Procedure

3.3.1 Multi-directional selection test

As part of the experiment, participants performed several multi-directional selection tests, also referred to as Fitts' test. The objective of using a multi-directional selection test was to direct participants' attention toward specific objects and ensure that they were not only attentive to the degradation of their visual qualities.

A multi-directional selection test is based on Fitts' law which models human performance, reflecting the speed and accuracy of a selection task (MacKenzie, 2013). The overall performance of the test is captured by the throughput index which is measured in bits per second (Soukoreff and MacKenzie, 2004). A higher throughput indicates a better efficiency to the task Soukoreff and MacKenzie (2004).

We based our experimental task on the Fitts' law evaluation procedure, standardized for a 2D performance evaluation (International Organization for Standardization, 2012). More precisely, the task consisted of sequentially selecting 15 targets of size W arranged in a circle with a distance D between two selections. This was repeated for nine different combinations of target size and distance, each defined by a difficulty index (ID). The ID ranged from 0.85 to 3.17, which is in line with Soukoreff and MacKenzie (2004) recommendations. All tasks were performed with targets located 0.5 m in front of the participant. The ID included three W of 0.05, 0.075, and 0.100 m, equivalent to visual angles of 343.5 arcmin, 514.7 arcmin, and 685.3 arcmin respectively. The D between targets were 0.080, 0.240, and 0.400 m, corresponding to visual angles of 548.9 arcmin, 1619.5 arcmin, and 2616.2 arcmin. Figure 3 shows different task IDs as presented in the virtual environment.

At the beginning of a trial, the 15 targets were displayed, disappearing after their respective selection. The target to select was opaque and yellow to distinguish it from the others which were transparent and navy blue. The first illuminated target was at the right

Figure 3

Representation of different multi-directional selection test IDs.

(a) **(b)** **(c)** **(d)**

Note. (a) shows an ID of 1.05 corresponding to a W of 0.075 m and a D of 0.08 m; (b) shows an ID of 2.07 corresponding to a W of 0.075 m and a D of 0.24 m; (c) shows an ID of 2.32 corresponding to a W of 0.1 m and a D of 0.4 m; and, (d) shows an ID of 2.54 corresponding to a W of 0.05 m and a D of 0.24 m.

side of the circle. After it was selected, the next target, located at about 180 degrees on the opposite side of the circle, became yellow and opaque. Selections then continued clockwise in a back and forth motion until all targets were selected. The first selected target was not included in the data analysis to minimize the impact of task initiation.

To make a selection, the participant had to place the tip of the controller in the target and pull the trigger. To help with the controller positioning, a red sphere was located at the tip of the virtual representation of the HTC Vive controller. Feedback confirming that the participant had made a selection was also provided, the sphere would turn green when the trigger was pulled.

It took a participant between 1 minute and 1 minute 30 seconds to complete a single multi-directional selection test.

3.3.2 Functional threshold's psychophysical evaluation

We evaluated the functional thresholds using an adaptation of the method of limits (Herrick, 1969) that incorporated a multi-directional selection test used to direct the observer's attention. The participants had to detect a visual degradation of the virtual environment.

The method of limits is based on the fact that there is a greater likelihood that a stimulus will be consciously detected if its intensity is greater (Herrick, 1969). It is a judgment task where participants are presented with a stimulus of incrementally varied intensity and need to tell if they can detect it or not.

Typically, the method of limits consists of two procedures: the ascending limit method and the descending limit method. The ascending limit method consists of increasing the intensity of a stimulus until it is detected by the participant. The descending limit method consists of decreasing the intensity of a stimulus until it is no longer detected by the participant. Experimenters then compute the combination of thresholds from both methods executed a multitude of times by many participants resulting in a distribution of perceptual thresholds (Herrick, 1969).

The main modification we brought to the method was to only increment the stimulus intensity in one direction; degrading the visual qualities. This way, the detection probability increases with each increment because the degradation becomes more obvious, even if the participants are engaged in a task. We believe that incrementing the stimulus intensity in the other way would not have yielded appropriate results. In a normal use case, by doing a conscious effort, observers using the HTC Vive Pro Eye set at its maximal perceptual values can detect that their visual qualities are degraded. This is because the technology is not yet able to reproduce the perceptual thresholds. As a consequence, since it is expected to be difficult for participants to judge when the image has reached the maximal perceptual value that the HMD can offer, we anticipated that most participants would not have been apt to tell that they weren't able to detect the degradation of their visual qualities.

Thus, the size of the FOV, the screen resolution, and the contrast offset value were all separately decreased during the experiment. The participants had to indicate when they were able to perceive a degradation of either their FOV, visual acuity, or contrast sensitivity and, if possible, identify which quality was degraded, without disrupting the multi-directional selection test being performed.

We calibrated our degradation rates and final visual quality values in pre-test sessions. Each degradation was calibrated so that an increment was imperceptible. This avoided underestimating the thresholds since a variation with a large disparity could have erroneously lead to earlier detection. We calibrated our final values by degrading the visual qualities until they reached a value judged detectable at a 100% rate. We made sure that this value was reached after around 2 minutes, giving each participant enough time to complete the multi-directional selection test while the degradation occurred.

Also, to identify functional thresholds, representing lower bound values of perception, we made sure that participants were naive to the degradation to avoid an overestimation of the threshold. In that regard, each visual quality degradation was only carried out once per participant. We also made sure that they were not aware of which visual quality would be degraded or if a degradation would happen.

The following sections present the method that we used to achieve the degradation of the three visual qualities targeted by this study. Note that the visual quality values were measured according to the system values and then converted to generally interpretable values, e.g. the visual acuity was measured in resolution and then expressed in arcmin to ensure their communicability.

3.3.3 Field of view

We reduced the 110-degree FOV offered by the HTC Vive Pro Eye with the *postprocessing* package from Unity by using the vignette function. Figure 4 shows an example of such degradation.

Measurement and degradation rate As the vignette function does not allow for a directly convertible value of the FOV in degrees, we measured it after the experiment. To do so, observers had to look at a point in the center of the screen and move their extended right arm holding the controller, away from the center toward the right side of their FOV.

Figure 4

Representation of FOV degradation.

(a) (b) (c)

Note. (a) shows the environment without any degradation, at an FOV of 110 degrees; (b) shows the FOV degraded to 102.4 degrees; and, (c) shows the FOV degraded to 50.6 degrees.

At the exact moment when the observers weren't able to see the controller, they pressed its trigger, recording the angle between the controller and the FOV's center. The total FOV corresponds to the recorded angle between the border of the vignette at the recorded participant's functional threshold value and the center of its FOV multiplied by 2. For each functional threshold, this measure was repeated three times. The results presented in this paper consist of the mean of those three measurements.

In the HMD, the vignette wasn't visible until it reached -0.7 arbitry unit (AU) after 19 seconds. Thus, the FOV varied according to the following equations :

$$FOV_{deg} = \begin{cases} 110, & \text{if } 0 < t < 19 \\ 126.76 - 0.88t, & \text{otherwise} \end{cases} \tag{1}$$

Where FOV_{deg} is the size of the field of view in degrees and t is the elapsed time since the beginning of the degradation in seconds.

3.3.4 Visual acuity

To reduce the visual acuity, we used the *SteamVR* package to virtually reduce the resolution of the presented image by a parameterizable factor. This resulted in the

Figure 5

Representation of visual acuity degradation.

(a) (b) (c)

Note. (a) shows the environment without any degradation, at a visual acuity of 4.6 arcmin; (b) shows the visual acuity degraded to 15.5 arcmin; and, (c) shows the visual acuity degraded to 50.7 arcmin.

presentation to the observer of what appears to be larger and larger pixels. Figure 5 shows an example of such degradation.

Measurement and degradation rate When the resolution parameter is set to 1, the system offers the maximum possible resolution with the equipment, in this case, 2880 x 1600 or a resolution acuity of 4.6 arcmin. The visual acuity varied according to the following equation :

$$Res_{deg} = 2880 - 37.43t \tag{2}$$

Where Res_{deg} is the horizontal resolution of the screen and t is the elapsed time since the beginning of the degradation in seconds. The vertical resolution was reduced by the same factor.

To convert the screen resolution in arcmin, we used the following equation, which is based on Maxwell et al. (2018)

$$VA_{arcmin} = 120 \arctan \frac{2Lens \tan \frac{V}{2}}{Res_{deg} Width} \tag{3}$$

Where VA_{arcmin} is the visual acuity in arcmin, and Res_{deg} is the horizontal pixel resolution, with $Lens$, the horizontal lens' size, here equal to 59.47 mm, V, the horizontal visual angle, equal to 110 degrees, and $Width$ the horizontal screen' size, equal to 88.9 mm.

3.3.5 Contrast sensitivity

Figure 6

Representation of contrast degradation.

(a) **(b)** **(c)**

Note. (a) shows the environment without any degradation, at a contrast of 31.2%; (b) shows the contrast degraded to 23.5 %; and, (c) shows the contrast degraded to 14.7 %.

We achieved contrast degradation by reducing the extreme luminance levels presented in the image. This was done with the *postprocessing* package from Unity that allows the contrast of the pixels presented to the user to be offset, varying between 100 AU and -100 AU. For this experiment, the initial contrast value was set to 0 and could be degraded up to -100 AU. At -100 AU the pixels all have the same dark gray color. Figure 6 shows an example of such degradation.

Measurement and degradation rate The contrast varied according to the following equation:

$$C_{deg} = -0.8t \tag{4}$$

Where C_{deg} is the contrast and t is the elapsed time since the beginning of the degradation in seconds.

The recommended measurement for natural contrast stimuli is RMS contrast (Pelli and Bex, 2013). It is expressed as the ratio of the standard deviation and the mean luminance of the image. For this environment, the contrast presented as a Unity parameter value can be converted in RMS contrast using the following equation.

$$C_{RMS} = 0.2882 + 0.0011(19.29 + C_{deg}) + 3.786 \times 10^{-7}(19.29 + C_{deg})^3 \tag{5}$$

This equation was obtained with a polynomial regression of order 3 with an $R^2 = 0.9983$. We calculated the RMS contrast of images of the virtual environment corresponding to each recorded functional contrast sensitivity threshold value to compute the regression.

3.4 Experimental procedure

One participant at a time took part in the experiment. After reviewing and signing the information and consent form, the participants sat in a chair positioned so that they could move their arms to the limits of their kinesthetic sphere without hitting surrounding objects. The experimenter then explained to the participants that they would have to perform four multi-directional selection tests.

To limit the possible bias linked to the participant's expectation to see a degradation, they were told that they would not necessarily see the degradation of each visual quality. They were given examples such as: they might see two times the degradation of contrast and two times no degradation or that they might see three times the degradation of FOV and once the degradation of visual acuity.

Participants were asked to put on the HMD and were given a controller before being asked to follow the HTC Vive Pro Eye eye-tracking camera calibration procedure to adjust their interpupillary distance. It should be noted that at this point, all visual qualities were set at their maximum value. The participants then performed a practice run of the multi-directional selection test for which no data was recorded. They were reminded that they had to do the task continuously, without stopping, until they had selected all targets. Before each multi-directional selection test, the participants were reminded that they had to focus on the multi-directional selection test, executing it as quickly and accurately as possible and that there may or may not be any degradation. They were also prompted to say out loud when they perceived a degradation if they perceived any.

The participants then executed four multi-directional selection tests. The moment when a participant identified a degradation was recorded. The participants had to select all targets for the task to be completed. In the case in which a participant did not notice the degradation, the visual quality value recorded was the one at the end of the last target selection. Once a multi-directional selection test was completed, the environment was returned to the initial visual condition.

3.5 Experimental design

Each participant performed a total of four multi-directional selection tests, one in which no degradation occurred and three in which either the FOV size, visual acuity, or contrast sensitivity was degraded. To control for a possible bias due to the presentation order the conditions were balanced according to a Latin square experimental design.

The dependent variables were the value for which an observer was able to perceive a degradation of its visual quality, i.e. the functional thresholds. Each experiment yielded three measures of functional threshold, one for each visual quality.

During a multi-directional selection test, the presentation of each ID was randomized between trials to control for bias due to participant expectations. Throughout the experiment, the participants made a total of 540 selections (15 targets x 9 IDs x 4 degradation conditions).

3.6 Data analysis

A psychometric function was computed for each degraded visual quality. This function provides a statistical estimate of the functional threshold distribution from which we estimated the point of subjective equality (PSE) and the just noticeable difference (JND).

The PSE is defined as the degradation value that will be detected by 50% of the observers. It is considered as a degradation level for which the observer will judge the environment with an equal likelihood to be degraded or not (Mania et al., 2004). The JND is the additional amount of degradation needed to increase an observer's detection rate from 50% to 75%. It is considered to reflect the observer's sensitivity to a degradation difference (Ellis et al., 2004).

Each psychometric function was modeled as a 2-parameter cumulative Weibull distribution. Their parameters were estimated with the *fitdist* function from the *fitdistrplus* library in R using the maximum likelihood estimator method. The 95% confidence intervals (CI) of the distribution were estimated with the *confint* function from the same library using likelihood profiling.

Between-subject analyses were performed to take into account the potential effect of age, gender, dominant hand, visual correction devices (glasses and contact lenses), experience in virtual reality, and order of presentation on the measured functional threshold. We compared the mean degradation value of FOV, visual acuity, and contrast sensitivity for the previously enumerated variables. For gender (female or male), dominant hand (left or right), correction devices (wearing or not wearing), and experience in VR (less than 2 hours or more than 2 hours), we performed an independent-samples t-test when the data met the assumptions for a parametric test and a Mann-Withney U test otherwise. For age (less than 30 y.o., between 30 and 40, or more than 40) and order (first, second, third, or fourth) we performed a one-way analysis of variance (ANOVA) when the data met the assumptions for a parametric test, and Kruskal-Wallis test otherwise.

Within-subject analyses were performed to compare the results of the multi-directional selection test and verify the potential effect of the measured functional threshold on the task. We compared the mean throughput for order (first, second, third, or fourth) and visual degradation condition (FOV, visual acuity, contrast sensitivity, or no degradation), performing a one-way repeated measures ANOVA when the data met the assumptions for a parametric test, and a Friedman test otherwise.

Between and within-subject analyses, significance was evaluated with an $\alpha = 0.05$.

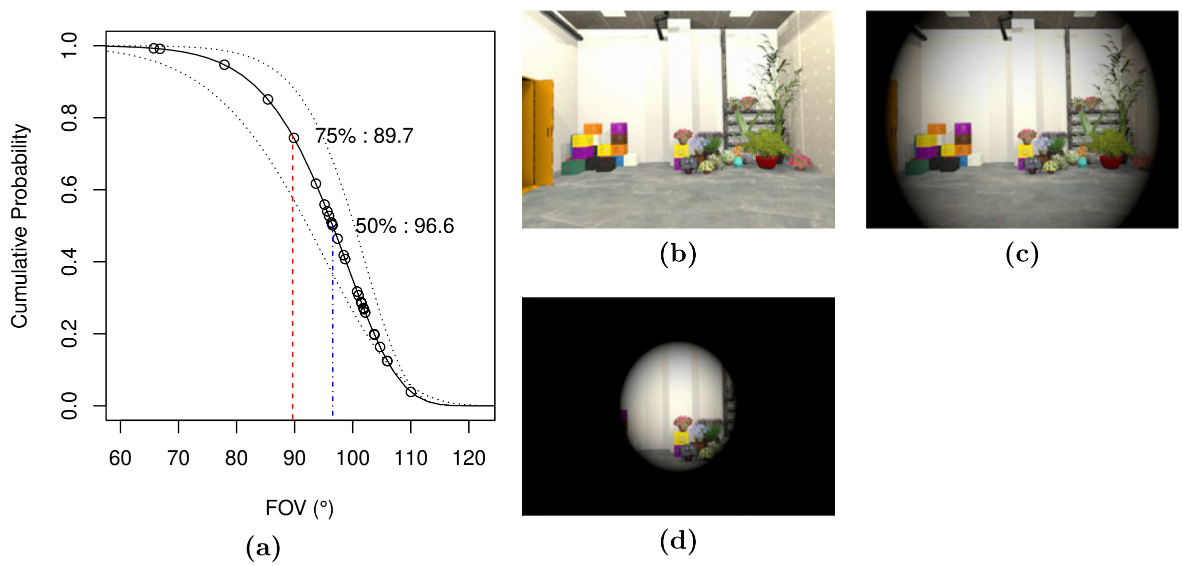
4 Results

This section first presents the results for the three measured functional thresholds. For each, it reports the PSE, JND, the scale and shape parameters of the estimated Weibull distribution along with the parameters of the confidence intervals. Then, the significant results of the between-subject analyses and within-subject analyses are reported.

4.1 Field of view

Figure 7

Results of FOV degradation.



Note.(a) Psychometric function showing the probability of a participant to be able to detect the degradation of its FOV in a virtual environment; (b) maximum recorded value; (c) PSE; (d) minimum recorded value.

The PSE for FOV is 96.6 degrees with a 95% CI [92.3, 100.3] and the JND is 6.9 degrees, as illustrated in figure 7a. The computed cumulative Weibull distribution has a shape parameter $\beta = 11.86$ and a scale parameter $\eta = 99.59$. The lower bound of the 95% confidence interval has a shape $\beta = 8.15$ and a scale $\eta = 96.50$, and the higher bound has a shape $\beta = 15.58$ and a scale $\eta = 102.69$. Figures 7b to 7d show respectively a representation of what was seen by an observer's left eye at the maximum value allowed by the system, the PSE, and the minimal value recorded during the experiment.

A Kruskal-Wallis test revealed a significant effect of age on the detection of FOV degradation $\chi^2(2) = 7.624$, $p < 0.05$. Post hoc pairwise comparisons using a Mann-Withney U test: $U = 66$, $n1 = 23$, $n2 = 3$, $p < 0.05$, showed that there was a significant difference between participants aged younger than 30 years old ($M = 97.34$ degrees, $SD = 8.98$ degrees) and older than 40 years old ($M = 67.59$ degrees, $SD = 17.43$ degrees).

Also, a Mann-Withney U test revealed that wearing a visual correction device had a significant effect on the detection of FOV degradation:

$U = 2.037, n1 = 14, n2 = 16, p < 0.05$, participants wearing visual correction device ($M = 89.57$ degrees, $SD = 16.76$ degrees) detected the FOV degradation at a lower threshold than the participants not wearing them ($M = 99.56$ degrees, $SD = 7.01$ degrees). Further analysis did not reveal any significant differences between glasses and contact lenses wearer.

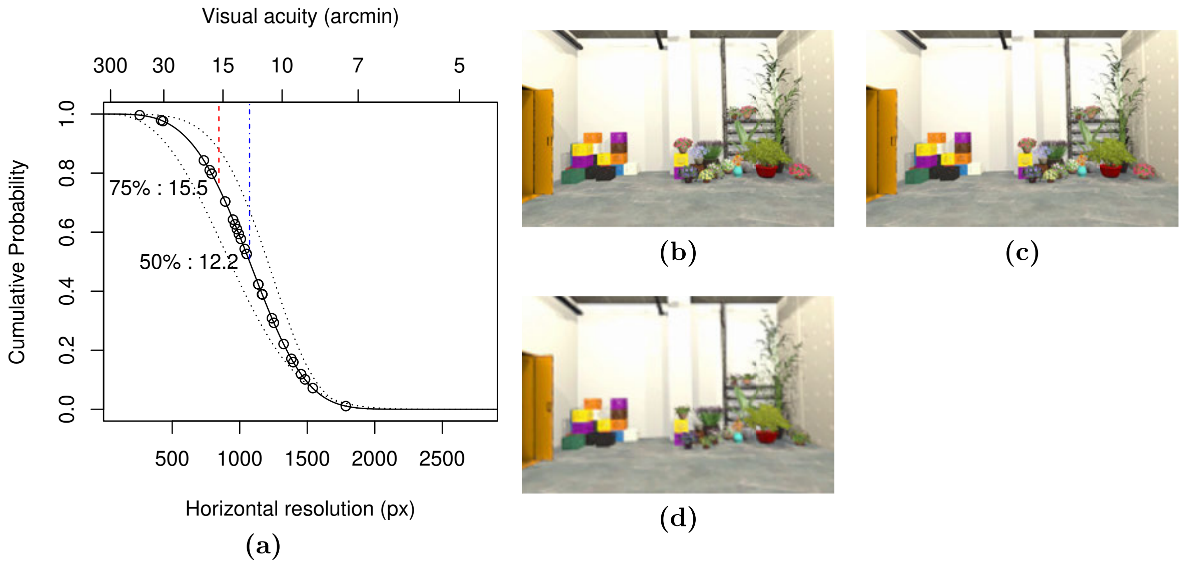
A Kruskal-Wallis test revealed no significant difference of the identified functional FOV threshold while comparing the order of execution of the multi-directional selection test $\chi^2(3) = 1.6376, p > 0.10$.

4.2 Visual acuity

The PSE for visual acuity is 12.2 arcmin with a 95% CI [10.9, 14.2] and the JND is 3.3 arcmin as illustrated in figure 8a. The computed cumulative Weibull distribution has a shape parameter $\beta = 3.69$ and a scale parameter $\eta = 1185.04$. The lower bound of the 95% confidence interval has a shape $\beta = 2.64$ and a scale $\eta = 1064.96$, and the higher bound has a shape $\beta = 4.75$ and a scale $\eta = 1305.12$. Figure 8b to 8d show a representation of what was seen by an observer's left eye at the maximum value allowed by the system, the PSE,

Figure 8

Results of visual acuity degradation.



Note. (a) Psychometric function showing the probability of a participant to be able to detect the degradation of its visual acuity in a virtual environment; (b) maximum recorded value; (c) PSE; (d) minimum recorded value.

and the minimal value recorded during the experiment. We reported no significant between-subject analyses for this visual quality.

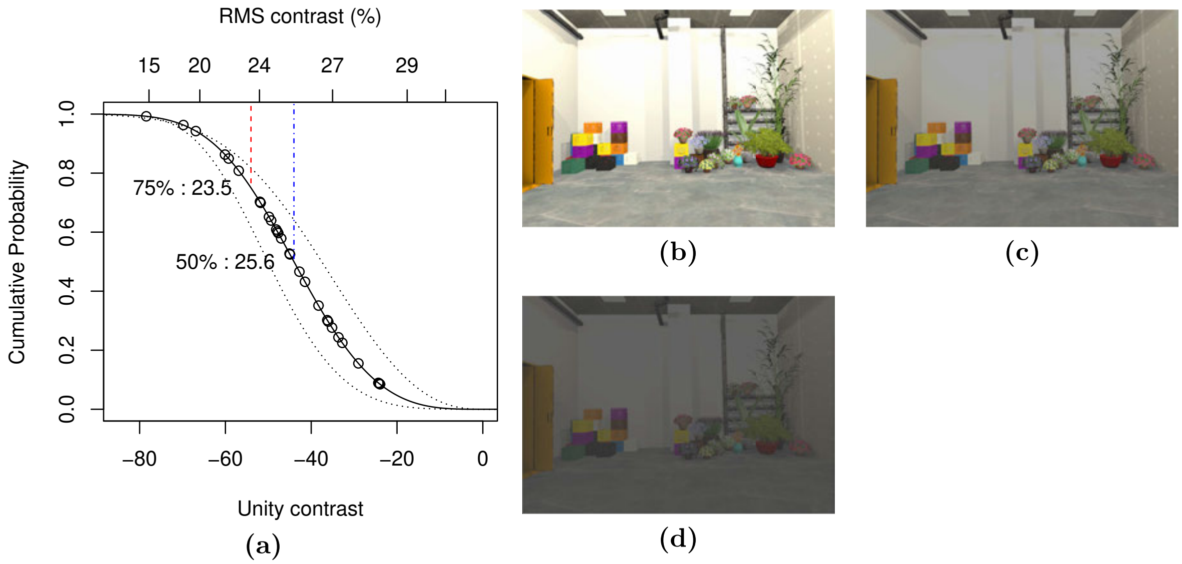
Moreover, a one-way ANOVA revealed no significant difference of the identified functional visual acuity threshold while comparing the order of execution of the multi-directional selection test $F(3, 26) = 0.359$, $p > 0.10$.

4.3 Contrast sensitivity

The PSE for contrast sensitivity is 25.6% with a 95% CI [24.4, 26.6] and the JND is 2.1% as illustrated in figure 9a. The computed cumulative Weibull distribution has a shape parameter $\beta = 3.38$ and a scale parameter $\eta = 49.07$. The lower bound of the 95% confidence interval has a shape $\beta = 2.46$ and a scale $\eta = 43.57$, and the higher bound has a shape $\beta = 4.30$ and a scale $\eta = 54.56$. Figure 9b to 9d show a representation of what was seen by an observer's left eye for the maximum value allowed by the system, the PSE, and

Figure 9

Results of visual acuity degradation.



Note. (a) Psychometric function showing the probability of a participant to be able to detect the degradation of its contrast sensitivity in a virtual environment; (b) maximum recorded value; (c) PSE; (d) minimum recorded value.

the minimal value recorded during the experiment.

A two sample t-test showed that there was a significant effect of gender on the detection of contrast degradation $t(28) = -2.075$, $p < 0.05$, women (M = 23,7%, SD = 3,9%) detected the contrast degradation at a lower threshold than men (M = 26,1%, SD = 1,8%).

A one-way ANOVA revealed no significant difference of the identified functional contrast sensitivity threshold while comparing the order of execution of the multi-directional selection test $F(3, 26) = 0.2118$, $p > 0.10$.

4.4 Multi-directional selection test throughput analysis

A one-way repeated measures ANOVA showed a significant order effect on the throughput $F(3, 87) = 6.698$, $p < 0.001$. A post hoc comparison with 6 paired sample t-tests corrected with Holm-Bonferroni method to counteract multiple comparisons problem, was used to identify the differences between the conditions. As shown in table 1, the task executed first

Table 1

Summary of post hoc pairwise comparison of task execution order

Comparison Group	Mean Difference	Sig.	Effect Size	DF
First and Second	-0.27	< 0.05	-2.768	29
First and Third	-0.41	< 0.001	-4.384	29
First and Fourth	-0.33	< 0.05	-2.828	29

had a significantly lower throughput (M = 4.68 bps, SD = 0.68 bps) than the tasks performed second (M = 4.96 bps, SD = 0.51 bps), third (M = 5.09 bps, SD = 0.64 bps), or fourth (M = 5.01 bps, SD = 0.62 bps). Significant differences were only found between the first and the other order rank of execution of the task. Other comparisons revealed no significant differences.

Table 2

Summary of post hoc pairwise comparison of throughput between visual degradation conditions

Comparison Group	Mean Difference	Sig.	Effect Size	DF
FOV size and no degradation	-0.16	> 0.1	-1.563	29
Visual acuity and no degradation	0.09	> 0.1	0.9771	29
Contrast sensitivity and no degradation	-0.23	> 0.1	-2.114	29
Visual acuity and FOV size	0.25	< 0.05	2.825	29
Visual acuity and contrast sensitivity	0.32	< 0.01	3.373	29
FOV size and contrast sensitivity	0.07	> 0.1	0.612	29

Also, a one-way repeated measures ANOVA showed a significant effect of the degradation of visual qualities on the throughput $F(3, 87) = 4.2739$, $p < 0.01$. However, a post hoc comparison with 6 paired sample t-tests corrected with the Holm-Bonferroni method showed no difference between a degraded visual quality's throughput and the reference condition's throughput with no degradation. As shown in table 2 significant differences in throughput were only found between the visual acuity degradation and, respectively, the FOV size degradation, and the contrast sensitivity degradation. The visual acuity degradation condition had a significantly higher throughput (M = 5.10 bps, SD = 0.56 bps) than the FOV size condition (M = 4.85 bps, SD = 0.61 bps) and the contrast sensitivity condition (M = 4.78 bps, SD = 0.71 bps).

5 Discussion

The goal of this experiment was to measure the functional thresholds of FOV, visual acuity, and contrast sensitivity. These thresholds represent the values below which virtual environment's users can perceive a degradation of their visual qualities while their attention is not directed toward the discrimination of the degradation. In this section we analyze the results obtained following the experiment, first focusing on each targeted visual quality, and then in a more general manner.

5.1 Functional thresholds

To determine the functional thresholds we used the PSE measured using a 2-parameter cumulative Weibull function based on the experimental data collected from 30 participants for each visual quality degradation. Since functional thresholds represent a lower bound of perception, we chose this conservative value to assess them.

In this section, we provide a critical analysis of the results obtained. For each targeted visual quality, we examine the relation between the measured functional threshold to the information identified in the related work and we discuss our study limitations.

5.1.1 Field of view

For FOV, we identified a functional threshold of 96.6 degrees. It represents a degradation of 13.4 degrees from the initial setting of 110 degrees. This functional threshold is smaller than the FOV covered by both eyes when an observer looks straight ahead (Blissing, 2020; Riecke et al., 2006).

This result seems to corroborate the statement that methods assessing the UFOV neglect the contribution of peripheral information (Wolfe et al., 2017). Using the common UFOV value of 40 degrees and adding 50 degrees to include the fixed foveated region, the region where an observer can sustain its eye gaze, we would expect a maximal value of 90 degrees for the FOV in which information can be perceived and processed. Yet, our measured FOV functional threshold is larger and seems to correspond to the UFOV of 80 degrees from the adapted UFOV of Itoh et al. (2009). When taking into account the fixed foveated region this value spans up to 130 degrees.

We suggest that the difference between the value of 130 degrees and our measured functional threshold of 96.6 degrees is because the portion of the FOV that is used by an observer depends on the task performed (Kemeny, 1999). During the multi-directional selection test, the maximum portion of the FOV covered by the targets is estimated at 55 degrees which is encompassed in the measured functional threshold value. We consider that the measured value discussed in this section should be interpreted as the FOV functional threshold when the elements related to the task can all be included in the visual field of the observer. However, our experiment does not allow us to issue recommendations on the relation of the size of an object in the FOV and the FOV functional threshold. Further experiments should be carried out to determine to what extent this functional threshold can be applied.

Some factors could have biased our functional threshold assessment of FOV. These factors essentially consist of a glitch in the virtual environment and the effect of age and visual correction devices on the identified degradation value.

In figure 7a we can observe that the confidence interval is relatively large and becomes narrower around a FOV size of 100 degrees. This may be due to a glitch that occurred when the vignette parameter was dynamically reduced, causing a lag in the update of the pixels near the border of the vignette. This glitch was only apparent when the participant's head moved. One participant described it as what one sees when its head emerges from the water while wearing swimming goggles. In total, six participants reported seeing this glitch, but it wasn't clear if it was before or after perceiving the degradation, and most of them mentioned that they reported the degradation because they saw the tunneling vision effect. Therefore, the functional FOV threshold measured may be overestimated and a lower threshold could be expected for the same task.

Also, our analysis revealed a significant effect of age and visual correction devices on the measured FOV functional threshold.

Participants older than 40 years old ($n = 3$) identified a significantly lower FOV functional threshold than participants aged younger than 30 years old ($n = 23$). Since our population of participants older than 40 years old is relatively low, we assume that more participants would be necessary to make a reliable conclusion on the effect of age.

Finally, participants wearing visual correction devices identified lower FOV functional thresholds. We hypothesize that this effect may be due to the glasses' frame that could have veiled the degradation, yielding a possible underestimation of the measured functional threshold value of FOV.

5.1.2 Visual acuity

For visual acuity, we identified a functional threshold of 12.2 arcmin. The range of the 95% CI [10.9-14.2] corresponds to the visual acuity at 20 to 30 degrees of eccentricity from the fovea (Siekawa et al., 2019; Thibos et al., 1996). This visual acuity functional threshold seems to reflect the fact that most of the perception is done through the peripheral vision field (Lee, 2004).

At the value of 12.2 arcmin, the smallest object relevant to the multi-directional selection test was still clearly visible, as the smallest target width was 343.5 arcmin. The degradation of visual acuity mostly made the object in the background difficult to discern. Most participants measured the extent of this degradation only at the end of the task. At that point, they were able to take the time to look at the environment in general and not only focus on the targets.

5.1.3 Contrast sensitivity

For contrast sensitivity, we identified a functional threshold of 25.6% RMS contrast. The identified RMS contrast was 5.6% less than the initial situation of 31.2 %. This functional threshold is close to the contrast sensitivity threshold of 19% to 23% at 30 degrees of eccentricity from the fovea (Thibos et al., 1996). It is also slightly higher than the threshold of 23.3% that we found when applying Dolhasz et al. (2016) contrast offset threshold equation to our context. The difference may be the corollary of the poorer contrast sensitivity of observers using an HMD (Sproule et al., 2019). Yet, the value we measured seems to also indicate that perception is done through the peripheral vision field (Lee, 2004).

Three participants were not able to identify the contrast degradation during the experiment. However, they reported that the degradation was obvious either at the moment when they finished the task, or just after the environment was set back to its original state. We suggest two explanations for this phenomenon. First, we assume that they might have perceived the degradation and weren't just confident enough in their judgment to report it at this moment. Otherwise, this inability to identify the degradation may be due to the degradation rate established for the experiment, which was too slow to reach a value sufficiently low to be detected in the allotted time.

We found a significant effect of gender on the detection of contrast degradation. In our experiment, women ($n = 13$) were less sensitive to contrast degradation than men ($n = 17$). According to our debriefing interviews, we believe that our male population had more video gaming experience. Hence, they were more likely to have already practiced tasks similar to the one executed during the experiment, making them better at shifting their attention between the multi-directional selection test and the contrast discrimination.

5.2 Multi-directional selection test

In this section, we discuss the impact of the multi-directional selection test on our experiment. We examine the practice effect identified with a one-way repeated measures ANOVA, then we analyze the effect of the task on the degradation detection.

A one-way repeated measures ANOVA revealed an order effect on the throughput that can be attributed to the participants' lack of practice. The first multi-directional selection test performed by the participants had a significantly lower throughput than the three others. As Soukoreff and MacKenzie (2004) recommend, we asked our participants to do a practice test that was not recorded to familiarize themselves with the multi-directional selection test. They had the option to do as many practices run as they needed to feel comfortable with the task and every participant chose to do only one. As reflected by this analysis, participants should have completed two practice tests to eliminate the practice effect on the throughput. However, the throughput was not the variable of interest in this study.

To verify that this order effect did not impact the identification of the degradation, we performed a between-subject analysis of every visual quality degradation value. These tests revealed no significant difference when comparing the order of execution of a degradation. Thus, we cannot assume that the order in which the visual qualities were degraded affected the detection of the degradation

Also, to determine the effect of the multi-directional selection test on the degradation condition, we compared the throughput of each degradation condition. A one-way repeated measures ANOVA revealed no difference in performance between a task with degradation and without degradation. We suggest that this indicates that the multi-directional selection test was well suited for the identification of a functional threshold.

Surprisingly, we found that the throughput was significantly higher for the visual acuity condition than the FOV and the contrast sensitivity conditions. When looking at the mean difference in table 2, we see that the visual acuity is the only degradation to have a positive mean difference with the no degradation condition. We would have expected the degradation to affect all the throughput scores in the same way. Since the visual environment was degraded, we expected that all throughput scores would be less than the no degradation condition.

We suggest that the loss in visual acuity may have enhanced the precision of the participants in the multi-directional selection test. As the boundary between the target and the background became less clear, the participant had to make selections more toward the center to keep up with the instructions related to the multi-directional selection test, i.e. to select targets as quickly and as accurately as possible. We think that the degradation of FOV size and contrast sensitivity did not allow for such a shift in the participant's behavior. Thus, the significant differences between these conditions could be explained by the fact that the visual degradation affected the participant just enough to have a significantly different behavior between visual acuity degradation and FOV size or contrast sensitivity, but not enough to affect significantly these conditions when compared to the no degradation condition.

5.3 Implication of functional threshold

The functional thresholds identified in this paper could serve as constraining values for the assessment of immersion in a virtual environment. For a virtual experience to be considered realist, it must be able to simulate each visual quality with values comprised in the range bounded by the perceptual threshold as the upper limit and the functional threshold as the lower limit. The values for the visual qualities targeted by this study are included in table 3.

Yet, these functional thresholds need to be used carefully. We assume that functional thresholds are task-dependent. Different task parameters could have yielded different thresholds, e.g. small characters in a reading task would have necessitated a greater visual acuity. Thus, we argue that when such task parameters are known, they should constitute the functional threshold for the said task.

Table 3

Perceptual and functional thresholds of FOV, visual acuity and contrast sensitivity

Visual quality	Measure	Perc. threshold	Funct. threshold
FOV	FOV size	290 degrees horizontally, 190 degrees vertically (Riecke et al., 2006)	96.6 degrees vertically and horizontally
Visual acuity	Resolution acuity	1 arcmin (Livingston, 2006)	12.2 arcmin
Contrast sensitivity	RMS Contrast	0.2% (Thibos et al., 1996)	25.6%

Note. The FOV sizes presented in table 3 take into account the eye movements.

However, we believe that using a multi-directional selection test to measure the visual qualities allowed us to estimate a limit that can be used in diverse situations since no particular sensitivity to the tested visual qualities was required by the task. The FOV size encompassed the entire task elements and the targets that needed to be selected were large enough and clearly distinct from other objects in the scene to be seen in any degradation condition. Therefore, we suggest that the measured values constitute general indications that can be used when the perception of a task is not limited by known values. To measure the extent of this assessment, the results obtained in this paper should be evaluated with other tasks that require no particular sensitivity to the tested visual qualities.
Otherwise, as mentioned in the related work section, visual qualities are not independent of each other. For example, a low-resolution value combined with a low contrast value might yield a lower functional threshold. However, we think that by using a conservative approach, i.e. assessing the functional thresholds with the identified PSE, we were able to present valid values of the concept of functional thresholds. Still, more experiments are needed to confirm the robustness of the information presented in this paper.

6 Conclusion

In this paper, we introduced the notion of functional threshold which constitutes a limit below which perception should be considered functionally degraded. Thus, below these values, a virtual environment seems unlikely to be appropriate to carry out equivalent real-world tasks.

We measured the functional thresholds for three visual qualities. We've assessed that for real-world tasks executed in a virtual environment a FOV of 96.6 degrees, a visual acuity of 12.2 arcmin, and a contrast of 25.6% constitute limiting values.

We support the idea that values farther from the perceptual thresholds than the functional thresholds will surely be considered as degraded by an observer, while value closer might not.

These results can be used as constraints to establish an objective measure of immersion for the emergence of presence in a virtual environment. In a future study, we will use them as a reference to evaluate the impact of the extensiveness of visual qualities simulation on presence in a virtual environment. Moreover, they will be used to make recommendations on the technological requirements for a virtual environment. For this purpose, the functional threshold reported in this paper shall be considered as minimal values to be used so observers are unable to detect with certitude that their visual qualities are degraded while in a virtual environment.

7 Acknowledgments

This work was supported by Mitacs, the Consortium for Aerospace Research and Innovation in Canada (CARIC), the Consortium de recherche et d'innovation en aérospatiale au Québec (CRIAQ) and by its partners under GRANT CARIC-CRIAQ MDO-1649-TRL4+. We would also like to express our sincerest gratitude to Antoine Martin for his technical contribution to the software development.

References

Barten, P. G. (1989). The effects of picture size and definition on perceived image quality. *IEEE Transactions on Electron Devices*, 36(9):1865–1869.

Bennett, C. R., Bex, P. J., Bauer, C. M., and Merabet, L. B. (2019). The assessment of visual function and functional vision. In *Seminars in Pediatric Neurology*, volume 31, pages 30–40. WB Saunders.

Blissing, B. (2020). *Driving in virtual reality: Requirements for automotive research and development*, volume 2085. Linköping University Electronic Press.

Clausen, O., Fischer, G., Fuhrmann, A., and Marroquim, R. (2019). Towards predictive virtual prototyping: Color calibration of consumer VR HMDs. In *16th GI AR/VR Workshop*.

Deniaud, C., Mestre, D., Honnet, V., and Jeanne, B. (2014). The concept of "presence" used as a measure for ecological validity in driving simulators. In *Proceedings of the 2014 European Conference on Cognitive Ergonomics*, pages 1–4, Vienna, Austria. ACM Press.

Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., and Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pages 222–228. IEEE.

Dolhasz, A., Williams, I., and Frutos-Pascual, M. (2016). Measuring observer response to object-scene disparity in composites. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 13–18. IEEE.

Duchowski, A. T., Bate, D., Stringfellow, P., Thakur, K., Melloy, B. J., and Gramopadhye, A. K. (2009). On spatiochromatic visual sensitivity and peripheral color LOD management. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–18.

Duh, H. B.-L., Lin, J. J., Kenyon, R. V., Parker, D. E., and Furness, T. A. (2002). Effects of characteristics of image quality in an immersive environment. *Presence: Teleoperators & Virtual Environments*, 11(3):324–332.

Elbamby, M. S., Perfecto, C., Bennis, M., and Doppler, K. (2018). Toward low-latency and ultra-reliable virtual reality. *IEEE Network*, 32(2):78–84.

Ellis, S. R., Mania, K., Adelstein, B. D., and Hill, M. I. (2004). Generalizeability of latency detection in a variety of virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 2632–2636. SAGE Publications Sage CA: Los Angeles, CA.

Felton, W. M. and Jackson, R. E. (2021). Presence: A review. *International Journal of Human–Computer Interaction*, 0(0):1–18.

Gardner, H. J. and Martin, M. A. (2007). Analyzing ordinal scales in studies of virtual environments: Likert or lump it! *Presence: Teleoperators and Virtual Environments*, 16(4):439–446.

Hansen, T., Pracejus, L., and Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of vision*, 9(4):26–26.

Herrick, R. M. (1969). Psychophysical methodology: Comparisons within the method of limits. *Perceptual and Motor Skills*, 28(2):503–514.

Hoffman, D., Meraz, Z., and Turner, E. (2018). Limits of peripheral acuity and implications for VR system design. *Journal of the Society for Information Display*, 26(8):483–495.

Hvass, J., Larsen, O., Vendelbo, K., Nilsson, N., Nordahl, R., and Serafin, S. (2017). Visual realism and presence in a virtual reality game. In *2017 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4. IEEE.

International Organization for Standardization (2012). Ergonomics of human-system interaction — part 411: Evaluation methods for the design of physical input devices. Standard ISO/TS 9241-411:2012, International Organization for Standardization, Geneva, CH.

Iskander, J., Hossny, M., and Nahavandi, S. (2018). A review on ocular biomechanic models for assessing visual fatigue in virtual reality. *IEEE Access*, 6:19345–19361.

Itoh, N., Sagawa, K., and Fukunaga, Y. (2009). Useful visual field at a homogeneous background for old and young subjects. *Gerontechnology*, 8(1):42–51.

Kemeny, A. (1999). Simulation and perception. In *Introduction de la Conférence sur la Simulation de Conduite*.

Kim, J., Jeong, Y., Stengel, M., Akşit, K., Albert, R., Boudaoud, B., Greer, T., Kim, J., Lopes, W., Majercik, Z., et al. (2019). Foveated AR: dynamically-foveated augmented reality display. *ACM Transactions on Graphics*, 38(4):1–15.

Kress, B. C. (2019). Digital optical elements and technologies (EDO19): Applications to AR/VR/MR. In *Digital Optical Technologies 2019*, volume 11062, page 1106222. International Society for Optics and Photonics.

Langlois, S. (2013). ADAS HMI using peripheral vision. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 74–81, New York, NY, USA. Association for Computing Machinery.

Ledda, P., Chalmers, A., and Seetzen, H. (2004). HDR displays: a validation against reality. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 3, pages 2777–2782. IEEE.

Lee, K. M. (2004). Why presence occurs: Evolutionary psychology, media equation, and presence. *Presence: Teleoperators & Virtual Environments*, 13(4):494–505.

Lidestam, B., Eriksson, L., and Eriksson, O. (2019). Speed perception affected by field of view: Energy-based versus rhythm-based processing. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65:227–241.

Lin, J.-W., Duh, H. B.-L., Parker, D. E., Abi-Rached, H., and Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE Virtual Reality 2002*, pages 164–171. IEEE.

Livingston, M. A. (2006). Quantification of visual capabilities using augmented reality displays. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 3–12. IEEE.

MacKenzie, I. S. (2013). *Human-Computer Interaction: An Empirical Research Perspective*. Morgan Kaufmann, San Francisco, CA, USA, first edition.

Mania, K., Adelstein, B. D., Ellis, S. R., and Hill, M. I. (2004). Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pages 39–47.

Maxwell, D., Oster, E., Lynch, S., Maxwell, D., Oster, E., and Lynch, S. (2018). Evaluating the applicability of repurposed entertainment virtual reality devices for military training. *MODSIM World*, (0028):1–10.

Milleville-Pennel, I. and Charron, C. (2015). Driving for real or on a fixed-base simulator: Is it so different? An explorative study. *Presence*, 24(1):74–91.

Oberhauser, M. and Dreyer, D. (2017). A virtual reality flight simulator for human factors engineering. *Cognition, Technology & Work*, 19(2):263–277.

Ogawa, M. and Shidoji, K. (2011). Relationship between perception of image resolution and peripheral visual field in stereoscopic images. In *Stereoscopic Displays and Applications XXII*, volume 7863, page 78630S. International Society for Optics and Photonics.

Ommerli, C. (2020). Examining the effects of perceived telepresence, interactivity, and immersion on pilot situation awareness during a virtual reality flight exercise. Master's thesis, Carleton University.

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics*, 35(6):1–12.

Pelli, D. G. and Bex, P. (2013). Measuring contrast sensitivity. *Vision research*, 90:10–14.

Perroud, B., Regnier, S., Kemeny, A., and Merienne, F. (2019a). Application of the relative visual performance model in a virtual reality immersive system. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3128–3132.

Perroud, B., Régnier, S., Kemeny, A., and Mérienne, F. (2019b). Model of realism score for immersive VR systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, 61:238–251. Special TRF Issue: Driving Simulation.

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., and McMahan, R. P. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7):794–807.

Riecke, B. E., Nusseck, H.-G., and Schulte-Pelkum, J. (2006). Selected technical and perceptual aspects of virtual reality displays. Technical Report 154, Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

Riley, J. M., Kaber, D. B., and Draper, J. V. (2004). Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 14(1):51–67.

Ringer, R. V., Johnson, A. P., Gaspar, J. G., Neider, M. B., Crowell, J., Kramer, A. F., and Loschky, L. C. (2014). Creating a new dynamic measure of the useful field of view using gaze-contingent displays. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 59–66.

Safi, M., Chung, J., and Pradhan, P. (2019). Review of augmented reality in aerospace industry. *Aircraft Engineering and Aerospace Technology*, 91(9):1187–1194.

Siekawa, A., Chwesiuk, M., Mantiuk, R., and Piórkowski, R. (2019). Foveated ray tracing for VR headsets. In *International Conference on Multimedia Modeling*, volume 11295, pages 106–117. Lecture Notes in Computer Science, Springer.

Skarbez, R., Brooks, F., and Whitton, M. (2020). Immersion and coherence: Research agenda and early results. *IEEE Transactions on Visualization and Computer Graphics*.

Skarbez, R., Brooks, Jr., F. P., and Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys*, 50(6):1–39.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557.

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics*, 29(4):1–9.

Soukoreff, R. W. and MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International journal of human-computer studies*, 61(6):751–789.

Sproule, D., Jacinto, R. F., Rundell, S., Williams, J., Perlmutter, S., and Arndt, S. (2019). Characterization of visual acuity and contrast sensitivity using head-mounted displays in a virtual environment: A pilot study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 547–551. SAGE Publications Sage CA: Los Angeles, CA.

Thibos, L. N., Still, D. L., and Bradley, A. (1996). Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision research*, 36(2):249–258.

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence*, 9(5):497–503.

Vivas-Mateos, G., Boswell, S., Livingstone, I. A., Delafield-Butt, J., and Giardini, M. E. (2020). Screen and virtual reality-based testing of contrast sensitivity. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pages 6054–6057. IEEE.

Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240.

Wolfe, B., Dobres, J., Rosenholtz, R., and Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. *Applied ergonomics*, 65:316–325.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., Lederman, S. J., and Merfeld, D. (2006). *Sensation & perception*. Sinauer Sunderland, MA, fifth edition.

Yan, Z., Song, C., Lin, F., and Xu, W. (2018). Exploring eye adaptation in head-mounted display for energy efficient smartphone virtual reality. In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications*, pages 13–18.