Model of Realism Score for Immersive VR Systems

Benoit Perroud[1,2] Stéphane Régnier[2] Andras Kemeny[1,2]

Frédéric Mérienne[1]

(1) Le2i FRE 2005, **Arts et Métiers**, CNRS, Univ. Bourgogne Franche-Comté, HeSam, Institut Image 2 rue T. Dumorey 71100 Chalon-sur-Saone France. E-mail: {benoit.perroud, frederic.merienne}@ensam.eu

(2) VR and Immersive Simulation Center, **Renault**, Guyancourt, France. E-mail: {benoit.perroud, stephane.regnier, andras.kemeny}@renault.com

Abstract

A model of a realism score for immersive virtual reality and driving simulators is presented. First, we give an outlook of the different definitions of what "realism" is and the different approaches that exist in the literature to objectively quantify it. Then, we present the method, the theoretical development of the score and the results proposed. This realism score system aims to objectively quantify the characteristics of the visual perception happening for a perfect (non-altered vision) observer when experiencing an immersive VR system, as compared to the human visual system in a real (non-VR) situation. It addresses not only the visual perception but also the immersivity of the experience. The approach is different from the signal detection theory and the quantum efficiency theory that both rely on probabilities computation. It is made of several items, graded between 0 and 100, and divided in two sections: vision cues and immersion cues. These items represent, and are based on, the different skills of the human visual system. Realism score could be used as a helping tool in many applications such as objectively grading the performance of a VR system, defining the specifications of a new display, or choosing a simulator between several others available for a given experiment.

Keywords - Virtual Reality, Realism Assessment, Objective Score, Human Visual System

1 Introduction

In the automotive industry, simulators and immersive display systems are largely used for relative and, sometimes, absolute validation of vehicle systems designs, such as Human Machine Interface (HMI) and Advanced Driver Assistance Systems (ADAS), as well as human factors studies [Kemeny, 2009]. Manufacturers hence need trustful tools that can achieve realism in order to make decisions. But what is realism ? A major issue lies within the very conception of every Virtual Reality (VR) system: they are usually designed as individual and unique systems, thus providing various level of performances.

Visual information is often claimed to be 90% of the information used by the driver, though it is yet to be fully proved numerically [Sivak, 1996]. The measure of its quality appears to be essential. In addition, previous studies indicate that the perceptual world of the human differs of that of the real world, driver perceiving a more contracted world relative to the reality [Gilinsky, 1951].

1.1 What is Realism ?

We shall give a definition of realism in the field that matters to us: virtual reality and driving simulation. Realism is often defined as something beautiful looking, for its common use is entertainment [Ferwerda *et al.*, 1996]. A more complete definition can be found in [Fuchs *et al.*, 2003] ; realism takes five possible acceptations:

- Realistic looking: very detailed shaders and materials, hard work on lights in the scene and other artistic tricks.
- Realistic construction of the virtual world: what's implemented is based on scientifically proven models (gravity, dynamics, ...)
- Physiologic realism: the inputs received by the body are the same that those it would receive in a real situation, even if overall it seems odd to the observer.
- Psychological realism: what's implemented seems realistic to the observer, even if it is, in fact, over or under-powered (walking speed, field of view, ...)
- Presence: even if the scene is only made of non textured polygons, the maximum the presence, the better.

Last but not least, and quite in the same idea of the last of the five previous acceptations, [Hoorn *et al.*, 2003] give a different definition of realism. In a simulation, if the goal is achieved, it is said to be realistic enough. For example, in a learning application, if the skill taught is learned and transferable to a real situation, the simulation is said to be realistic.

1.2 Scope of Work

The scope of this work is oriented toward objective physiological realism with respect to the human visual system (HVS) [Perroud *et al.*, 2016]. To our knowledge, literature on establishing an objective model of physiological realism based on vision skills criteria is very sparse. It is not to be mingled with quality assessment of images which is based on the comparison of a source file with the same file that went through a transformation process (like video encoding) [Čadík, 2004]. Presently, there already exists numerous methods for quality assessment of 2D images/video streams, and for the last few years many resources have been engaged on 3D objective and subjective image quality assessment. Different reviews were done earlier

[Moorthy & Bovik, 2013; Moorthy *et al.*, 2013] and completed with a perceptual dimension [Beghdadi *et al.*, 2013].

Realism-wise, the quantifying mostly comes with a question about the realistic-ness of the experimentation a subject just gone through (which is to be tied to the fourth acceptation, "psychological realism"). A more detailed approach goes through questionnaires [Fiard *et al.*, 2014; Fucentese *et al.*, 2015]. There are though different other approaches that are to be highlighted:

- The Rose model [Burgess, 1999; Rose, 1948] is an attempt (at that time for black and white television) to characterize the ideal "image pickup device", to capture a single object. The model can be summarized with the equation $BC^2\alpha^2 =$ constant with B the luminance of the object in footlamberts[1], C the contrast of the object compared to the background (in %) and α the visual angle under which the object is seen.
- The definition of an Ideal Observer [Geisler, 2003] is based on the Signal Detection Theory[2] and the Bayesian theory[3]. It includes probabilities of detection and thresholds of vision. Ideal Observers are though devoted to a specific task like photon detection, pattern discrimination, identification, ...

2 Methodology

2.1 Aim of Study

In an industrial environment, where strategic decisions are to be made, to create aesthetically beautiful experiences is, unless specific applications such as design or perceived quality, (only) a plus. What is needed is to have trustful systems upon which one can rely in a development process. We need to make sure that an engineer will be able to work with a digital mock-up with as much performance as needed.

The aim of this study is to quantify the bias between what's to be displayed in an immersive environment and how well the user will perceive it from the human visual model perspective. How the visual system will interact with the immersive display. How different is this interaction (HVS and display) with respect to the one that would occur if the scenario was happening in real life. The quantification must be fully objective and depends on the physical specifications of the display system. We came with the idea of a score, that would picture, for a given display system, and in a given situation, how efficient is the simulator to broadcast the good level of visual information and immersion factors: this is the so-called "realism score".

[1]NB: $1\ fL \approx 3.426\ cd/m^2$

[2]Ability to discern a specific pattern between different other patterns or noise

[3]Interpretation of the concept of probabilities as states of knowledge -visible or not- instead of a frequency/propensity of a phenomenon

2.2 General Method

Since the score system is to be based on the human visual system, we built our work on the different specifications that can be extracted from vision skills. In addition, other criteria (that we generally call *items*) were added from the different possible immersion cues. Vision skills involve acuity, color and contrast vision, but also depth comprehension [Gross *et al.*, 2008]. Immersion is made possible through techniques like position tracking, stereoscopy and such. Establishment of the different items should be, as far as it is possible, based on values (key-values) or models that are already well described and reputed in the literature. Breaking down vision skills and immersion cues into elementary bricks lead us to establish a model made of 12 items.

We used three different approaches to grade items. The first and main one is to declare that, on our score scale of 100, 0 is the point which is correlated to the value under which (or over which, depending on the case) the vision skill doesn't work anymore, whether because the input is too weak of because the HVS cannot achieve useful processing. The grade of 100 is given to the point at which the HVS is at its maximum capacity, at its finest resolution. Finally, the grade of 80 is given to the value of the HVS that represent its nominal performance.

3 Model Proposition

We propose a score model divided in twelve items distributed over two sections: vision cues and immersion cues. The approach is to be very pragmatical, the model does not pretend to fully and exhaustively describe the vision behavior but to draw links between the human visual model and an immersive VR system's display hardware. The criteria division was made for one category to represent the information an eye would get inside a simulator, while the second category is about what the system provides to the user to help immersion. Immersion is obviously achieved through other processes and we only categorized the ones linked to the system/hardware.

The items are distributed as follows:

- **Vision cues:** contrast and luminosity (luminance), frames per second, number of different colors achievable, field of view, and both monoscopic and stereoscopic acuities.
- **Immersion cues:** latency, field of regard, stereoscopy, tracking, uniformity and camera convergence (which is directly related to eye-tracking).

The whole model overview is presented in Fig. 1. A detailed presentation of the different items is given in the following sections.

Each of these items benefits from its own weighting. Depending on the use-case, the visual system will not be solicited the same way. For example, a design review and a highway driving simulation will not need maximum quality on the same items. The first one will need more colors or a better field of view, while the other will need less latency and smaller pixels to be able to see far away, sharply. Attributing sets of weighting coefficients, per use-case,

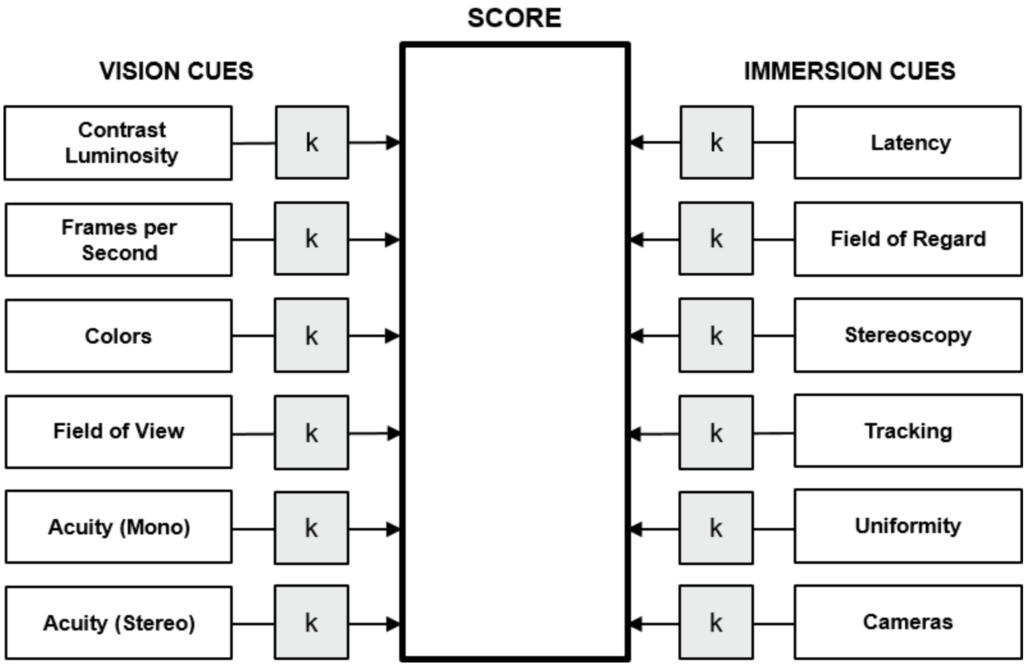


Figure 1: Overview of the proposed Realism Score Model

allows to respect this necessity. The coefficients are to be established through literature and experimentation. At the end of the day, the score is computed through a weighted sum.

When designing an immersive VR system, the goal would be to reach a score around 80-90. A score of 100 could be thought as the ultimate goal but only represents the limits of the HVS, and should not be something to be reached at any costs.

4 Theoretical Development

The following items are part of the model proposition. We present the different values from the literature that lead us to include these items into our model.

4.1 Vision Cues

4.1.1 Contrast & Luminance

Contrast is theoretically defined as being the difference in brightness/color between the light and the dark parts of an image or of an object. Contrast is a very important feature of the HVS as the later is more sensitive to the contrast than to an absolute level of brightness. As any other in psychophysics, this stimulus can be divided in two parts: its magnitude (that we categorized here as Luminance item) and its resolution (Contrast item).

Relative Visual Performance (RVP) is defined as the ability of the HVS to perform a given task, translated to a value between 0 and 1. It takes contrast and luminance as entry parameters and returns a single value. Many different models of RVP were developed, mostly in the eighties, but two of them should be put in the spotlight: the CIE[4] model [Blackwell, 1981] and the Rea model [Rea, 1986].

On one hand, the CIE model was designed through the sum of many experimental results from many researchers. Their goal was to achieve a tool that would help choose the best lightning conditions in factories to make the workers work at maximum yield. Beside of the contrast and the luminance it also requires the age and the task demand level as entry parameters.

On the other hand, Rea model is simpler but they claim to be more accurate. It takes this time only the luminance from the background and the one from the task as entry parameters. An evolved version of this model can be found at [Rea & Ouellette, 1991]. In this case, three more entry parameters were added: age, task size and adaptation luminance (luminostiy at which the eye adapted its pupil size).

Nevertheless, those models are not designed for VR and may not fit directly in our model. An experimentation is currently ran to answer this question. The aim is to directly use the RVP value scaled from 0 to 100. The CIE defines the standard performance with a RVP value of 0.8, which fits perfectly in our model.

4.1.2 Frames per Second

Setting specific values to vision phenomena can be rather unnatural since the process of vision is continuous [Bear *et al.*, 2007]. But there's also exists some notable effects that only occur at specific frame rates.

First of all, motion perception is not based on persistence of vision (as it was long believed) but on two perceptual illusions: phi phenomenon and beta movement [Nichols & Lederman, 1980]. These effects start at 16 frames per second (FPS). Under this threshold, no movement is perceived, only a suite of distinct images.

Secondly, another major feature of vision is flickering. Flickering happens when the frame rate is too slow and a fading effect between the images becomes visible. [Driscoll *et al.*, 1978] aggregated the work of Landis [Landis, 1954] and de Lange [de Lange Dzn, 1958a,b] about the determination of the critical flicker frequency for the eye, beyond which flickering would disappear. The critical flicker frequency seems to be based on the ripple ratio[5] r. Based on the temporal-modulation transfer-function curves, for all ripple ratios, the critical flicker frequency, at any luminance, is equal to 70Hz.

Other values come from the dorsal and ventral pathways. Ventral and dorsal pathways is the main theory on how the brain is dealing with the incoming visual stream. The information is divided in two distinct computation patterns [D'hondt, 2011; Ungerleider & Mishkin, 1982]. The first one (dorsal or parietal pathway) is the "where" loop and is sensitive to the direction and to the movement. The second one (ventral or temporal) is the "what" loop and is sensitive

[4] Commission Internationale de l'Eclairage

[5] Ripple ratio: $r = \frac{\text{Amplitude Of The Fundamental Harmonic}}{\text{Mean Luminance}}$

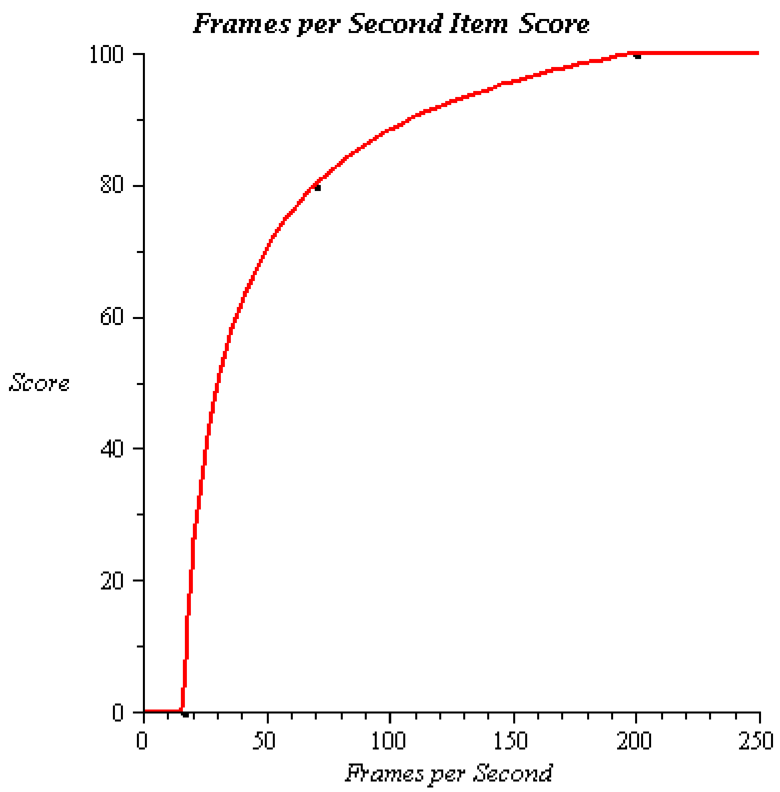


Figure 2: Plot of the Frames per Second Item Score Function

to shape, color and texture. The two pathways are allegedly running at around respectively 200Hz and 25Hz, which is correlated by the fact that one is fast and the other slow [D'hondt, 2011], and by the number of cortexes the pathways run through [D'hondt, 2011] and the own distinct latencies of the different cortexes [Bullier, 2001] ; but no real scientific proven values were found yet. US Air Force is said to have done experiments with jet pilots that were able to recognize a plane with images flashed at $1/220th$ of a second only[6].

Thus, by using the three following key values: minimum frame rate for movement illusion (16 Hz), maximum flicker frequency (70 Hz) and alleged dorsal frequency (200 Hz), we can fit a curve and solve the coefficients of the equation. For this item, with f being the number of frames per second, we come with the following equation (Eq. 1, plot given in Fig. 2.):

$$F_{FPS}(f) = \begin{cases} 0 & f < 16 \\ 126.5 - \frac{367.1}{\sqrt{f-7.6}} & else \\ 100 & f > 200 \end{cases} \tag{1}$$

Finally, it is to be noted that all those key values are per eye, thus that they must be multiplied by two to be used in stereo for 3D imaging. Moreover these frequencies are the effective ones and not the hardware ones, which are more numerous because of the Shannon principle.

[6]Human Eye Frames Per Second. In *AMO.net America's Multimedia Online*. Seen at http://amo.net/nt/02-21-01fps.html

4.1.3 Visual Acuity

Monoscopic Acuity Monoscopic acuity is the accuracy of the human eye, its resolution. This means how small are the smallest things that can be seen or how thin is the smaller perceived step between two patterns. This concept can be directly linked with the size of the pixels in a driving simulator or immersive display: the smaller the pixel the more accurate the image. But at some point, the pixel size can become smaller than the resolution of the eye and then become invisible, hence, somehow less useful.

Usually, it is said that the HVS has a monoscopic acuity between 30" of arc and 2' of arc, with a practical mean of 1' of arc [Fuchs *et al.*, 2003]. Those values can be, in photopic conditions, refined per task [Gross *et al.*, 2008] (Cf. Table 1).

Table 1: Acuity of the eye, [Gross *et al.*, 2008]

Task	Acuity
Pattern recognition	5'
Grating resolution	2'
Two-points (same colors) resolution	1'
Two-points (inverted colors) resolution	30"
Vernier acuity (small parallel straight lines)	10"
Stereoscopic resolution of depth	5"

Vernier resolution can only be achieved in specifics conditions and stereoscopic acuity is to be adressed with another approach that is described later on. Moreover, Deering [Deering, 1998], showed that the smallest resolution achievable on a screen is 28" of arc. Thus, only values from 5' to 30" of arc are to be retained for the study.

The equation is proposed from these values, with α being the angle under which the pixel is seen, in $arcmins$ (Eq. 2, plot given in Fig. 3.):

$$F_{mono\_acuity}(\alpha) = \begin{cases} 0 & \alpha > 3.5 \\ 128.9 - 68.8\sqrt{\alpha} - \frac{0.1}{\alpha} & else \\ 100 & \alpha < \frac{1}{6} \end{cases} \tag{2}$$

Stereoscopic Acuity Stereo acuity is the ability of the Human Visual System to perceive the difference in depth between two planes, at a given distance. It is well known and described in literature. Stereo acuity follows a geometric pattern that is demonstrated in [Fuchs *et al.*, 2003; Gross *et al.*, 2008]. The generally accepted model is as shown in Eq. 3.

$$\Delta r = 0.001 r^2 \tag{3}$$

Where Δr is the theoretical minimum observable difference in depth (in millimeters) and r the distance of observation (in meters). The 0.001 factor is the fraction between the physiological threshold of stereoscopic vision ($\Delta\nu_{min}$) and the inter-pupillary distance (IPD).

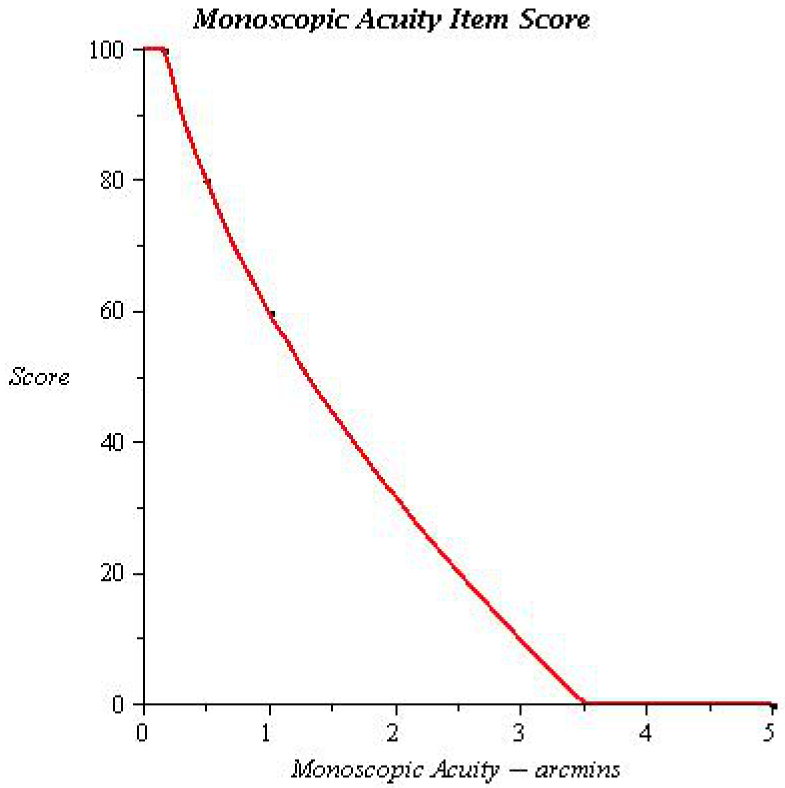


Figure 3: Plot of the Monoscopic Acuity Item Score Function

However, these two parameters can vary significantly. Inter-pupillary distance ranges from 52 to 78mm over the population [Dodgson, 2004] while physiological threshold of vision is dependent on the luminance, especially at very low levels [Gross *et al.*, 2008]. Plus, a general routine to measure the Δr value in any cases would be difficult to set up as it would be the mean value taken from experimentation with subjects through the routine. Exploiting the relation in [Gross *et al.*, 2008] (Eq. 4) that describes the same stereoscopic acuity model, physiological threshold of vision appears to be a core value of the stereoscopic vision and the only parameter linked with both the human visual perception and the system's display that can easily and objectively be measured.

$$\Delta\nu_{min} = \frac{d_{IPD} * \Delta r}{r^2} \tag{4}$$

With $\Delta\nu_{min}$ the limiting angle for stereoscopic vision, d_{IPD} the inter-pupillary distance, Δr the difference of depth that can be perceived at a distance of r meters.

In fine, the item score equation is the fraction between the limiting stereoscopic angle at the lowest luminance the VR display should have ($\Delta\nu_{min}$ taken from the graph in [Gross *et al.*, 2008], in $arcsecs$) and the actual angular resolution the VR display can achieve (α, in $arcsecs$) (Eq. 5):

$$F_{stereo\_acuity}(x, r) = \begin{cases} 100 & \alpha < \Delta\nu_{min} \\ 100 * \frac{\Delta\nu_{min}}{\alpha} & else \end{cases} \tag{5}$$

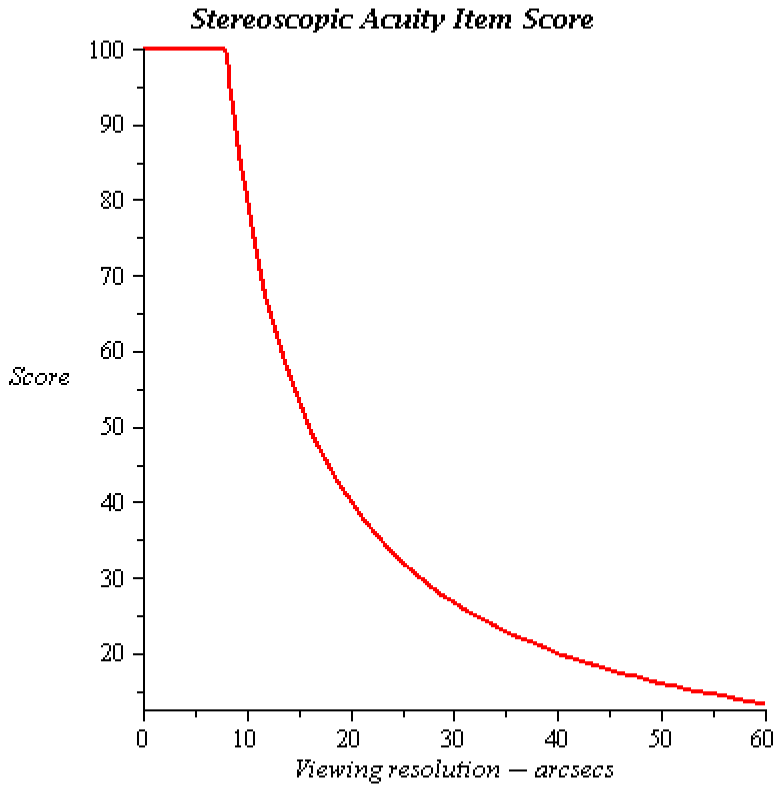


Figure 4: Plot of the Stereoscopic Acuity Item Score Function ($\Delta\nu_{min} = 8\ arcsecs$)

In a CAVE application, the luminance ranges generally from 10^{-1} to $10^{+1}\ cd/m^2$, which gives a $\Delta\nu_{min}$ average value of $8\ arcsecs$[7]. The plot of the function in this case can be seen in Fig. 4.

4.1.4 Colors

In 1931, The International Commission on Illumination (CIE), established the definition of the RGB Color Space. This color space represents the colors that can be seen by a normal 3-cones human eye. Though, no technique has been developed to render 100% of the 1931 CIE Color Space, yet. All the different systems use a fraction of the Color Space, which is called gamut, that depends on the three (or more) primary colors that are chosen. See in Table 2 some typical values of coverage by gamuts of the CIE 1931 Color Space (note that for the ProPhoto RGB gamut, 13% of the colors are imaginary (the primary colors are taken outside of the 1931 Color Space)). Empirically, acceptation begins at Adobe RGB.

As we could not find any paper on how many colors are enough for the human visual system -even the number of discernible colors is still not precisely determined with values between 100.000 and 10 millions colors[8]-, the item proposition is a linear function between the score

[7]This value was obtained by averaging the value of the limiting stereoscopic angle, in the range of luminance of the CAVE. The evolution of the limiting stereoscopic angle with respect to the luminance can be found p.67 in [Gross *et al.*, 2008]

[8]Number of Colors Distinguishable by the Human Eye. In *The Physics Factbook*. Seen at `hypertextbook.com/facts/2006/JenniferLeong.shtml`

Table 2: Gamuts Coverage of 1931 Color Space

BR.709 (HDTV)	35.9%
Adobe RGB	52.1%
Digital Cinema	53.6%
BT.2020 (UHD)	75.8%
Wide-Gamut RGB	77.6%
ProPhoto RGB	90.0%

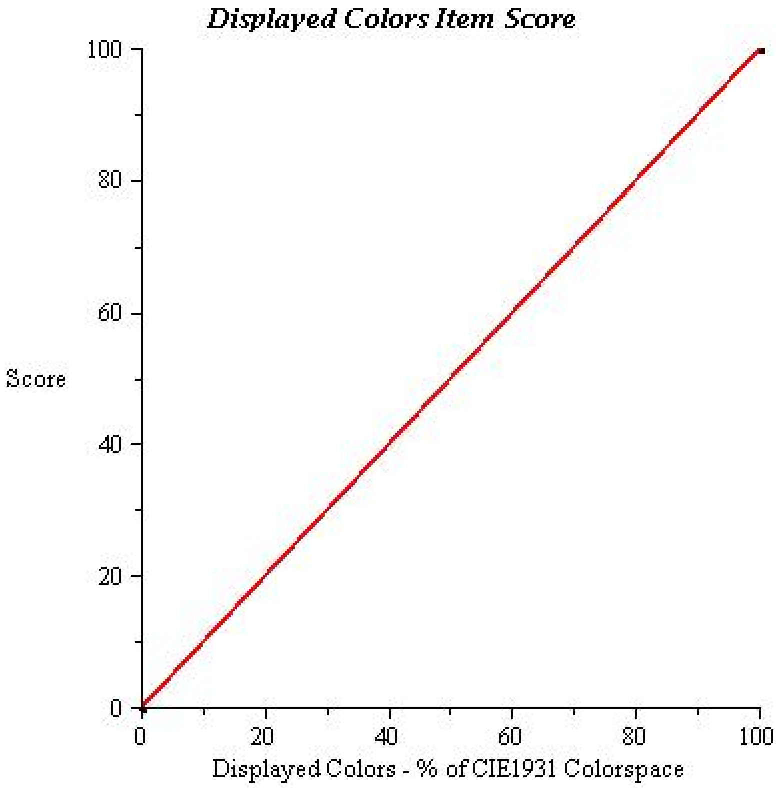


Figure 5: Plot of the Color Item Score Function

and the percentage of the color space that can be rendered by the system (See Fig. 5). Item score equation is then, with c the percentage of the gamut area compared to the color space:

$$F_{color}(c) = \begin{cases} c \\ 100 & c > 100 \end{cases} \tag{6}$$

4.1.5 Field of View

Field of view (FOV) is defined as being the portion of space that one can see at a given time, without moving the head. It is not to be confused with the Field of Regard (FOR) which is the total portion of space that can be viewed with the movements of the head and of the eyes over the time.

The field of view is made of two different axes: vertical and horizontal. We propose here two sets of values, one for each axis. These axes will then be weighted relatively to each other. We make the hypothesis that it is with respect to their relative size. Field of view is to be measured when the subject is in the medium/standard position (both body placement and head orientation).

Horizontal Axis Some of the most common values can be found in [Devisme, 2004]. On the azimuth (horizontal) angle, with both eyes, a standard human being is able to see from 170 to 190 degrees ; while on the elevation (vertical) angle it is only 130 to 145 degrees. Within the azimuth angle, 120 degrees (only) are capable of binocular vision. Binocular vision is possible because of the overlapping of the portions of space seen by each of the two eyes. Maximum acuity is achieved within the foveal area, which is a 3-5 degrees portion of space, at the center. Reading can only be achieved in a cone of 20 degrees while patterns recognition needs 40 degrees and color discrimination 60 degrees. Local equation becomes, with h the value (in degrees) of the horizontal field of view (H-FOV) calculated in the immersive system (Eq. 7):

$$F_h(h) = \begin{cases} 0 & h < 20 \\ 19.6\sqrt{h} - 0.5 \cdot h - 78.3 & otherwise \\ 100 & h > 180 \end{cases} \tag{7}$$

Vertical Axis Three characteristic values can be retained out of the many that one can found in the literature for the vertical field of fiew (V-FOV). We shall consider the values of lateral vision (130 degrees), induced impression (85 degrees) and lookout (20 degrees) [Langlois, 2013]. Local equation becomes, with v the value (in degrees) of the vertical field of view (Eq. 8):

$$F_v(v) = \begin{cases} 0 & v < 20 \\ 32.0\sqrt{v} - 1.1 \cdot v - 121.1 & otherwise \\ 100 & v > 130 \end{cases} \tag{8}$$

Relative Weighting We make our own hypothesis of the equal importance of the vertical axis compared to the horizontal axis in the field of view. Hence we weight them in the item function with respect to their size (180 degrees on horizontal axis and 130 degrees on the vertical one)(Eq. 9).

$$\begin{cases} k_h = \frac{180}{180+130} = 0.58 \\ k_v = 1 - k_h = 0.42 \end{cases} \tag{9}$$

At the end, the field of view item function comes like this (Eq. 10):

$$F_{FOV}(h, v) = k_h \cdot F_h(h) + k_v \cdot F_v(v) \tag{10}$$

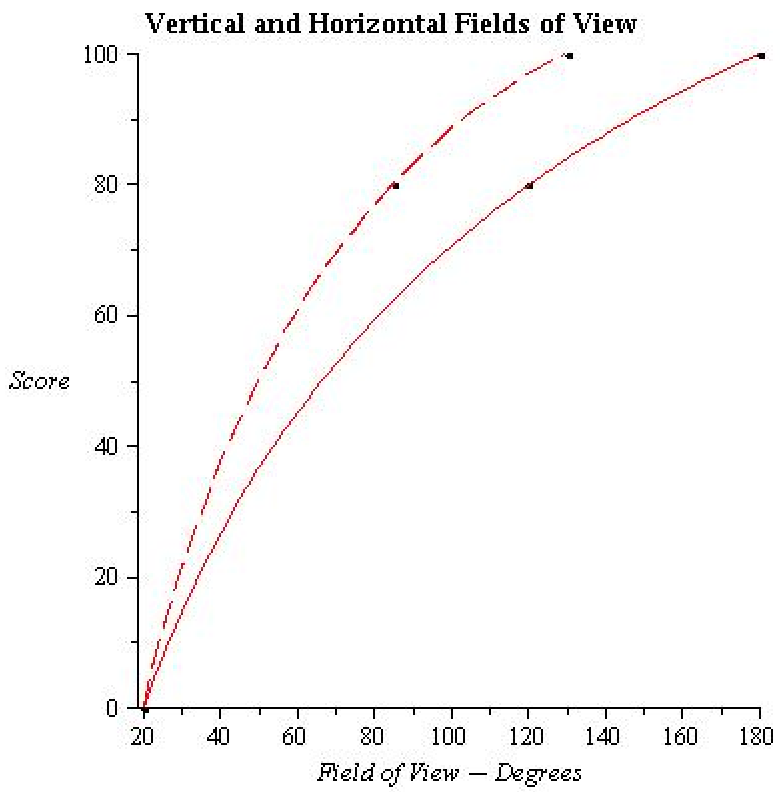


Figure 6: Plot of the Field of View Item Score Sub-Functions F_h (solid line) and F_v (dashed line)

4.2 Immersion Cues

4.2.1 Latency

Latency can have different definitions that are as follows:

- Movement to photon: from the movement of a tracked user to the display of the new point of view. Can be thought of as the "total" latency of a system and is mostly used.
- Movement to pre-calculation: time elapsed between the movement of a tracked observer and the order to re-compute the image at a new position.
- Pre-calculation to calculation: time elapsed to go through the rendering pipeline.
- Computation to photon: time elapsed to display the newly computed image.

Though there exists perceived latency thresholds measurements in the literature [Brooks, 1999; Kemeny, 2014], and values of visual latency for bright (74 ms) and dark (106 ms) images [Feng Han *et al.*, 2010], we would like to enhance these information and correlate them with the HVS.

4.2.2 Field of Regard

Field of regard (FOR) is the extension of Field of View. It is defined as the portion of space that one can see over the time, including the movements of the head and of the eyes. Alike the previous item, Field of View, we must elaborate the score function in two steps: horizontal (H-FOR) and vertical field of regard (V-FOR) and then their relative weighting.

Horizontal and Vertical Fields With respect to the possible movements of the head and of the eye in its orbit (up to 15 degrees), each eye has a field of regard of: over than 200 degrees on temporal side and around 130 degrees on nasal side on the horizontal axis ; and 310 degrees on the vertical axis, distributed as $+140$ and -170 degrees [Fuchs *et al.*, 2003]. Hence, on one hand, and since visual fields overlaps, we set the maximum value of H-FOR at 360 degrees (h_{max}). V-FOR maximum is set at the maximum physical value, 310 degrees (v_{max}. On the other hand, the minimum values cannot be lower than the values of the Field of View item (that we call here h_0 and v_0). The evolution is linear within each sub-item.

The equations become (Eq. 11), for H-FOR and V-FOR, with h and v the measured values on horizontal and vertical axes ; and h_0, v_0, h_{max}, v_{max} previously defined:

$$\begin{cases} F_{H-FOR}(h) = \frac{100}{h_{max}-h_0} \cdot (h - h_0) \\ F_{V-FOR}(v) = \frac{100}{v_{max}-v_0} \cdot (v - v_0) \end{cases} \tag{11}$$

Relative Weighting With the same hypothesis as for the FOV weighting, the relative weighting of the two sub-items of Field of Regard is based on their relative maximum size: 400 degrees for horizontal axis and 310 degrees for vertical axis. With the same methodology as for FOV, we obtain (Eq. 12):

$$\begin{cases} k_h = \frac{400}{400+310} = 0.56 \\ k_v = 1 - k_h = 0.44 \end{cases} \tag{12}$$

Which are very similar to the previous case. At the end of the day, the final Field of Regard score equation becomes (Eq. 13):

$$F_{FOR}(h, v) = k_h \cdot F_{H-FOR}(h) + k_v \cdot F_{V-FOR}(v) \tag{13}$$

4.2.3 Stereoscopy

Stereoscopy is the way of giving depth and relief to standard 2D images. It can be achieved through many different techniques (HMD, shutter glasses, polarized glasses, anaglyph, ... a review can be found at [Fauster & Wien, 2007]). There also exists other means to re-create the binocular vision, such as auto-stereoscopic displays, holographic displays, ... As for tracking, the proposed rating method is based on presence (100) and absence (0).

4.2.4 Tracking

Tracking is one of the most (if not THE most) important feature of immersion as it enables movement in the simulation (body and/or head), the inclusion of external devices and such. We propose to rate it only based on its presence (100) or absence (0).

4.2.5 Uniformity

The Uniformity item can be divided in two: (Single-)Screen Uniformity and System Uniformity, when it applies. It involves the ability of the user to perceive a difference of contrast, luminance or color between different areas of the same -or different- screen(s).

4.2.6 Camera Convergence

The vergence of cameras (toe-in cameras) is the opposite of parallel-cameras image capture technique. Toe-in cameras do generate distortions [Woods *et al.*, 1993], but only when the viewing point is not known. In this case, a viewing point has to be assumed to configure the convergence of the cameras. Whenever looking at a point that is at a different distance that of the assumed point, distortions occurs. The assumed distance of the viewing point can be corrected in real time with eye-tracking. Another approach is to use parallel cameras, assuming the viewing point is at infinity. This hypothesis is good whenever looking at far objects, when eyes' directions would be indeed parallels. But at the moment one knows the viewing point, with the correct image processing to transform non-parallel captured images into alongside displayed images, it becomes beneficial to the user as it reproduces the movement of the eyes into the orbits and adds vertical disparities [Aurat, 2016].

Hence, the item is divided into the three possible outcomes: toe-in cameras without knowledge of the viewing point (0), parallel cameras (80) and toe-in cameras with knowledge of the viewing point (100).

5 Experimental Study

5.1 Aim of Experimentation

An experimentation was carried out to compare the grades given by our score model to acceptation grades given by subjects using an immersive VR system. This experiment cannot validate our model (see section 6.2) as it compares data of two different nature. This is only a comparison between a theoretical model that provides a score which reflects how close to the visual model the hardware of an immersive VR system is, and a score given subjectively by subjects -through experimentation- that reflects how much the subjects subjectively evaluate the different capabilities of the system. However, this experimentation allows us to show how different can be the appreciation of an item compared to its real life characteristics, and

INSUFFICIENT		GOOD		MAX
1	2	3	4	5
			X	

Figure 7: 1-to-5 aided scale used to subjectively rate the different criteria of the model

provides hints on how users behave in the system in the very use-case we used ; which can help in some of our hypotheses.

5.2 Apparatus

The experimentation involved 32 young subjects: 23 males and 9 females, between the ages of 20 and 27 (M = 25, SD = 1.8) years. It was carried out in a 4-faces CAVE display system. The dimensions of the CAVE are given in the Table 3. The pixel size was $2.25\ mm$ and the subject was seated $2\ m$ away from the front face. The experimentation was ecological: subjects had to drive in the simulation. The driving was made with a Logitech G25 wheel+pedals connected to a SCANeR Studio simulation. The simulation environment was a simple riviera tour with different landscapes (sea, lakes, city and mountains).

Table 3: Dimensions of the CAVE-like immersive VR system

Face	Physical size	Display resolution
Front	3.60 x 2.70 m	1600 x 1200 px
Left	4.20 x 2.70 m	1920 x 1200 px
Right	4.20 x 2.70 m	1920 x 1200 px
Floor	4.20 x 2.70 m	1600 x 1200 px

Every subjects had to do the same task: a eight-minutes drive with no particular other task than looking around while driving. The speed of the car was constrained under 30 km/h to stay in a use-case of low-speed driving. After the eight minutes, the subjects were asked to rate on a 1-to-5 aided scale (see Fig. 7) the different criteria of our model, compared to what they would have had in real life. A comparison was then made between the mean of the subjects' subjective ratings and the objective values from the score model. No weighting on the items were applied.

To represent the span of different heights of subjects, the field of view and field of regard theoretical values were computed twice, with a $1.10\ m$ and a $1.15\ m$ eye height from the ground, and then averaged.

Only young subjects (ages from 20 to 27) with perfect (or corrected to perfect) vision were taken. It was made, though the Rea and CIE models take age as a parameter, because of the aging of the visual system that begins around 30 y.o., to be sure to have subjects with full vision capacities.

5.3 Results

In order to analyze the results, the population of subjects was then divided into 8 sub-groups. The characteristics of the different sub-groups can be found at Table 4. No statistical results can be withdrawn from the *women gamer* $(W + G]$ sub-group as there was only 1 subject that fitted the category.

Table 4: Population and age characteristics of the sub-groups of subjects

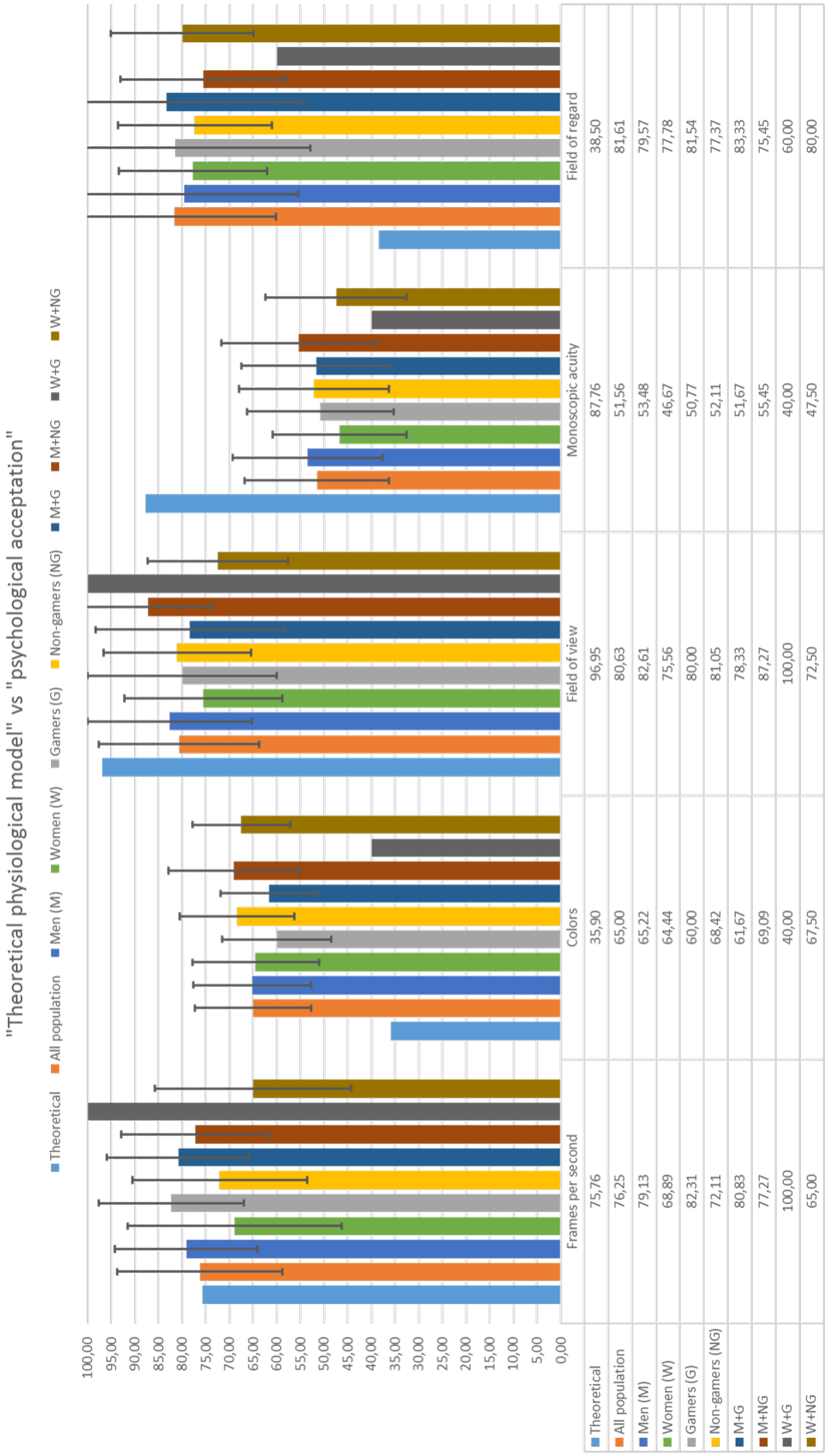
Group	Population	Average age	SD
All subjects	32	24.8 y.o.	1.8
Men (M)	23	24.8 y.o.	2
Women (W)	9	24.8 y.o.	1.7
Gamers (G)	13	24.5 y.o.	2.3
Non-gamers (NG)	19	25 y.o.	1.5
M+G	12	24.7 y.o.	2.3
M+NG	11	24.9 y.o.	1.6
W+G	1	22 y.o.	0
W+NG	8	25.1 y.o.	1.5

Five theoretical score values from the model were computed and graded by subjects: frames per second, colors, monoscopic acuity, field of view and field of regard. The other items from the model were either not yet gradable theoretically (luminance, contrast, latency and uniformity) and thus cannot be compared to their subjective equivalent, or difficult to rate through a simple questionnaire without specific task in the simulation to evaluate it (stereoscopic acuity, tracking and stereoscopy). Values of means and SD (except for theoretical value and the Woman+Gamer sub-group in which there was only one subject) are to be found in Fig. 8.

5.4 Discussion

The results from the comparison between the theoretical model and the psychological acceptation can be divided in three outcomes. First, the *frames per second* measured values fit well with the theoretical value. Secondly, and opposed to the first outcome, two items are strongly different from the acceptation values ; in one way (monoscopic acuity: theoretical value greater than acceptation values) and the other (colors: acceptation values greater than theoretical value). Last but not least, the third outcome gives similar acceptation values with very different theoretical values (96.95 for Field of View and a poor 38.50 for Field of Regard). We will discuss the first and second outcomes.

Colors and monoscopic acuity Both items show a great difference between the grade based on physiology and the grade based on acceptation. Nevertheless, within the sub-groups, means and standard deviations are somewhat close. On one hand, the subjects were mostly satisfied with the quantity of colors though they are very few with respect to all the colors a human being can see (CIE XYZ 1931 Color Space). A possible cause may be the absence



	Frames per second	Colors	Field of view	Monoscopic acuity	Field of regard
Theoretical	75,76	35,90	96,95	87,76	38,50
All population	76,25	65,00	80,63	51,56	81,61
Men (M)	79,13	65,22	82,61	53,48	79,57
Women (W)	68,89	64,44	75,56	46,67	77,78
Gamers (G)	82,31	60,00	80,00	50,77	81,54
Non-gamers (NG)	72,11	68,42	81,05	52,11	77,37
M+G	80,83	61,67	78,33	51,67	83,33
M+NG	77,27	69,09	87,27	55,45	75,45
W+G	100,00	40,00	100,00	40,00	60,00
W+NG	65,00	67,50	72,50	47,50	80,00

Figure 8: Results of the theoretical computed values and the subjectively rated values

of real-world reference to compare the colors and by the frequent use of displays using the same color space (smart phones, televisions, cinemas ...). On the other hand, monoscopic acuity was poorly graded by the subjects though it should be sufficient. At the end of the day, these differences shall help us on designing the coefficients needed to complete the final score model.

Field of view and field of regard Though a bit higher, the field of view theoretical value is still close to the acceptation values and contained in most of the standard deviations of the sub-groups. However, the Field of Regard theoretical value is dramatically low compared to the acceptation values and outside of all the standard deviations ranges. The low theoretical value comes from that, even if there is a very good horizontal viewing angle in the CAVE, the lack of ceiling face is heavy for the vertical component of the grade. Hence, the global field of regard grade is deteriorated because of the close weighting of the two components. A driving task requires full attention on the front side of the car and involves mostly the horizontal part of vision. That's why subjects had not to use frequently out-of-screen parts of the CAVE and graded the field of regard item closely to the field of regard item. This implies that, in this specific case, the hypothesis of size-based weighting for the horizontal and vertical field of regard we made is not efficient. A specific set of weights should be applied, per use-case, based on the needs of visual space for the application.

Lastly, it is interesting to point out a tendency (even though not statistical). Women (women, and non-gamer women) seem to almost always give lower grades than men, whatever the theoretical grade is.

6 Discussion & Future Work

6.1 Going Further

As it is only a first proposal of this model, some items are not developed enough and must be enhanced. For example, the Stereoscopy item which is now only rated by the presence or absence of the technique could integrate the hardware behind: anaglyphs may not be as efficient as shutter glasses.

Also, other limitations are known, like for monoscopic acuity, which values were shown to be dependent from contrast, speed, luminance and even the age of the observer [Gross *et al.*, 2008]. Though it is well studied [Cavonius & Robbins, 1973; Millodot, 1969; Owsley & Sloane, 1987], no theoretical models have emerged yet.

6.2 Validation and weighting coefficients

Finally, the weighting of the model (through coefficients) must be determined based on use-cases. The same weights cannot be used for different applications and sets of coefficients must be proposed to cover use-cases. However, a generic grade could be given trough an equal weighting of all the parameters. This score would only allow to compare systems' hardware and couldn't be related to any use-case.

A definitive and complete validation of the model is not yet possible as it would need the model to be fully finished. Though quite advanced, the current proposition is still under construction. Nevertheless, a first outlook was given through an experimentation that compared our model to the subjective acceptation of its criteria.

7 Conclusion

We described the different possible meanings of the word realism and showed that there is a lack of methods to objectively quantify one of its acceptations. We hence proposed a model of score to quantify physiological realism (only). It aims to evaluate the bias between the virtual scene and the perception of a perfect observer, through an immersive display system, compared to the perception this observer would have in real conditions. This model is an extended version of a previous model. The score is made of weighted items that represent the different perks of vision and immersion. We proposed equations and plots for most of the items, based on key-values and models from the literature. An experimentation was conducted to compare the graves given by the model to grades of acceptation given by subjects. This experimentation was conducted in a particular application context (of a low-speed driving simulation).

Main differences with the previous version of the model are that more items are implemented, those above being distributed in two different branches. Still, some items need to be fully established through experimentation and some others must be enhanced with a more precise and detailed oncoming. The final steps will be to determine, through experimentation and literature, different weighting coefficients sets of every items and then to fully validate the score system. Other acceptation experimentation could be performed to quantify the differences between different scenarii.

Bibliography

Čadík, Martin. 2004. Human Perception and Comupter Graphics. *In: in Czech Technical University Postgraduate Study Report.*

Aurat, David. 2016. *Immersion Visuelle Tridimensionnelle Haute Performance par Suivi du Mouvement des Yeux.* Informatique Temps Réel, Robotique & Automatique, Université de Recherche Paris Scienes et Lettres (PSL Research University), Paris, France.

Bear, Mark F, Connors, Barry W, & Paradiso, Michael A. 2007. *Neurosciences: à la découverte du cerveau.* Paris: Éditions Pradel.

Beghdadi, A., Larabi, M.-C., Bouzerdoum, A., & Iftekharuddin, K.M. 2013. A survey of perceptual image processing methods. *Signal Processing: Image Communication*, **28**(8), 811–831.

Blackwell, H. R. 1981. *An Analytic model for describing the influence of lighting parameters upon visual performance*. Publication CIE, nos. no. 19/2.1–19/2.2. Paris, France: Bureau central de la CIE.

Brooks, F.P. 1999. What's real about virtual reality? *IEEE Computer Graphics and Applications*, **19**(6), 16–27.

Bullier, Jean. 2001. Integrated model of visual processing. *Brain Research Reviews*, **36**(2-3), 96–107.

Burgess, Arthur E. 1999. The Rose model, revisited. *Journal of the Optical Society of America*, **16**(3), 633–646.

Cavonius, C. R., & Robbins, D. O. 1973. Relationships between luminance and visual acuity in the rhesus monkey. *The Journal of Physiology*, **232**(2), 239.

de Lange Dzn, H. 1958a. Research into the Dynamic Nature of the Human Fovea-Cortex Systems with Intermittent and Modulated Light I Attenuation Characteristics with White and Colored Light. *Journal of the Optical Society of America*, **48**(11), 777.

de Lange Dzn, H. 1958b. Research into the Dynamic Nature of the Human Fovea-Cortex Systems with Intermittent and Modulated Light II Phase Shift in Brightness and Delay in Color Perception. *Journal of the Optical Society of America*, **48**(11), 784.

Deering, Michael F. 1998. The Limits of Human Vision. vol. 2.

Devisme, Céline. 2004. *Optimisation du Traitement des Indices de Profondeur pour Application à la Vision Artificielle - Cas de la Disparité Horizontale*. Mémoire.

D'hondt, Fabien. 2011. *Emotion et Espace Visuel: Approche Neuromagnétique, Neurosomatique et Comportementale*. Ph.D. thesis.

Dodgson, Neil A. 2004 (May). Variation and Extrema of Human Interpupillary Distance.

Driscoll, Walter G., Vaughan, William, & Optical Society of America (eds). 1978. The Eyes and Vision. *In:* Driscoll, Walter G., Vaughan, William, & Optical Society of America (eds), *Handbook of optics*. New York: McGraw-Hill.

Fauster, L., & Wien, T. U. 2007. Stereoscopic Techniques in Computer Graphics. *Tu Wien*.

Feng Han, Tao Xu, Chaoping Tian, & Zhenjie Hou. 2010. Investigation on human visual response latency. IEEE.

Ferwerda, James A., Pattanaik, Sumanta N., Shirley, Peter, & Greenberg, Donald P. 1996. A Model of Visual Adaptation for Realistic Image Synthesis. ACM.

Fiard, Gaelle, Selmi, Sonia-Yuki, Promayon, Emmanuel, Vadcard, Lucile, Descotes, Jean-Luc, & Troccaz, Jocelyne. 2014. Initial Validation of a Virtual-Reality Learning Environment for Prostate Biopsies: Realism Matters! *Journal of Endourology*, **28**(4), 453–458.

Fucentese, Sandro F., Rahm, Stefan, Wieser, Karl, Spillmann, Jonas, Harders, Matthias, & Koch, Peter P. 2015. Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. *Knee Surgery, Sports Traumatology, Arthroscopy*, **23**(4), 1077–1085.

Fuchs, Philippe, Moreau, Guillaume, & Berthoz, Alain. 2003. *Le Traité de la Réalité Virtuelle. Volume 1.* Vol. 1. Paris: Les Presses de l'École des Mines.

Geisler, Wilson S. 2003. Ideal Observer Analysis. *The Visual Neurosciences*, **10**(7).

Gilinsky, Alberta S. 1951. Perceived Size and Distance in Visual Space. *Psychological Review*, **58**(6), 460.

Gross, Herbert, Blechinger, Fritz, Achtner, Bertram, & Gross, Herbert. 2008. Human Eye. *In: Survey of Optical Instruments*. Handbook of Optical Systems, no. ed. by Herbert Gross ; Vol. 4. Weinheim: Wiley-VCH.

Hoorn, J.F., Konijn, E.A., & Van der Veer, G.C. 2003. Virtual Reality: Do Not Augment Realism, Augment Relevance. *Upgrade-Human-Computer Interaction: Overcoming Barriers*, **4**(1), 18–25.

Kemeny, Andras. 2009. Driving Simulation for Virtual Testing and Perception Studies. *In: Proceedings of the DSC 2009 Europe*.

Kemeny, Andras. 2014. From driving simulation to virtual reality. ACM Press.

Landis, Carney. 1954. Determinants of the Critical Flicker-Fusion Threshold. *Physiological Reviews*, **34**(2), 259–286.

Langlois, Sabine. 2013. ADAS HMI using peripheral vision. ACM Press.

Millodot, Michel. 1969. Variation of Visual Acuity With Contact Lenses: A Function of Luminance. *Archives of Ophthalmology*, **82**(4), 461.

Moorthy, Anush K., & Bovik, Alan C. 2013 (Mar.). A survey on 3D quality of experience and 3D quality assessment.

Moorthy, Anush Krishna, Su, Che-Chun, Mittal, Anish, & Bovik, Alan Conrad. 2013. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, **28**(8), 870–883.

Nichols, Bill, & Lederman, Susan J. 1980. Flicker and Motion in Film. *Pages 96–105 of:* de Lauretis, Teresa, & Heath, Stephen (eds), *The Cinematic Apparatus*. London: Palgrave Macmillan UK.

Owsley, C., & Sloane, M. E. 1987. Contrast sensitivity, acuity, and the perception of 'real-world' targets. *British Journal of Ophthalmology*, **71**(10), 791–796.

Perroud, Benoit, Régnier, Stéphane, Kemeny, Andras, & Mérienne, Frédéric. 2016. Realism Score for Immersive Virtual Reality Systems and Driving Simulators. *Pages 213–215 of: Proceedings of the Driving Simulation Conference DSC 2016 Europe VR*.

Rea, Mark S. 1986. Toward a Model of Visual Performance: Foundations and Data. *Journal of the Illuminating Engineering Society*, **15**(2), 41–57.

Rea, Mark S., & Ouellette, Michael J. 1991. Relative visual performance: A basis for application. *Lighting Research and Technology*, **23**(3), 135–144.

Rose, Albert. 1948. The Sensitivity Performance of the Human Eye on an Absolute Scale*. *Journal of the Optical Society of America*, **38**(2), 196.

Sivak, Michael. 1996. The Information That Drivers See: Is It Indeed 90% Visual ? *Perception*, **25**(9), 1081–1089.

Ungerleider, Leslie G., & Mishkin, Mortimer. 1982. Two Cortical Visual Systems. *In:* Ingle, David, Goodale, Melvyn A., & Mansfield, Richard J. W. (eds), *Analysis of visual behavior*. Cambridge, Mass: MIT Press.

Woods, Andrew J., Docherty, Tom, & Koch, Rolf. 1993 (Sept.). Image distortions in stereoscopic video systems.