

THÈSE DE DOCTORAT

**Immersion Visuelle Hyper-Réaliste et**

**Multi-Sensorielle 3D**

# RÉSUMÉ

La thèse « Immersion visuelle hyper-réaliste et multi-sensorielle 3D » a été réalisée dans le cadre d'un contrat CIFRE établit entre les Arts et Métiers d'une part et Renault SAS d'autre part. Elle propose un modèle de score permettant d'évaluer objectivement de la capacité d'un système d'affichage immersif à reproduire le bon niveau de stimulation sensorielle pour un utilisateur, par rapport à ce qu'il recevrait dans la réalité et à la modélisation du système visuel humain.

Dans un premier temps nous nous sommes intéressés à poser les bases du modèle : celui-ci est composé de douze critères, répartis équitablement en une somme d'indices de vision et d'indices d'immersion. Chaque critère se voit attribuer, dans la mesure du possible, une note de 0 à 100. La note de 0 représente l'incapacité du système visuel à percevoir ou à utiliser les informations visuelles, tandis que la note de 100 incarne la capacité maximale. Une note de 80 est également assignée pour la performance standard. Chaque critère se voit assigner une pondération en fonction de la tâche réalisée en environnement virtuel.

Nous réalisons dans un second temps une série d'expérimentations afin de compléter les informations disponibles dans la littérature, pour l'établissement des critères. On s'intéresse plus particulièrement au contraste et à la latence. La première expérimentation consiste en la transposition et la validation d'un modèle de performance visuelle en Réalité Virtuelle. Dans la seconde, on compare les effets de la latence sur une tâche donnée, entre un simulateur de type CAVE et un casque de Réalité Virtuelle.

La thèse propose un certain nombre de résultats tant théoriques et méthodologiques qu'expérimentaux. En condition d'immersion virtuelle, les critères de contraste et de luminance sont importants pour la perception visuelle. Les expériences réalisées montrent qu'un modèle adapté aux conditions d'immersion virtuelle est nécessaire. Par ailleurs, on montre que des seuils d'influence de la latence sur la performance semblent exister. On vérifie également la pertinence de corrélations entre performance, présence et mal du simulateur, en fonction de la latence et du système immersif.

# ABSTRACT

The thesis "Hyperrealistic and multi-sensorial 3D visual immersion" was carried out within the framework of a CIFRE contract established between Arts et Métiers on the one hand and Renault SAS on the other. It proposes a score model to objectively evaluate the ability of an immersive display system to reproduce the right level of sensory stimulation for a user, compared to what he would receive in reality and to the modeling of the human visual system.

First, we were interested in laying the foundations of the model : it is composed of twelve criteria, equitably divided into a sum of vision indices and immersion indices. Each criterion is given, as far as possible, a score from 0 to 100. A score of 0 represents the visual system's inability to perceive or use visual information, while a score of 100 represents maximum capacity. A score of 80 is also assigned for standard performance. Each criterion is assigned a weight according to the task performed in the virtual environment.

We then carry out a series of experiments to complete the information available in the literature, to establish the criteria. Contrast and latency are of particular interest. The first experiment consists in the transposition and validation of a visual performance model in Virtual Reality. In the second, we compare the effects of latency on a given task, between a CAVE simulator and a HMD.

The thesis proposes a number of theoretical, methodological and experimental results. In virtual immersion conditions, contrast and luminance criteria are important for visual perception. Experiments show that a model adapted to virtual immersion conditions is necessary. Furthermore, we show that thresholds of latency influence on performance seem to exist. The relevance of correlations between simulator performance, presence and cybersickness is also checked, depending on the latency and the immersive system.

# **REMERCIEMENTS**

Tarifs pour apparaître dans les remerciements :

- Citation générale dans un groupe : gratuit.
- Citation nominale : 5 euros.
- Paragraphe personnel : 15 euros.
- Photo personnelle avec des cœurs : 106 euros.

# TABLE DES MATIÈRES

<b>Table des figures</b>	<b>viii</b>
<b>Liste des tableaux</b>	<b>x</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Indices de vision</b>	<b>3</b>
1.1 Contraste & luminosité . . . . .	3
1.1.1 Fonction de sensibilité au contraste (CSF) . . . . .	4
1.1.2 Performance visuelle relative . . . . .	5
1.2 Images par seconde . . . . .	6
1.2.1 Minimum de fonctionnement . . . . .	6
1.2.2 Phénomène de scintillement (flickering) . . . . .	6
1.2.3 Maximum de fonctionnement . . . . .	7
1.2.4 Fonction de notation du critère . . . . .	8
1.3 Couleurs . . . . .	8

1.3.1	Dénombrement des couleurs visibles . . . . .	8
1.3.2	Les espaces colorimétriques . . . . .	9
1.3.3	Indice de rendu des couleurs (IRC) . . . . .	10
1.3.4	Fonction de notation du critère . . . . .	11
1.4	Champ de vision . . . . .	11
1.4.1	Axe horizontal . . . . .	12
1.4.2	Axe vertical . . . . .	13
1.4.3	Pondération . . . . .	13
1.4.4	Fonction de notation du critère . . . . .	14
1.5	Acuité monoscopique . . . . .	14
1.5.1	Première approche . . . . .	14
1.5.2	Deuxième approche . . . . .	15
1.5.3	Fonction de notation du critère . . . . .	16
1.6	Acuité stéréoscopique . . . . .	16
1.6.1	Première modélisation . . . . .	16
1.6.2	Fonction de notation du critère . . . . .	19
<b>2</b>	<b>Indices d'immersion</b>	<b>21</b>
2.1	Latence . . . . .	21
2.2	Champ de regard . . . . .	22
2.2.1	Champ horizontal & champ vertical . . . . .	22
2.2.2	Pondération . . . . .	22
2.3	Stéréoscopie . . . . .	23
2.3.1	Fonctionnement . . . . .	23
2.3.2	Fonction de notation du critère . . . . .	24
2.4	Tracking . . . . .	25

2.5 Uniformité . . . . .	26
2.6 Orientation des caméras . . . . .	27
2.6.1 Mimétisme du fonctionnement oculaire . . . . .	27
2.6.2 Fonction de notation du critère . . . . .	29
<b>Discussion générale et perspectives</b>	<b>30</b>
<b>Bibliographie</b>	<b>37</b>

# TABLE DES FIGURES

1.1 Fonctions de sensibilité au contraste pour les domaines photopiques, scotopiques et mésopiques. . . . .	4
1.2 Exemple de performance visuelle pour deux tailles de cible différentes . . . . .	6
1.3 Illustration du « mouvement beta » et de l’« effet phi » . . . . .	7
1.4 Tracé de la fonction de notation du critère « images par seconde » . . . . .	9
1.5 Tracé de différents gamuts sur l'espace colorimétrique CIE RGB 1931. . . . .	10
1.6 Tracé de la fonction de notation du critère « couleurs » . . . . .	11
1.7 Répartition des zones visuelles sur l'axe horizontal du champ de vision. . . . .	12
1.8 Répartition des zones visuelles sur l'axe vertical du champ de vision. . . . .	13
1.9 Tracé de la sous-fonction horizontale de notation du critère « champ de vision » (ligne continue) et de la sous-fonction verticale (pointillés). . . . .	14
1.10 Classification ICD-9-CM . . . . .	15
1.11 Tracé de la fonction de notation du critère « acuité monoscopique » . . . . .	17
1.12 Variation de la constante du modèle d'acuité stéréoscopique en fonction de la distance interoculaire. . . . .	18
1.13 Variation de la distance interoculaire sur la population. Image tirée de [Dodge- son, 2004] . . . . .	18

1.14 Tracé de la fonction de notation du critère « acuité stéréoscopique » ( $\Delta\nu_{min} = 8 \text{ arcsecs}$ ) . . . . .	20
2.1 Recouvrement des demi-champs de regard (temporal et nasal). . . . .	23
2.2 Lunettes de stéréoscopie. . . . .	24
2.3 Illustration d'une notation du critère « Stéréoscopie » en fonction de la technologie . . . . .	25
2.4 « Body » de tracking monté sur des lunettes de stéréoscopie & ensemble caméra-émetteur infrarouge pour le tracking. . . . .	26
2.5 Exemple de système de tracking de regard portatif. . . . .	28
2.6 Le carré est vu comme deux rectangles dans une configuration « caméras à convergence nulle » et comme deux trapèzes lorsque les caméras sont convergentes. Il faut donc retordre artificiellement les trapèzes pour qu'ils correspondent à la projection qui se fait sur une surface rectangulaire (image tirée de [Aurat, 2016]). . . . .	28
I Rappel de la modélisation du score de réalisme. . . . .	31
II Score de réalisme pour les quatre systèmes testés. . . . .	33
III Diagramme radar des critères du score pour IRIS. . . . .	34
IV Diagramme radar des critères du score pour P3I. . . . .	34
V Diagramme radar des critères du score pour le casque VIVE. . . . .	35
VI Diagramme radar des critères du score pour le casque Hololens. . . . .	35

# LISTE DES TABLEAUX

1.1 Gamuts Coverage of 1931 Color Space . . . . .	10
1.2 Acuity of the eye, [Gross <i>et al.</i> , 2008] . . . . .	16

# INTRODUCTION GÉNÉRALE

LES travaux de la thèse « Immersion visuelle hyper-réaliste et multi-sensorielle 3D », dont les résultats sont présentés ici, ont été réalisés dans le cadre d'un contrat CIFRE établi entre les Arts et Métiers et Renault, au sein du laboratoire commun LIV. Ils sont le fruit d'un travail de trois ans, réalisé en grande partie au Technocentre Renault à Guyancourt, sous l'encadrement de Stéphane Régnier, expert maquette virtuelle pour l'ingénierie et le design ; Fédéric Mérianne, professeur aux Arts et Métiers et directeur de l'Institut Image ; et Andras Kemeny, expert-leader simulation immersive et réalité virtuelle.

Renault, fabricant français de voitures bien connu, investit de plus en plus de moyens dans la réalité virtuelle, et même depuis peu, dans la réalité augmentée. Ces technologies viennent compléter les outils de développement des nouveaux véhicules. Ils permettent de tester, d'afficher ou de rouler sans avoir besoin de passer par des prototypes physiques et dans les conditions que l'on souhaite (brouillard, nuit, pluie, ...). Ces procédés permettent de nombreux gains de temps et d'argent, composantes vitales pour l'entreprise.

L'Institut Image, institut de recherche rattaché au campus Arts et Métiers de Cluny, a quant à lui pour objectifs le développement d'outils et de méthodes pour l'immersion virtuelle, au service de l'ingénieur. Ses missions sont celles de la formation, de la recherche et de l'innovation. Dans le but de collaborer à des projets industriels concrets, Renault et les Arts et Métiers, via l'Institut Image, créent en 2011 un laboratoire commun : le LIV (Laboratoire d'Immersion Virtuelle). C'est dans le cadre de ce partenariat qu'à lieu la présente thèse.

Notre étude est née de la volonté de Renault d'offrir à ses équipes des moyens sûrs, permettant la prise de décisions parfois stratégiques, sans craindre un biais causé par la technique. Les outils de travail doivent proposer des situations les plus fidèles à la réalité possible, c'est à dire « réalistes ». L'objectif de la thèse est donc de formuler ce qu'est le réalisme et de quel manière il peut être atteint ou réalisé. Si le sujet peut sembler un

peu large initialement, nous concentrons nos efforts sur la partie matérielle des moyens immersifs, avec en priorité ceux disponibles dans le laboratoire commun. Renault s'associe donc avec un partenaire universitaire et c'est ici qu'entre en jeu la spécificité d'un contrat de type CIFRE : les travaux ont une vocation universitaire, c'est à dire qu'ils sont appelés à être communiqués, publiés ; mais ils possèdent également une réelle vocation applicative, avec des attendus concrets de la part de l'entreprise, qui peuvent être confidentiels. Il est alors dévolu au doctorant de trouver le juste équilibre entre les attentes des deux parties.

Le manuscrit est présenté en cinq grandes parties : deux parties à dominante théorique, une partie modélisation puis deux parties expérimentales.

La première partie présente la synthèse de l'état de l'Art qui a été effectué afin de comprendre en profondeur les différents mécanismes et modélisations qui prennent part au fonctionnement de l'œil, de la vision, et des premières phases de la perception visuelle.

La deuxième partie s'intéresse spécifiquement aux modèles de vision et retrace la première approche qui a été envisagée, puis abandonnée, pour la thèse.

Ensuite, la troisième partie présente la proposition de modèle au cœur de nos travaux : ses objectifs, sa mise en place et ses différentes composantes.

D'abord très théorique, la thèse se dirige ensuite vers deux segments plus expérimentaux. Dans la quatrième partie, on décrit la traduction et l'implémentation d'un modèle de performance visuelle, en réalité virtuelle.

Enfin, dans la cinquième et dernière partie, on s'intéresse à la latence, ses effets sur la performance et l'expérience utilisateur, en comparant deux moyens immersifs.

## CHAPITRE

# 1

## INDICES DE VISION

Nous allons maintenant présenter les différents indices placés dans la catégorie « vision », c'est à dire les indices fournis par l'environnement immersif qui participent uniquement à la vision et pas à l'immersion. Ces indices sont les caractéristiques perçues par un œil fixe et immobile à un instant  $t$  dans le simulateur. On présente d'abord les valeurs clefs trouvables dans la littérature puis leur association aux valeurs clefs (0, 80 et 100) du modèle, la génération de l'équation permettant d'attribuer la note entre 0 et 100 au critère et, quand c'est possible, le tracé de cette fonction.

### 1.1 Contraste & luminosité

Le contraste est théoriquement défini comme la différence en luminosité (ou en couleur) entre les parties claires et les parties sombres d'une image ou d'un objet. Le contraste est un élément très important du système de vision humain, notamment car ce dernier est plus sensible au contraste entre deux niveaux de lumière qu'à un niveau absolu. Comme il peut être fait classiquement en psychophysique, ce stimulus peut être divisé en deux parties : sa magnitude (la valeur brute) et sa résolution (son « pas » d'évolution).

On a pu voir les différentes approches théoriques pour le définir dans la première partie de ce manuscrit. Néanmoins il existe une d'autres approches de la gestion du contraste et de la luminance comme par exemple la fonction de sensibilité ou des modèles de performance visuelle. C'est vers cette piste qu'on s'oriente initialement.

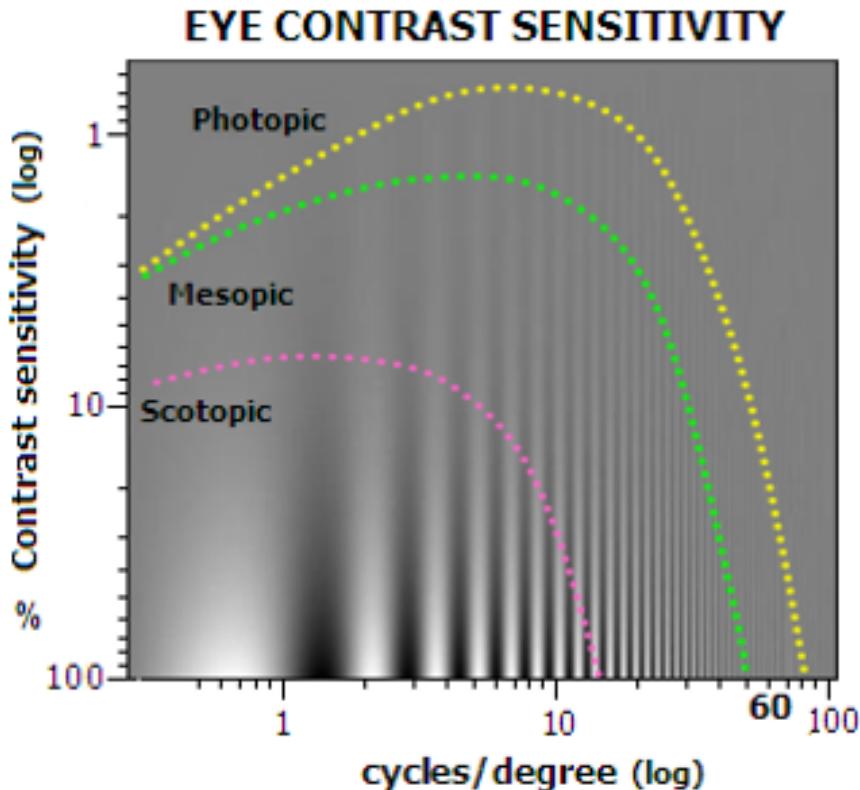


FIGURE 1.1 – Fonctions de sensibilité au contraste pour les domaines photopiques, scotopiques et mésopiques.

### 1.1.1 Fonction de sensibilité au contraste (CSF)

On a déjà donné une brève description des fonctions de sensibilité (CSF - Contrast Sensitivity Functions en anglais) à la fin du chapitre sur le fonctionnement de la vision. Pour rappel, les CSF sont des courbes traçant le seuil à partir duquel l'œil humain est capable de distinguer une différence entre deux niveaux de gris. La Fig. 1.1 en est un exemple pour les trois domaines de luminosité : photopique (jour), scotopique (nuit) et mésopique (entre-deux). Tout ce qui se trouve sous la courbe est visible.

Bien que très intéressantes, ces courbes ne sont pas suffisantes car elles ne donnent qu'un seuil en dessous duquel on voit et au dessus duquel on ne voit pas. On ne peut en tirer aucune information sur la facilité ou la qualité de vision en dessous du seuil. Que l'on soit très proche ou très éloigné du seuil revient au même : on est dans la catégorie « voit ». On s'intéresse donc à d'autres approches sur le contraste qui nous permettrait de détailler ce qu'il se passe dans la partie visible du contraste. On va donc maintenant présenter le concept de performance visuelle.

### 1.1.2 Performance visuelle relative

La performance visuelle relative (RVP - Relative Visual Performance en anglais) est définie comme la capacité du système visuel humain à réaliser une tâche donnée, capacité transposée dans une valeur comprise entre 0 et 1. Les paramètres d'entrée du modèle sont la luminance et le contraste pour une valeur de sortie unique et sous la forme d'une valeur (comprise entre 0 et 1, donc). De nombreux modèles de performance visuelle ont été développés, notamment dans les années 80, mais deux d'entre eux peuvent être particulièrement mis en avant : le modèle de la CIE (Commission International de l'Eclairage) [Blackwell, 1981] et le modèle de Rea [Rea, 1986].

#### Modèle de la CIE

Le modèle de la CIE a été construit sur la base d'un grand nombre de résultats expérimentaux issus de nombreux chercheurs différents. Leur but était d'arriver à un outil pour aider à choisir les meilleures conditions d'illumination pour les ateliers d'usine de manière à ce que les ouvriers travaillent au rendement maximal. En plus de la luminance et du contraste, le modèle nécessite également l'âge et le « niveau de demande de la tâche » comme paramètres d'entrée.

#### Modèles de Rea et Ouellette

Le modèle de Rea est plus simple mais ses auteurs le déclarent plus précis que celui de la CIE. Leur modèle ne prend en entrée que la luminance de la tâche et la luminance du background de la tâche. Mark S. Rea présente une première version de son modèle en 1986 [Rea, 1986] puis une version raffinée en 1991 [Rea & Ouellette, 1991]. Dans ce dernier modèle, trois paramètres de plus ont été rajoutés : l'âge, la taille de la tâche à réaliser et la luminance d'adaptation (quantité de lumière à laquelle les yeux ont adapté la taille de leur pupille). Ce modèle permet de tracer les graphes visibles en Fig. 1.2. Ces courbes représentent la performance visuelle en fonction de la luminosité et du contraste pour une taille d'objet vu donnée : la courbe de gauche, plus prononcée, concerne un objet plus petit que celui pour la courbe de droite, moins prononcée.

#### Vers une vérification expérimentale

Néanmoins, ces deux modèles n'ont pas été pensés pour la Réalité Virtuelle et pourraient ne pas s'inscrire dans une utilisation directe dans notre modèle de score de réalisme, en lieu et place du critère de luminance et de contraste. Une expérimentation a été réalisée pour confirmer ou infirmer cette hypothèse, notamment à propos du modèle de Rea (*cf.* partie suivante). L'objectif serait de mettre à l'échelle entre 0 et 100 les résultats entre 0 et 1 du modèle de performance visuelle. Il est à noter que dans son modèle la CIE définit la valeur de 0.8 (80 ramené à notre échelle) comme la performance standard, ce qui colle parfaitement avec nos bases de modélisation.

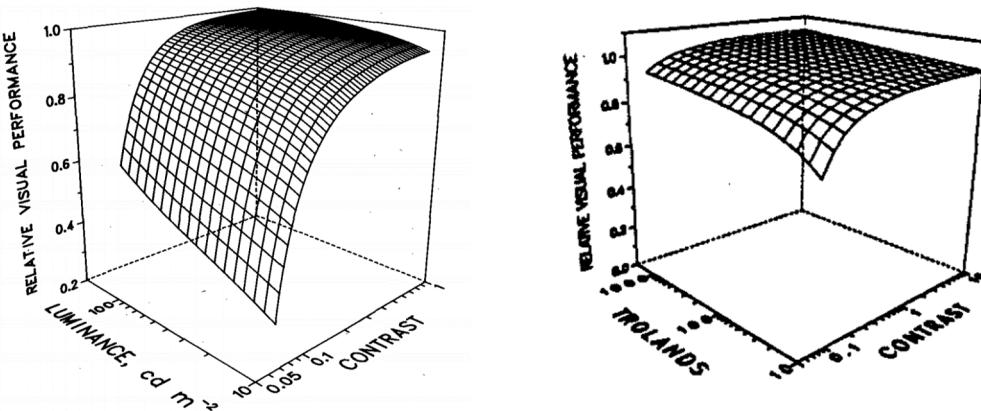


FIGURE 1.2 – Exemple de performance visuelle pour deux tailles de cible différentes

## 1.2 Images par seconde

Attribuer des valeurs spécifiques, discrètes, aux phénomènes qui composent la vision peut sembler peu naturel tant le processus de vision est continu [Bear *et al.*, 2007]. Néanmoins, c'est un exercice auquel on se risque et il existe un certain nombre d'effets notables qui n'apparaissent qu'à certains nombres d'images par seconde (frame rate).

### 1.2.1 Minimum de fonctionnement

Premièrement, la perception du mouvement n'est pas basée (comme cela a été suspecté pendant longtemps) sur la persistance rétinienne mais sur deux illusions perceptuelles : l'effet *phi* et le mouvement *bêta* [Nichols & Lederman, 1980], voir Fig. 1.3. La persistance rétinienne est un phénomène passif laissant une image rémanente sur la rétine pendant un court instant. L'effet *phi* est une illusion de mouvement, cette fois active dans le cerveau et le traitement des images, pour un système en boucle fermée, comme les guirlandes de noël. Le mouvement *bêta* est lui aussi une illusion d'optique active mais pour les systèmes ouverts. Ces effets commencent à partir de 16 images par seconde, en dessous de cette valeur, aucun mouvement n'est perçu, seulement une suite distincte d'images.

### 1.2.2 Phénomène de scintillement (flickering)

Deuxièmement, un autre grand phénomène induit par la vision en informatique est le « flickering », le clignotement/scintillement de l'écran. Ce phénomène intervient quand le frame rate est trop lent et que l'œil devient capable de percevoir un effet de diminution/changement de luminosité (fading) entre les images. Driscoll [Driscoll *et al.*, 1978], en s'appuyant sur les travaux de Landis [Landis, 1954] et de de Lange [de Lange Dzn, 1958a,b], a travaillé sur la détermination de la fréquence critique de clignotement pour l'œil, c'est à dire la fréquence au delà de laquelle le cerveau ne perçoit plus de clignotement. La fréquence critique de clignotement semble être basée sur un ratio appelé « ratio d'ondulation » ou « ratio ondulatoire » (*ripple ratio* en anglais) qui se calcule de la manière

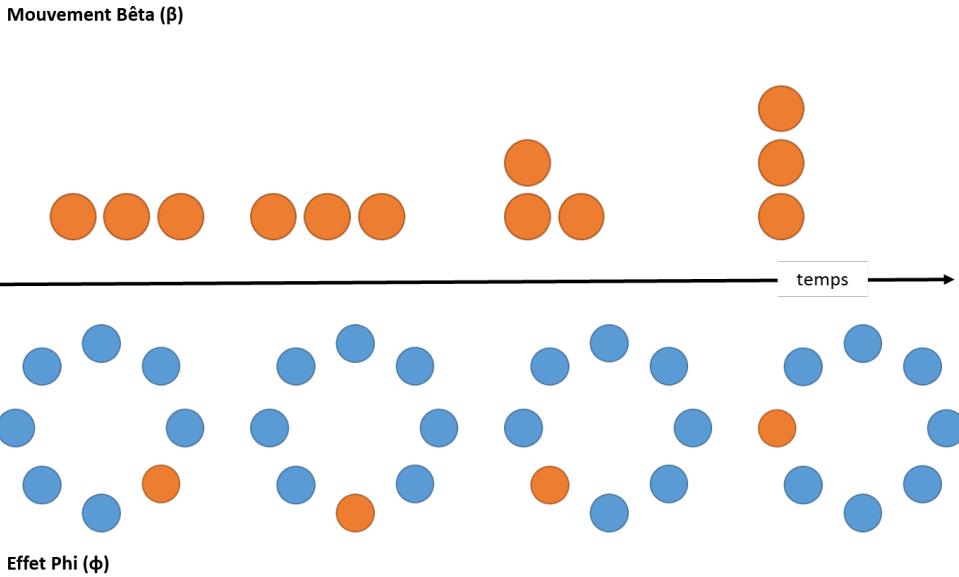


FIGURE 1.3 – Illustration du « mouvement beta » et de l'« effet phi »

suivante (Eq. 1.1) :

$$r = \frac{\text{Amplitude de l'harmonique fondamentale}}{\text{Luminance moyenne}} \quad (1.1)$$

En utilisant les courbes établies par Driscoll, les courbes de la fonction de transfert de modulation temporelle, quel que soit le ratio ondulatoire, c'est à dire quelle que soit la nature de l'onde lumineuse, la fréquence critique de clignotement, quelle que soit la luminance, est de 70 Hz.

### 1.2.3 Maximum de fonctionnement

#### L'hypothèse des deux voies

D'autres valeurs viennent des voies dorsale et ventrale. Comme pour les fonctions de sensibilité au contraste, la théorie des deux voies a déjà été abordée dans la première partie, chapitre sur le fonctionnement de la vision. Pour rappel, la théorie de deux voies est le postulat principal pour expliquer la manière dont le cerveau traite les informations visuelles arrivant des nerfs optiques. Une fois dans le cerveau, l'information est divisée en deux boucles de traitement [D'hondt, 2011; Ungerleider & Mishkin, 1982]. La première, la voie dorsale ou pariétale, est la boucle « où » et s'occupe d'extraire la direction et le mouvement du flux d'image qui arrivent. L'autre voie, la voie ventrale ou temporale, dite boucle du « quoi », s'occupe d'extraire la forme, la couleur et la texture. Les deux voies fonctionnent supposément à des fréquences autour de, respectivement, 200 Hz et 25 Hz, ce qui est corrélé par le fait que l'une est rapide et l'autre lente [D'hondt, 2011], par le nombre de cortex différents par lesquels ces voies passent [D'hondt, 2011] et par les latences propres

de chaque cortex [Bullier, 2001]; cependant, aucunes valeurs précises n'ont été scientifiquement établies.

## Les pilotes de l'US Air Force

Parallèlement, l'US Air Force aurait conduit des expériences sur ses pilotes de chasse pendant lesquelles ceux-ci étaient capables de reconnaître un modèle d'avion sur des images flashées à 1/220ème de seconde<sup>1</sup>. Même si cela ne constitue pas une preuve scientifique en tant que telle, cela vient renforcer l'hypothèse des deux voies sur la fréquence maximale associable au traitement visuel. A la différence du minimum de fonctionnement où des phénomènes sont clairement établis avec des fréquences de fonctionnement claires, on ne peut ici fonctionner que sur des approximations.

### 1.2.4 Fonction de notation du critère

Ainsi, en utilisant les trois valeurs clefs précédemment présentées : le nom minimal de frame rate pour percevoir le mouvement (16 Hz), la fréquence critique de clignotement (70 Hz) et la fréquence supposée de la voie dorsale (200 Hz), on peut assembler un modèle mathématique et en tirer une courbe. Dans le cas de ce critère, avec  $f$  le nombre d'images par seconde du système de visualisation, on présente l'équation suivante (Eq. 1.2), dont le tracé est donné en Fig. 1.4 :

$$F_{FPS}(f) = \begin{cases} 0 & f < 16 \\ 126.5 - \frac{367.1}{\sqrt{f-7.6}} & \text{else} \\ 100 & f > 200 \end{cases} \quad (1.2)$$

Finalement, il doit être noté que toutes ces valeurs clefs sont définies pour un œil unique et donc qu'elles doivent être multipliées par deux dans le cas d'un usage en stéréoscopie pour une imagerie 3D. De plus, ces fréquences sont les nombres d'images par seconde effectifs, c'est à dire ceux qui sont réellement affichés par l'écran et non pas les nombres d'images par seconde du hardware (en sortie de carte graphique par exemple) qui sont en général plus élevés à cause de principe de Shannon.

## 1.3 Couleurs

### 1.3.1 Dénombrement des couleurs visibles

La première approche serait d'essayer d'estimer le nombre de couleurs réellement visibles (discernables) par un œil humain. Cette entreprise a été tentée un certain nombre de fois

---

1. Human Eye Frames Per Second. Dans *AMO.net America's Multimedia Online*. Vu sur <http://amo.net/nt/02-21-01fps.html>

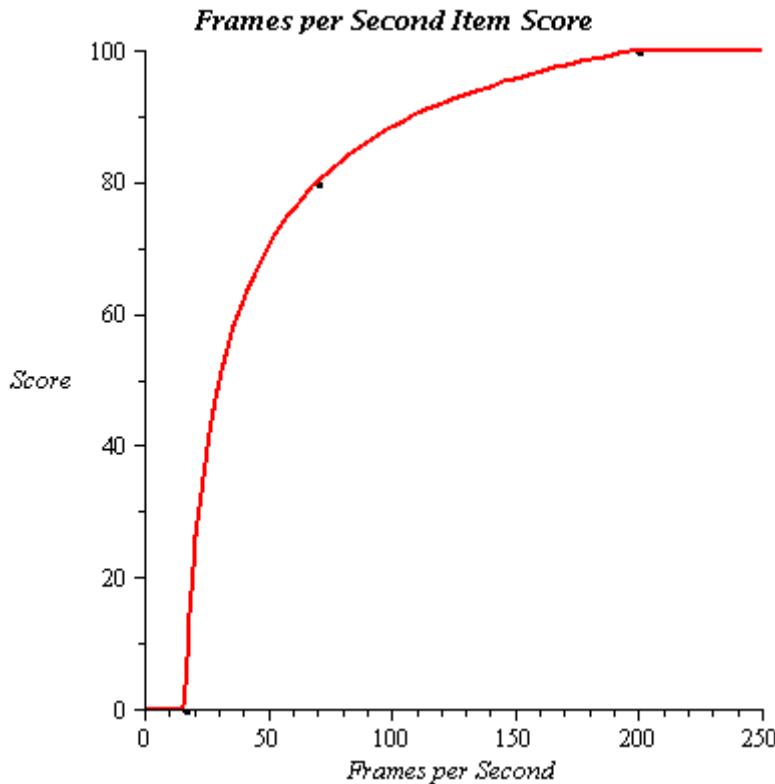


FIGURE 1.4 – Tracé de la fonction de notation du critère « images par seconde »

[Kuehni, 2015; Linhares *et al.*, 2008; Perales *et al.*, 2008; Pointer, 1980; Pointer & Attridge, 1998; Wen, 2006] sans jamais faire preuve d'un succès éclatant : les estimations varient entre 100.000 et 10 millions de couleurs visibles.

Les estimations se font mathématiquement : soit en utilisant les équations de différentiation des couleurs, soit en recalculant des observateurs réels pour les comparer à l'observateur idéal établi par la CIE (*cf.* première partie et le chapitre sur la couleur).

### 1.3.2 Les espaces colorimétriques

A défaut d'avoir une estimation validée du nombre de couleurs perceptibles et de pouvoir proposer une méthode pour calculer facilement ce même nombre dans un simulateur pour ensuite les comparer, on se tourne vers les espaces colorimétriques.

En 1931, la Commission Internationale de l'Eclairage (CIE) a établi la définition de l'espace colorimétrique RGB. Cet espace de couleur représente l'ensemble des couleurs (non dénombrées) qui peuvent être vues par un observateur normal possédant les trois types de cônes. Cependant, il n'existe encore aucune technique pour afficher 100% des couleurs théoriquement incluses dans l'espace RGB. Chaque système utilise une fraction de cet espace de couleur qu'on appelle « gamut » et qui dépend de trois (ou plus suivant les espaces) couleurs primaires choisies spécifiquement par la norme. Dans la table suivante (Table. 1.1), on présente un certain nombre de normes avec la proportion d'espace colorimétrique

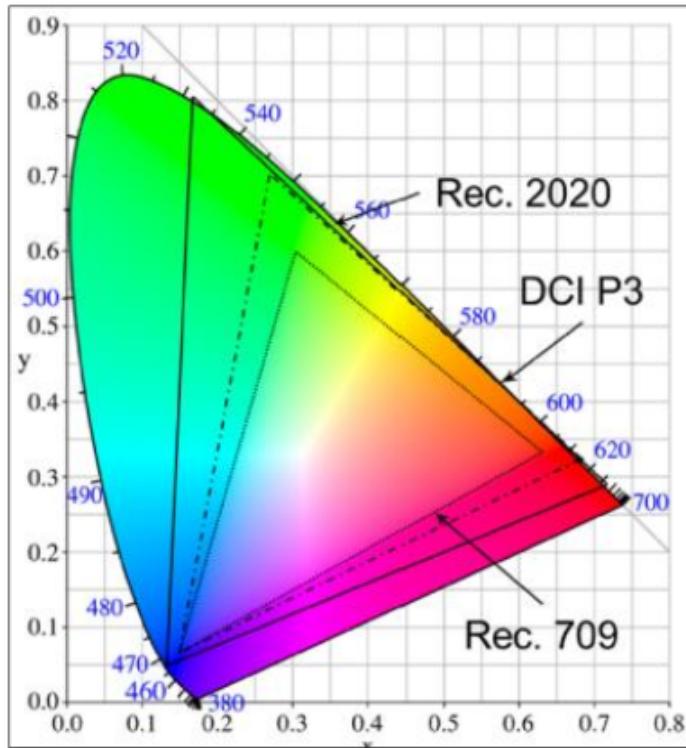


FIGURE 1.5 – Tracé de différents gamuts sur l'espace colorimétrique CIE RGB 1931.

couverte par leur gamut. Il est à noter que dans le cas de la norme ProPhoto RGB, 13% des couleurs atteignables sont en fait des couleurs imaginaires à cause de couleurs primaires qui sont prises hors de l'espace RGB de 1931. Empiriquement, l'acceptation commence à partir d'Adobe RGB. Fig. 1.5

TABLE 1.1 – 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%

### 1.3.3 Indice de rendu des couleurs (IRC)

Pas mal mais caractérise la lumières émise par le projo plutôt que les couleurs affichées sur l'écran (si l'écran n'est pas parfaitement blanc + influence des autres écrans que se passe-t-il ?)



FIGURE 1.6 – Tracé de la fonction de notation du critère « couleurs »

### 1.3.4 Fonction de notation du critère

Notre proposition de notation pour le critère de couleur est réduite à une fonction linéaire entre le score et le pourcentage de couverture de l'espace RGB 1931 auquel le système peut prétendre (Fig. 1.6). Ce pourcentage vient directement de-s espace-s colorimétrique-s impliqué-s dans la chaîne de rendu et d'affichage. La fonction de score pour ce critère est donc, avec  $c$  le pourcentage de couverture de RBB 1931 par le gamut :

$$F_{color}(c) = \begin{cases} c & \\ 100 & c > 100 \end{cases} \quad (1.3)$$

Il pourrait être intéressant de réaliser une expérimentation sur les besoins en couleurs (à différencier de l'appréciation du nombre de couleurs) en comparant par exemple des populations novices avec des populations expertes. Néanmoins faute de temps et de moyens (aujourd'hui encore aucun appareil commercial n'est capable d'afficher l'intégralité de l'espace RGB 1931) ce ne sera pas réalisable le temps de la thèse.

## 1.4 Champ de vision

Le champ de vision (FOV - Field of View en anglais) est défini comme la portion d'espace qu'une personne peut voir à un instant  $t$ , sans bouger la tête. Il ne doit pas être confondu

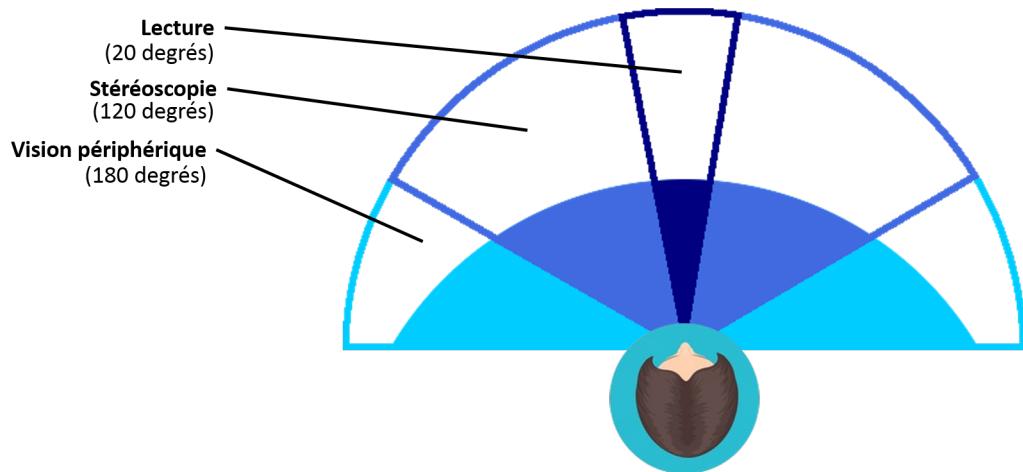


FIGURE 1.7 – Répartition des zones visuelles sur l’axe horizontal du champ de vision.

avec le champ de regard (FOR - Field of Regard en anglais) qui est la portion d'espace totale que l'on peut voir au cours du temps lorsque les mouvements de la tête et des yeux sont pris en compte.

Le champ de vision se décompose en deux orientations : l'axe vertical et l'axe horizontal. On propose deux sets de valeurs, un par axe, pour la notation. Les deux axes seront ensuite pondérés l'un par rapport à l'autre. On fait l'hypothèse que cette pondération se fait par rapport à leur taille respective. La valeur du champ de vision que l'on mesure pour introduire dans le score doit être prise pour la position standard du sujet dans l'environnement immersif (que ce soit la position du corps ou l'orientation de la tête).

### 1.4.1 Axe horizontal

Certaines des valeurs les plus communes admises pour le champ de vision horizontal sont recensées chez [Devisme, 2004]. Sur l'angle d'azimut (horizontal), avec les deux yeux, un être humain normal peut voir sur 170 à 190 degrés. A l'intérieur de cet angle d'azimut, seuls 120 degrés (centrés) permettent de voir binoculairement. La vision binoculaire est possible grâce à la superposition des portions d'espace vues par chaque œil simultanément. L'acuité maximale est atteinte dans la zone foveale soit entre 3 et 5 degrés de portion d'espace, au centre de la vision. La lecture n'est possible que dans un angle de 20 degrés tandis que la reconnaissance des formes est possible jusqu'à 40 degrés et la reconnaissance des couleurs jusqu'à 60 degrés. Ces différentes valeurs sont résumées dans la Fig. 1.7. L'équation propre à l'axe horizontal est, avec  $h$  la valeur en degrés du champ de vision horizontal (H-FOV) calculé dans le système immersif (Eq. 1.4) :

$$F_h(h) = \begin{cases} 0 & h < 20 \\ 19.6\sqrt{h} - 0.5 \cdot h - 78.3 & \text{otherwise} \\ 100 & h > 180 \end{cases} \quad (1.4)$$

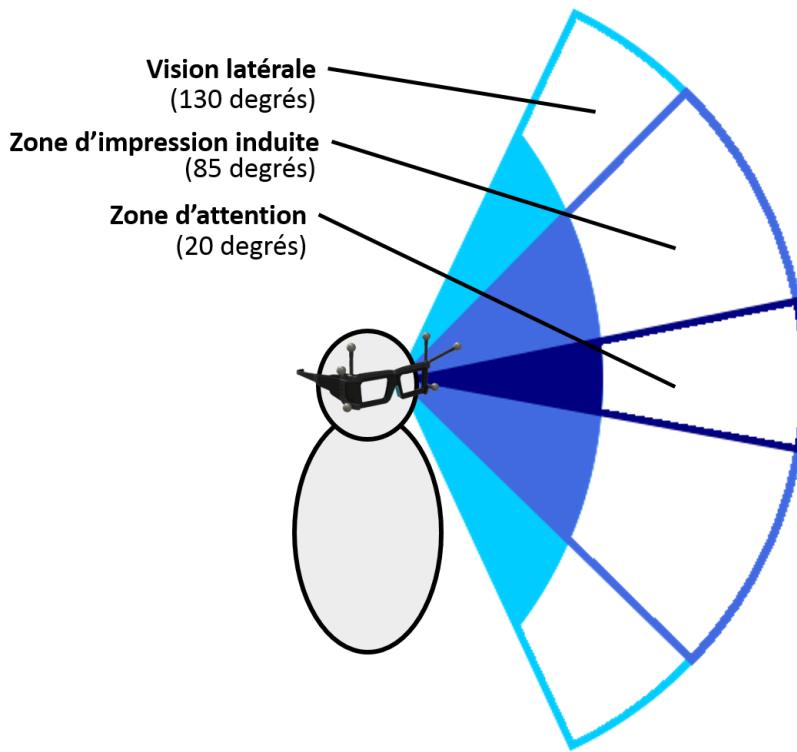


FIGURE 1.8 – Répartition des zones visuelles sur l’axe vertical du champ de vision.

### 1.4.2 Axe vertical

On retient trois valeurs caractéristiques parmi les nombreuses qu’on peut trouver dans la littérature pour le champ de vision vertical. La première correspond à la totalité de la vision latérale sur l’axe vertical : 130 degrés. On retient aussi les angles d’impression induite (85 degrés) et de vigilance (20 degrés) [Langlois, 2013] (Fig. 1.8). L’équation propre à l’axe vertical est, avec  $v$  la valeur en degrés du champ de vision vertical (V-FOV) (Eq. 1.5) :

$$F_v(v) = \begin{cases} 0 & v < 20 \\ 32.0\sqrt{v} - 1.1 \cdot v - 121.1 & \text{otherwise} \\ 100 & v > 130 \end{cases} \quad (1.5)$$

### 1.4.3 Pondération

On pose notre propre hypothèse d’égale importance des axes verticaux et horizontaux dans le champ de vision. On pondère alors les deux sous-critères en fonction de leur taille maximale respective (180 degrés pour l’axe horizontal et 130 pour l’axe vertical) (Eq. 1.6) :

$$\begin{cases} k_h = \frac{180}{180+130} = 0.58 \\ k_v = 1 - k_h = 0.42 \end{cases} \quad (1.6)$$

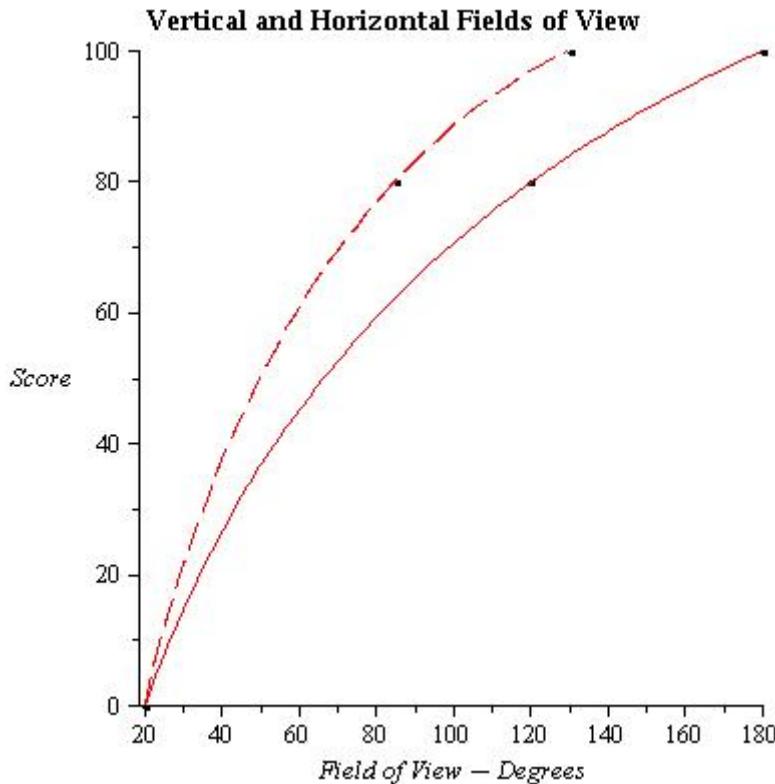


FIGURE 1.9 – Tracé de la sous-fonction horizontale de notation du critère « champ de vision » (ligne continue) et de la sous-fonction verticale (pointillés).

#### 1.4.4 Fonction de notation du critère

Au final, la fonction de notation du critère de champ de vision est l'agrégation des précédentes équations via une pondération et s'apparente comme suit (Eq. 1.7) :

$$F_{FOV}(h, v) = k_h \cdot F_h(h) + k_v \cdot F_v(v) \quad (1.7)$$

### 1.5 Acuité monoscopique

#### 1.5.1 Première approche

Notre première approche était de se raccrocher à un tableau établi par la « classification internationale des maladies » (CIM), publié par l'Organisation Mondiale de la Santé (OMS). L'objectif est de permettre l'analyse systématique des maladies et autres affections du corps. On s'intéresse ici à la partie relative aux trouble de la vision et plus particulièrement à l'acuité. On utilise la version cliniquement modifiée de la 9ème révision de cette classification, ICD-9-CM, sortie en 1975. La 10ème révision, ICD-10 ne faisant pas encore l'unanimité et ne rajoutant que d'autres entrées sans modifier celle ci.

RANGES (ICD-9-CM)		EQUIVALENT NOTATIONS		TRUE SNELLEN FRACTIONS (numerator = test distance)					Magnification Requirement	Visual Acuity Score (letter count)	
		Decimal	US	6.3 m	6 m (Britain)	5 m (Europe)	4 m (ETDRS)	1 m (Low Vision)	MAR (1/V)	Log MAR	
(Near-) Normal Vision	Range of Normal Vision	1.6	20/12.5	6.3/4	6/3.8	5/3.2	4/2.5	1/0.63	0.63	-0.2	110
	Mild Vision Loss	1.25	20/16	6.3/5	6/4.8	5/4	4/3	1/0.8	0.8	-0.1	105
Low Vision	Moderate Vision Loss	1.0	<b>20/20</b>	<b>6.3/6.3</b>	<b>6/6</b>	<b>5/5</b>	<b>4/4</b>	<b>1/1</b>	1.0	0	<b>100</b>
	Severe Vision Loss	0.8	20/25	6.3/8	6/7.5	5/6.3	4/5	1/1.25	1.25	+0.1	95
(Near-) Blindness	Profound Vision Loss	0.63	20/32	6.3/10	6/9.5	5/8	4/6.3	1/1.6	1.6	0.2	90
	Near-Blindness	0.5	20/40	6.3/12.5	6/12	5/10	4/8	1/2	2.0	0.3	85
Blindness	Blindness	0.4	20/50	6.3/16	6/15	5/12.5	4/10	1/2.5	2.5	0.4	80
		0.32	20/63	6.3/20	6/19	5/16	4/12.5	1/3.2	3.2	0.5	75
		0.25	20/80	6.3/25	6/24	5/20	4/16	1/4	4	0.6	70
		0.20	20/100	6.3/32	6/30	5/25	4/20	1/5	5	0.7	65
		0.16	20/125	6.3/40	6/38	5/32	4/25	1/6.3	6.3	0.8	60
		0.125	20/160	6.3/50	6/48	5/40	4/32	1/8	8	0.9	55
		<b>0.10</b>	<b>20/200</b>	<b>6.3/63</b>	<b>6/60</b>	<b>5/50</b>	<b>4/40</b>	<b>1/10</b>	<b>10</b>	<b>+1.0</b>	<b>50</b>
		0.08	20/250	6.3/80	6/75	5/63	4/50	1/12.5	12.5	1.1	45
		0.063	20/320	6.3/100	6/95	5/80	4/63	1/16	16	1.2	40
		0.05	20/400	6.3/125	6/120	5/100	4/80	1/20	20	1.3	35
		0.04	20/500	6.3/160	6/150	5/125	4/100	1/25	25	1.4	30
		0.03	20/630	6.3/200	6/190	5/160	4/125	1/32	32	1.5	25
		0.025	20/800	6.3/250	6/240	5/200	4/160	1/40	40	1.6	20
		0.02	0/1000	6.3/320	6/300	5/250	4/200	1/50	50	1.7	15
		0.016	20/1250	6.3/400	6/380	5/320	4/250	1/63	63	1.8	10
		0.0125	20/1600	6.3/500	6/480	5/400	4/320	1/80	80	1.9	5
		<b>0.01</b>	<b>20/2000</b>	<b>6.3/630</b>	<b>6/600</b>	<b>5/500</b>	<b>4/400</b>	<b>1/100</b>	<b>100</b>	<b>+2.0</b>	<b>0</b>
		---	---	---	---	---	---	---	---	---	---
		No Light Perception (NLP)									

FIGURE 1.10 – Classification ICD-9-CM

La figure 1.10 montre les différentes acuités possibles pour l’œil humain (encadré en bleu), associées avec le type de vision et surtout un score équivalent (encadré en bleu). Ce score est néanmoins peu adapté à notre utilisation, d’abord parce qu’il dépasse la note de 100 mais surtout parce qu’il englobe des acuités bien trop faibles pour être transposées en taille de pixel dans un simulateur. De plus il correspond à des défauts de vision plus qu’à des capacités de vision. On s’oriente alors vers une autre approche.

### 1.5.2 Deuxième approche

L’acuité monoscopique est la précision de l’œil humain, sa résolution ; c’est à dire à quel point la plus petite chose qui peut être vue doit être petit ou bien quelle est le plus petit écart perceptible entre deux motifs. Ce concept peut être directement lié à la taille du pixel sur un écran : plus le pixel est petit plus l’image est précise. Néanmoins, à partir d’une certaine taille, le pixel devient plus petit que la résolution de l’œil et devient alors invisible, et donc, pas forcément très utile.

Habituellement, on estime que le système visuel humain a une acuité monoscopique comprise entre 30 secondes d’arc et 2 minutes d’arc, avec une moyenne à 1 minute d’arc [Fuchs *et al.*, 2003]. Ces valeurs peuvent être, dans des conditions photopiques d’éclairage, raffinées en fonction de la tâche [Gross *et al.*, 2008] (Cf. Table 1.2).

La résolution de Vernier n’est atteignable que dans certains conditions spécifiques et l’acuité stéréoscopique est un sujet à part, traité avec une autre approche, et qui est traité dans la section suivante. De plus, Deering a montré que la plus petite résolution possible

TABLE 1.2 – 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"

sur un écran est de 28 secondes d'arc[Deering, 1998]. C'est pourquoi on ne retient que les valeurs entre 5 minutes d'arc et 30 secondes d'arc.

### 1.5.3 Fonction de notation du critère

L'équation qui est proposée depuis ces valeurs, avec  $\alpha$  l'angle sous lequel le pixel est vu, en minute d'arc (Eq. 1.8, tracé en Fig. 1.11) est la suivante :

$$F_{mono\_acuity}(\alpha) = \begin{cases} 0 & \alpha > 3.5 \\ 128.9 - 68.8\sqrt{\alpha} - \frac{0.1}{\alpha} & \text{else} \\ 100 & \alpha < \frac{1}{6} \end{cases} \quad (1.8)$$

## 1.6 Acuité stéréoscopique

### 1.6.1 Première modélisation

#### Modèle usuel

L'acuité stéréoscopique est la capacité du système visuel humain à percevoir une différence de profondeur entre deux plans, à une distance donnée. C'est une caractéristique qui est largement connue et décrite dans la littérature. Sa description mathématique vient d'une analyse géométrique simple [Fuchs *et al.*, 2003; Gross *et al.*, 2008]. Le modèle globalement accepté est le suivant (Eq. 1.9) :

$$\Delta r = 0.001r^2 \quad (1.9)$$

Avec  $\Delta r$  la différence minimale théorique de différence en profondeur (en millimètres)

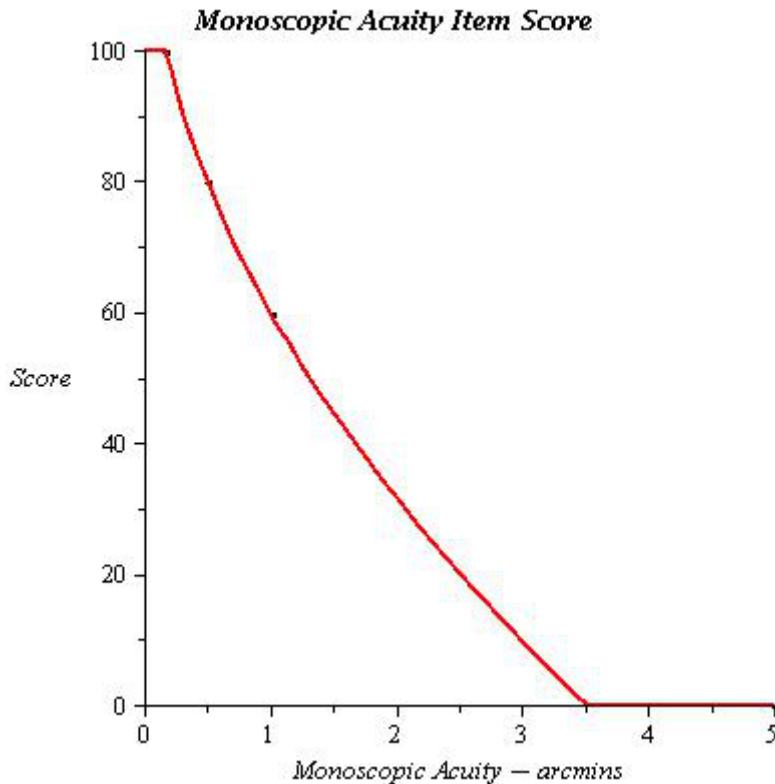


FIGURE 1.11 – Tracé de la fonction de notation du critère « acuité monoscopique »

et  $r$  la distance d'observation en mètres. Le facteur 0.001 est le rapport entre le seuil physiologique de la vision stéréoscopique ( $\Delta\nu_{min}$ ) et la distance inter-pupillaire (DIO).

### Limites et solution d'amélioration

Cependant, ces deux derniers paramètres peuvent varier de manière significative : la distance inter-pupillaire varie de 52 à 78 mm sur toute la population [Dodgson, 2004] (Fig. 1.13) tandis que le seuil physiologique de vision dépend de la luminance, notamment quand celle-ci est à un niveau très bas [Gross *et al.*, 2008]. On peut néanmoins conserver la valeur de 0.001 comme constante dans le modèle car la variation suivant les deux paramètres est assez faible (Fig. 1.12).

De plus, un processus standardisé pour mesurer la valeur de  $\Delta r$  d'un système serait difficile à mettre en place car il faudrait faire passer un test à un certain nombre de sujets puis en extraire le seuil moyen expérimental, ce qui sort du cahier des charges du score de réalisme. En travaillant la relation qui décrit ce même modèle d'acuité stéréoscopique (Eq. 1.10) [Gross *et al.*, 2008], le seuil physiologique de vision apparaît comme une valeur essentielle pour la vision stéréoscopique tout en étant relié à la fois au système visuel humain et au système d'affiche du système, le tout en étant facilement et objectivement mesurable.

$$\Delta\nu_{min} = \frac{d_{IPD} * \Delta r}{r^2} \quad (1.10)$$

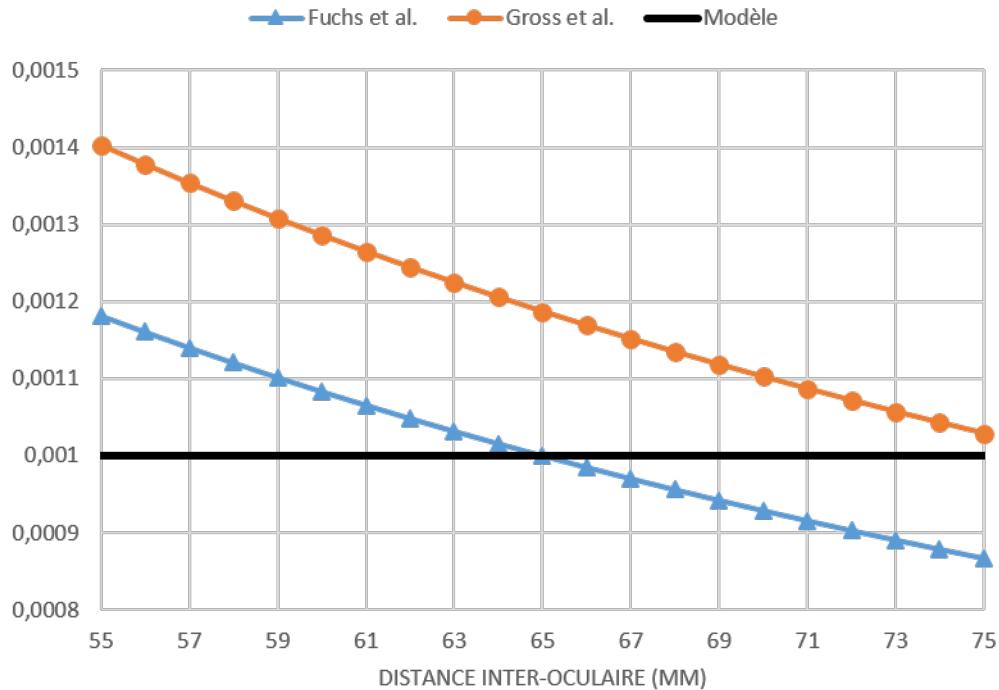


FIGURE 1.12 – Variation de la constante du modèle d’acuité stéréoscopique en fonction de la distance interoculaire.

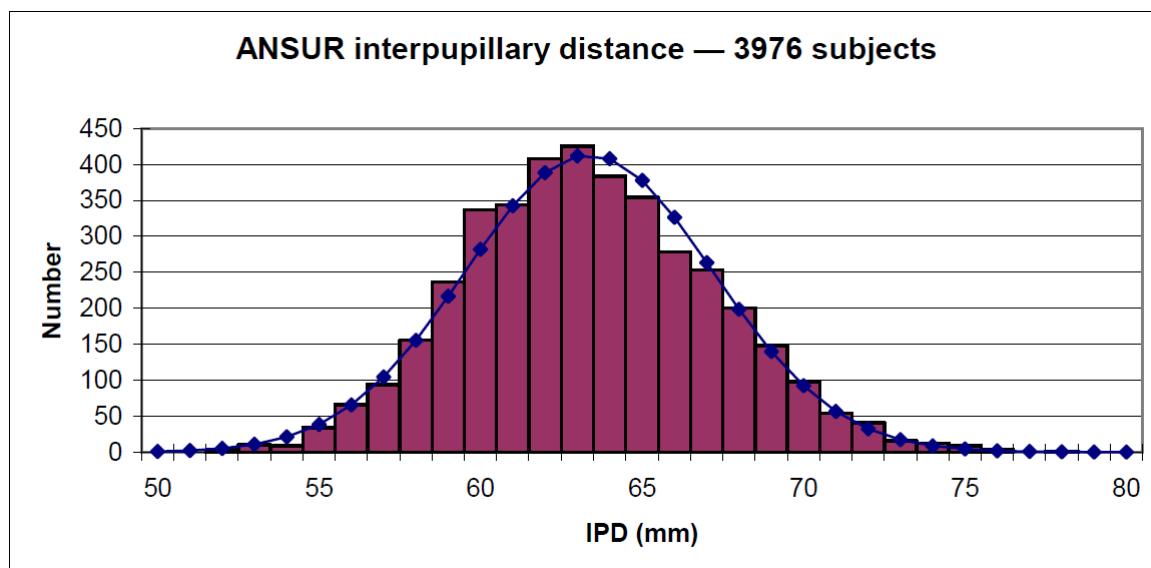


FIGURE 1.13 – Variation de la distance interoculaire sur la population. Image tirée de [Dodgson, 2004]

Avec  $\Delta\nu_{min}$  l'angle limite pour la vision stéréoscopique,  $d_{IPD}$  la distance inter-pupillaire et  $\Delta r$  la différence de profondeur qui peut être perçue à une distance  $r$  donnée.

## 1.6.2 Fonction de notation du critère

### Cas général

Au final, la fonction de notation du critère d'acuité stéréoscopique est le rapport entre l'angle limite pour la stéréoscopie à la plus basse luminance possible dans le système, et donc le plus critique, ( $\Delta\nu_{min}$  pris sur le graphe tiré de [Gross *et al.*, 2008], en *arcsecs*) et la résolution angulaire que possède le système de réalité virtuelle ( $\alpha$ , en *arcsecs*). Cette fonction est décrite par l'équation suivante (Eq. 1.11) :

$$F_{stereo\_acuity}(x, r) = \begin{cases} 100 & \alpha < \Delta\nu_{min} \\ 100 * \frac{\Delta\nu_{min}}{\alpha} & \text{else} \end{cases} \quad (1.11)$$

### Cas du CAVE

Dans un cas d'application CAVE, la luminance varie en général entre  $10^{-1}$  et  $10^{+1} \text{ cd/m}^2$ , ce qui donne un  $\Delta\nu_{min}$  moyen de 8 *arcsecs*. Cette valeur est obtenue graphiquement en moyennant les valeurs d'angle stéréoscopique limite pour toutes les valeurs de luminance possible dans un CAVE. L'évolution de l'angle stéréoscopique limite en fonction de la luminance vient d'une figure tracée p.67 dans [Gross *et al.*, 2008]. Dans ce cas d'utilisation, le CAVE, la fonction est tracée en Fig. 1.14.

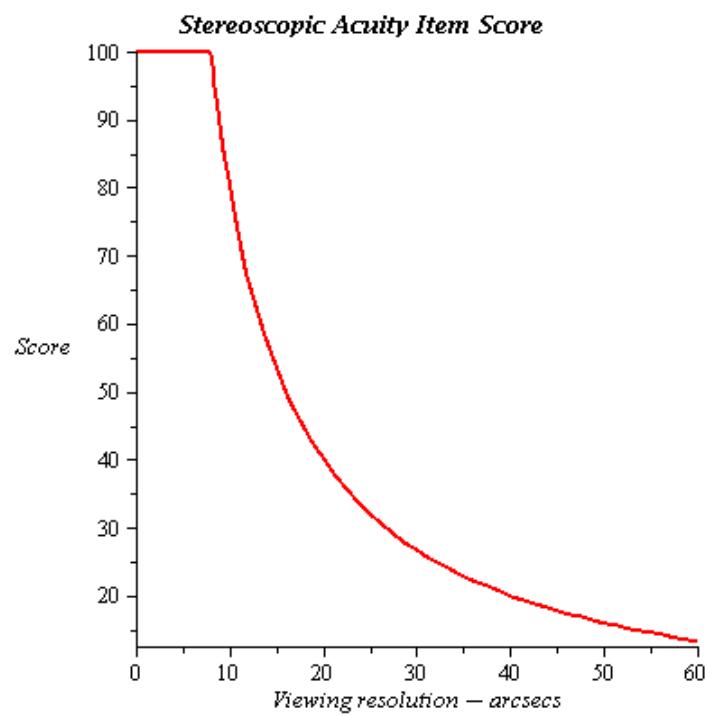


FIGURE 1.14 – Tracé de la fonction de notation du critère « acuité stéréoscopique » ( $\Delta\nu_{min} = 8 \text{ arcsecs}$ )

## CHAPITRE

# 2

## INDICES D'IMMERSION

### 2.1 Latence

La latence peut avoir différentes définitions, listées ci dessous :

- « Mouvement à photon » : depuis le mouvement d'un utilisateur tracké jusqu'à l'affichage du nouveau point de vue. Peut être simplifié comme la latence « totale » du système et est la définition généralement utilisée.
- « Mouvement à pré-calcul » : le temps écoulé entre un mouvement tracké et de l'ordre de recalculation de l'image à une nouvelle position. Plus simplement, c'est le temps d'acquisition des capteurs de mouvements.
- « Pré-calcul à calcul » : temps écoulé pour générer le rendu visuel, par le traitement informatique de toutes les informations.
- « Calcul à photon » : temps écoulé entre la fin du calcul de rendu et la fin de l'affichage sur l'écran. C'est le temps qui s'écoule entre l'ordre d'actualisation de l'image et la fin de la modification de tous les pixels.

Bien qu'il existe des mesures de seuil de perception de la latence dans la littérature [Brooks, 1999; Kemeny, 2014] et des valeurs de latence humaine pour les images claires (74 ms) et pour les images sombres (106 ms) [Feng Han *et al.*, 2010], on souhaite compléter ces informations. On met donc sur pied une expérimentation pour étudier ce phénomène en comparant son influence dans la réalisation d'une tâche de visée dans un CAVE et dans un casque.

## 2.2 Champ de regard

Le champ de regard (FOR - Field of Regard en anglais) est l'extension du champ de vision : il est défini comme la portion d'espace qui est visible, au cours du temps, en incluant la possibilité de bouger la tête (mais pas le corps) et les yeux dans leurs orbites. De même que précédemment pour le champ de vision, on élabore une fonction de score divisée en deux branches ; la branche horizontale (H-FOR) et la branche verticale (V-FOR). Ces deux axes sont ensuite pondérés l'un par rapport à l'autre.

### 2.2.1 Champ horizontal & champ vertical

En prenant en compte les mouvements possibles de la tête et des yeux (jusqu'à 15 degrés quand même), chaque œil peut couvrir un champ de regard de plus de 200 degrés du côté temporal et d'environ 130 degrés du côté nasal, sur l'axe horizontal. L'axe vertical représente lui environ 310 degrés répartis à +140 et -170 d'après [Fuchs *et al.*, 2003].

D'un côté, et parce que le champ visuel se recouvre, on limite la valeur maximale de H-FOR à 360 degrés ( $h_{max}$ ). Le maximum de V-FOR est lui laissé tel quel à 310 degrés ( $v_{max}$ ). De l'autre côté, les valeurs minimales ne peuvent pas être plus petites que les valeurs mesurées pour le champ de vision. On les nomme ici  $h_0$  et  $v_0$  mais valent les mêmes valeurs que les  $h$  et  $v$  du champ de vision. L'évolution de la notation est linéaire pour chaque axe entre le minimum et le maximum.

L'équation est donc (Eq. 2.1), pour le champ de regard horizontal (H-FOR) et (V-FOR), avec  $h$  et  $v$  les valeurs mesurées des axes horizontaux et verticaux ; et  $h_0, v_0, h_{max}, v_{max}$  définis précédemment pour le champ de vision :

$$\begin{cases} F_{H-FOR}(h) = \frac{100}{h_{max}-h_0} \cdot (h - h_0) \\ F_{V-FOR}(v) = \frac{100}{v_{max}-v_0} \cdot (v - v_0) \end{cases} \quad (2.1)$$

### 2.2.2 Pondération

En faisant la même hypothèse que pour la pondération du champ de vision, on base les poids relatifs des deux axes sur leurs tailles respectives : 400 degrés pour l'axe horizontal et 310 pour l'axe vertical. Les 400 degrés de l'axe horizontal diffèrent des 360 utilisés pour la notation de l'axe (pour laquelle on avait limité l'angle à 360 degrés) et peuvent intercaler. Si mécaniquement lorsque que l'on regarde autour de soi on regarde à 360 degrés, dans les faits comme les demi-champs de regard (celui du côté temporal et celui du côté nasal) sont supérieurs à 180 degrés, on peut voir une petite partie de l'autre demi-champ de regard : une partie de l'espace est visible de deux manières (*cf.* Fig. 2.1). En terme de possibilités donc le champ est plus grand que 360 degrés d'où la valeur de 400 degrés conservée. On utilise donc la même méthodologie que pour le champ de vision et on

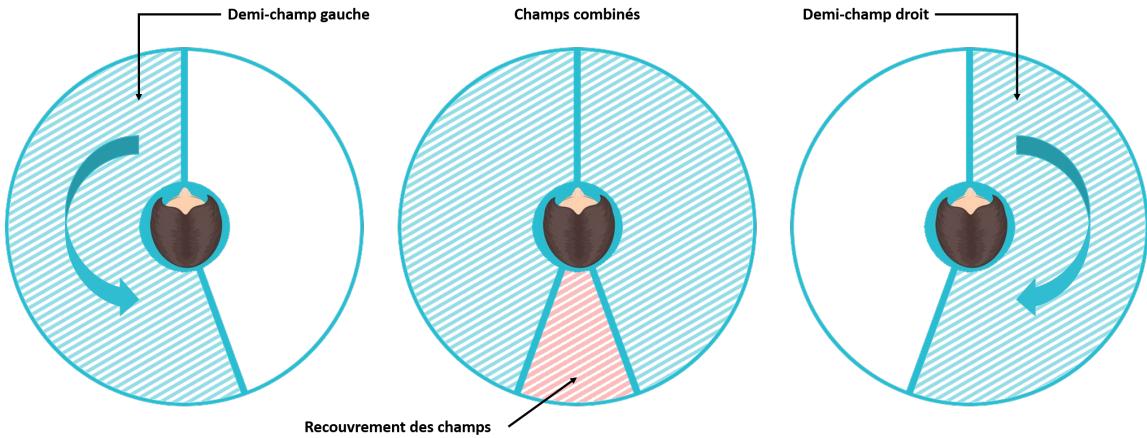


FIGURE 2.1 – Recouvrement des demi-champs de regard (temporal et nasal).

obtient (Eq. 2.2) :

$$\begin{cases} k_h = \frac{400}{400+310} = 0.56 \\ k_v = 1 - k_h = 0.44 \end{cases} \quad (2.2)$$

Ce qui est finalement assez proche de la pondération du champ de vision. Au final, la fonction de notation du champ de regard est définie par l'équation suivante (Eq. 2.3) :

$$F_{FOR}(h, v) = k_h \cdot F_{H-FOR}(h) + k_v \cdot F_{V-FOR}(v) \quad (2.3)$$

Cette hypothèse de pondération est valable pour le cas général. On a vu par la suite, qu'elle pouvait être mise à mal dans certains cas de figures (*cf.* chapitre sur l'expérimentation comparative entre les scores d'acceptation et les scores du modèle) et qu'il fallait, dans le cas par exemple d'une simulation de conduite, mettre la quasi intégralité de la pondération sur le champ de regard horizontal.

## 2.3 Stéréoscopie

### 2.3.1 Fonctionnement

La stéréoscopie est une méthode pour donner de la profondeur et du relief à des images standard en 2D. Les yeux, de part leur écartement, voient des images d'un point de vue différent. En combinant -notamment- les informations (et notamment les différences entre les deux points de vue) le cerveau est capable de récréer la profondeur. Le principe de la stéréoscopie est donc de fournir à chaque une image différente, calculée avec le bon point de vue. Cela peut être réalisé via un certain nombre de techniques avec des lunettes portées (Casque, lunettes obturantes, lunettes polarisées, anaglyphes, ...) qui sont présentées par Fauster [Fauster & Wien, 2007]. On rappelle ici les deux techniques les plus connues pour amener du relief.



FIGURE 2.2 – Lunettes de stéréoscopie.

Le type de stéréoscopie le plus connu du grand public est sûrement la technique dite anaglyphe qui fonctionne avec des lunettes aux verres rouge et bleu (*cf. Fig. 2.2*). Le principe est d'afficher les deux images nécessaires à la stéréoscopie en même temps mais chacune superposée par un filtre de couleur rouge ou de couleur bleue. Le verre rouge ne laisse passer que l'image rouge tandis que le filtre bleu ne laisse passer que son image de couleur correspondante. Chaque voit donc une seule image et le cerveau peut ainsi reconstruire la profondeur.

L'autre technique répandue est la stéréo dite « active » ; c'est celle utilisée par exemple dans les cinémas. Cette fois, les lunettes ont des filtres identiques et relativement transparents (*cf. Fig. 2.2*). Les images destinées aux yeux sont affichées l'une après l'autre (et pas en même temps comme la technique précédente) et ce sont les lunettes qui font le tri pour attribuer la bonne image au bon œil : de manière synchronisée avec l'affichage des images un verre devient opaque tandis que l'autre est transparent et ainsi de suite en alternance. Lorsque l'image pour l'œil gauche est affichée, le verre gauche laisse passer la lumière quand le verre droit la bloque, et inversement.

Il existe aussi d'autres moyens de recréer une vision binoculaire comme les affichages auto-stéréoscopiques ou les affichages holographiques mais qui sortent du cadre d'étude car on se concentre sur les simulateurs immersifs basés sur des techniques de stéréo active.

### 2.3.2 Fonction de notation du critère

En première approche, qui serait à étoffer, on propose de noter le critère sur sa présence (100) ou son absence (0). La piste d'amélioration à suivre pourrait être une découpe de la fonction de notation par rapport à la technique utilisée : toutes les techniques ne se valent pas et leurs performances pourraient être comparées. On propose en Fig. 2.3 un exemple à titre d'illustration uniquement.

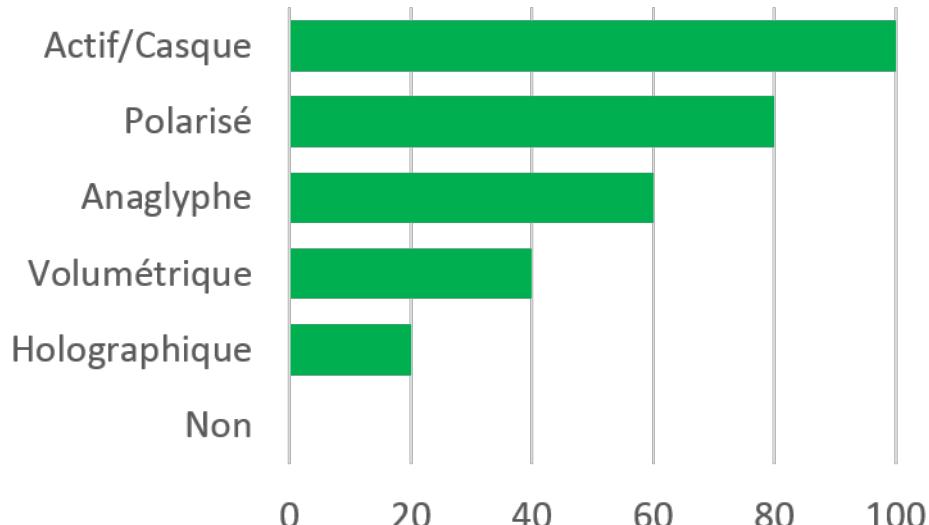


FIGURE 2.3 – Illustration d'une notation du critère « Stéréoscopie » en fonction de la technologie

## 2.4 Tracking

Le tracking est sûrement un des, si ce n'est LE, éléments les plus importants pour l'immersion. Le tracking permet d'inclure les mouvements de l'utilisateur dans la simulation, que ce soit du corps ou de la tête, mais aussi l'inclusion d'appareils extérieurs ou autres. Là encore, il existe un grand nombre de techniques, que ce soit embarqué dans les casques de réalité virtuelle ou bien des solutions externes. Dans le cas d'un CAVE on utilise en général des caméras infrarouge et un système de boules réfléchissantes sur les objets (ou parties du corps) à repérer dans l'espace (*cf.* Fig. 2.4). Pour chaque objet dont on entend connaître la position dans l'espace, on fixe au minimum trois boules réfléchissant les rayons infrarouges. Ces trois boules doivent être en permanence au même endroit sur l'objet et dans la même configuration (*cf.* Fig. 2.4 des deux côtés des lunettes). On balaye ensuite l'espace à tracker avec des caméras émettant et captant les rayons infrarouges. Le système de boules réfléchit ces rayons et est ainsi vu par les différentes caméras qui, comme elles sont placées à différents endroits dans l'espace, peuvent trianguler la position et l'orientation de l'objet étant donné que les boules réfléchissantes sont fixes par rapport à l'objet.

A défaut, et en attendant, de proposer une notation sur des critères plus développés, on propose de noter simplement ce critère sur sa présence (100) ou son absence (0). L'élément à creuser serait sa précision, c'est à dire son écart de position entre la position réelle de l'objet/body de tracking et la position vue par le système. Cette précision dépend néanmoins d'autres paramètres comme la calibration des caméras ou la calibration de la « room » (la définition des bordures de l'espace de travail). Il pourrait être aussi envisageable, au même titre que pour la stéréoscopie, de traiter le tracking en fonction de la technologie. Il faudrait alors comparer le tracking infrarouge avec le tracking magnétique et autre.



FIGURE 2.4 – « Body » de tracking monté sur des lunettes de stéréoscopie & ensemble caméra-émetteur infrarouge pour le tracking.

## 2.5 Uniformité

Bien que défini dans le modèle, le critère d'uniformité est le seul qui n'a pas été pleinement abordé. C'est un sujet complexe qui se divise en plusieurs parties. L'uniformité peut à la fois valoir à l'intérieur d'un même écran, c'est à dire entre différentes zones d'un écran, mais aussi -quand le système le permet de part sa construction- entre les différents écrans. Même si l'uniformité la plus évidente concerne la couleur (on souhaite qu'un rouge affiché soit identique de partout), il faut en fait l'élargir au d'autres critères :

- la couleur,
- la luminosité,
- le contraste.

La couleur est sûrement le critère le plus facile à vérifier car on sait mesurer des écarts entre les couleurs grâce aux équations de différentiations des couleurs (présentées dans la première partie du manuscrit). Ces équations donnent donc des écarts entre les couleurs qui, par construction doivent être inférieurs à une valeur de 1 pour qu'ils soient imperceptibles à l'œil humain. Néanmoins, en informatique notamment, le seuil réel est plus bas (la valeur en dessous de laquelle la différence est imperceptible est plus grande que le 1 théorique) pour les populations expertes dans le domaine de la couleur, et encore plus bas pour les populations néophytes [Vidal *et al.*, 2016]. On pourrait donc se baser sur ces trois valeurs (valeur théorique minimale, valeur de la population experte et valeur de la population néophyte) pour faire une première échelle de notation.

Néanmoins, l'uniformité concerne aussi la luminosité et le contraste qui sont des sujets bien plus compliqués à gérer, comme en témoigne la section qui leur est dévolue en début de partie.

## 2.6 Orientation des caméras

### 2.6.1 Mimétisme du fonctionnement oculaire

La capture d'image dans une scène virtuelle pour affichage sur un/des écran-s se fait au moyen d'une « caméra » virtuelle. Dans le cas de la stéréoscopie, comme il est nécessaire d'avoir une image par œil, c'est deux caméras virtuelles qui entrent en jeu. Ces caméras sont l'équivalent virtuel de nos yeux. Il semblerait donc naturel que les caméras aient un comportement relativement proche de ces derniers et pourtant ce n'est majoritairement pas le cas.

Lorsque l'on regarde un objet proche de nous, les yeux s'orientent dans leur orbite et se tournent vers cet objet. Plus l'objet est proche, plus il est nécessaire de converger. Au contraire, quand l'objet recule, et à partir d'une certaine distance, celui-ci est considéré comme « à l'infini » et les directions des yeux sont parallèles, soit une convergence nulle. Dans la majeure partie des cas, la convergence des caméras n'est pas implémentée : on parle alors de caméras « parallèles » (ou à « convergence » nulle). La raison vient du fait qu'utiliser des caméras arbitrairement convergées (et qui resteraient convergées, donc à « convergence fixe ») entraîne automatiquement un certain nombre de distorsions [Woods *et al.*, 1993]. Si l'utilisateur regarde un objet à une « convergence » différente de celle qui est prévue, l'orientation des caméras n'est pas la bonne. Ainsi, plutôt par soucis de simplicité, on fait en général l'hypothèse que tous les objets (ou en tout cas la majorité) seront suffisamment éloignés pour être dans une situation de convergence nulle.

Néanmoins on peut optimiser l'utilisation de ces distorsions si l'on est capable de régler la convergence des caméras en temps réel en fonction de la distance à laquelle l'utilisateur regarde. Cet ajout s'avère même bénéfique car il ajoute des disparités verticales qui font partie du processus de vision [Aurat, 2016]. On peut alors parler de « caméras à convergence variable ». Il faut alors acquérir la position du regard en temps réel, ce qui peut être fait soit matériellement en ajoutant par exemple un système de tracking du regard (*cf.* Fig. 2.5), soit en faisant des hypothèses telles que la personne regarde droit devant elle (et non en coin), donc la direction du regard est la même que la direction de la tête. La mise en œuvre de la convergence n'est également pas facilitée par le fait de capturer des images dans des plans non parallèles au plan de projection final. Les caméras convergées capturent des images planes mais dans un plan qui ne correspond pas forcément à la surface sur laquelle il faudra la projeter (*cf.* Fig. 2.6). Il faut donc passer par une étape supplémentaire de warping de l'image (en plus du warping qui peut exister pour que l'image d'un projecteur colle à l'écran de projection) [Aurat, 2016].

Cette dichotomie entre la convergence réelle des yeux et la convergence artificielle des caméras virtuelles n'est pas anodine car elle est à l'origine d'un des plus grands maux dans les simulateurs : le conflit accommodation-vergence qui génère notamment des fatigues, fatigues visuelles, participe au mal du simulateur et dégrade grandement la qualité d'immersion. En effet, les images sont affichées sur un écran à une certaine distance auquel les yeux s'accommodent tandis que les images jaillissent (ou inversement, s'enfoncent) de l'écran, obligeant les yeux à converger dessus. L'accommodation et la convergence qui sont normalement liées sont dès lors désynchronisées et génèrent un conflit pour le cerveau.



FIGURE 2.5 – Exemple de système de tracking de regard portatif.

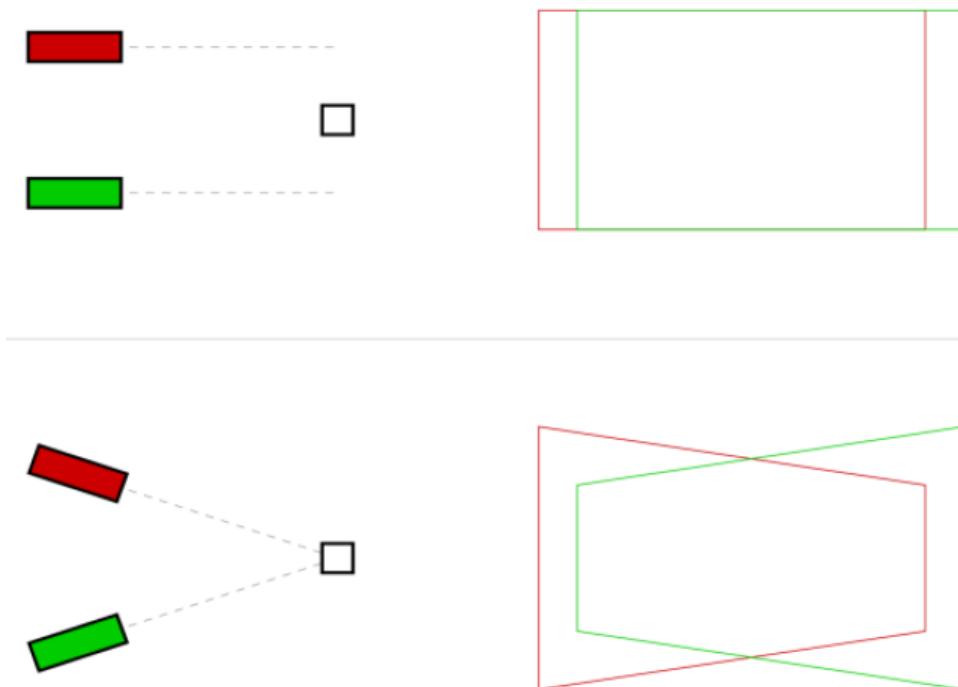


FIGURE 2.6 – Le carré est vu comme deux rectangles dans une configuration « caméras à convergence nulle » et comme deux trapèzes lorsque les caméras sont convergentes. Il faut donc retordre artificiellement les trapèzes pour qu'ils correspondent à la projection qui se fait sur une surface rectangulaire (image tirée de [Aurat, 2016]).

## 2.6.2 Fonction de notation du critère

Ainsi, le critère est divisé en trois, soit le nombre de possibilité qui existent :

- caméras à convergence fixe (c'est à dire sans connaître le point de regard), 0.
- caméras à convergence nulle (caméras parallèles), 80.
- caméras à convergence variable (avec connaissance du point de regard), 100.

# DISCUSSION GÉNÉRALE ET PERSPECTIVES

On conclue dans ce dernier chapitre l'ensemble des travaux relatifs à la thèse. On fait, dans un premier temps, la synthèse des résultats (modélisation et expérimentations) que l'on a obtenu, permettant ainsi d'avoir un modèle de score de réalisme. On applique ensuite ce dernier à différents systèmes présents chez Renault : casques de Réalité Virtuelle, casque de Réalité Augmentée, CAVEs, ... Pour terminer, on met en avant des applications concrètes qui pourraient utiliser notre modèle de score.

## Synthèse des résultats

### État des lieux de la modélisation

On a donc proposé un modèle de score de réalisme, basé sur le système visuel humain. Ce score ne se veut pas une nouvelle modélisation des fonctions visuelles humaines mais plutôt un guide pragmatique, pour l'ingénieur, pour estimer la qualité d'un système et sa capacité à envoyer des signaux de manière réaliste, c'est à dire conforme aux stimulation sensorielle dont on a l'habitude, indépendamment de ce qui est affiché.

Ce score est divisé en douze critères (Fig. I), répartis en deux entités de six : un groupe représentant les indices nécessaires à la vision, et un autre groupe représentant les indices d'immersion. Chaque critère est alors noté entre 0 et 100, basé sur les grandeurs qui caractérisent le système visuel humain. Une pondération des critères vient compléter le modèle et permet de prendre en compte l'utilisation que l'on fait du moyen immersif : en fonction de l'application, certains critères seront plus limitant que d'autres.

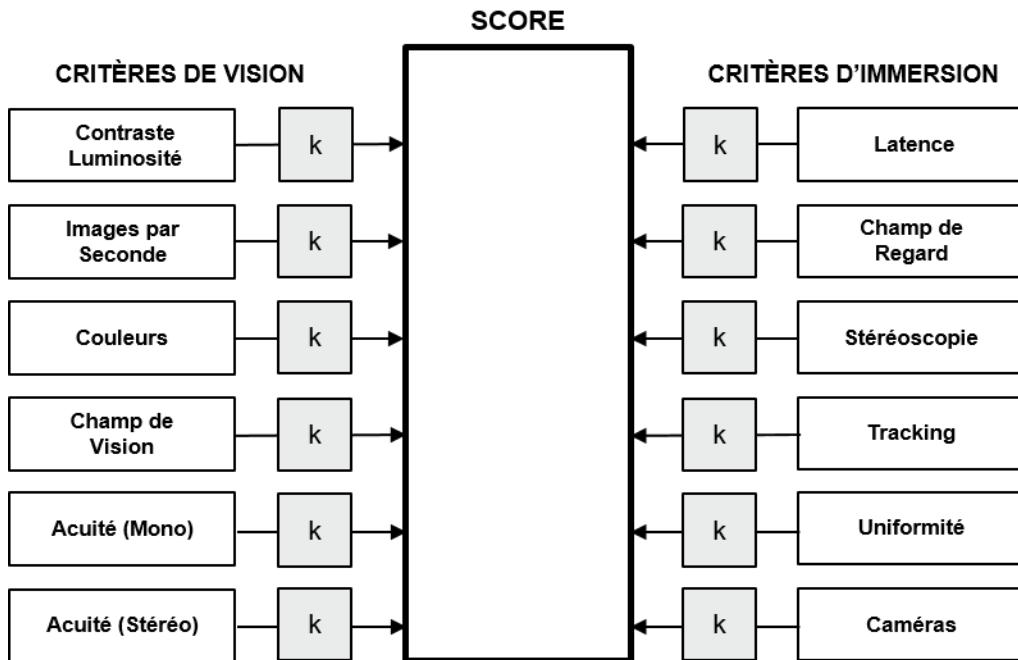


FIGURE I – Rappel de la modélisation du score de réalisme.

Le groupe d'indices de vision se compose des critères suivants : contraste et luminance, nombre d'images par seconde, quotité de couleurs affichables, taille du champ de vision, acuités monoscopique et stéréoscopique. Parallèlement, les critères d'immersion sont les suivants : latence, taille du champ de regard, stéréoscopie, tracking, uniformité des surfaces d'affichage et convergence des caméras. Enfin, on propose une pondération par ordres de grandeur (fort ou faible) pour chaque critère.

## Résultats expérimentaux

Dans le cadre de la thèse, on a réalisé trois expérimentations : la validation d'un modèle de performance visuelle pour le critère de contraste et luminance, la mesure de performance pour le critère de latence et la comparaison entre les notations subjectives des critères du modèle par des sujets et les notes objectives données par notre modèle. Cette dernière, de moindre ampleur par rapport aux autres, dégage néanmoins quelques tendances. La pondération des critères de champ de vision et de champ de regard est très dépendante de l'application : dans le cas de la conduite, l'utilisation du champ de regard est quasi-nulle et les sujets ne font pas la différence entre les deux champs. L'appréciation de la quotité de couleur semble sur-évaluée par habitude. Enfin, les femmes font preuve de plus de sévérité dans leur notation que les hommes.

Pour l'expérimentation sur le critère de contraste et de luminance, on a cherché à traduire une expérimentation en Réalité Virtuelle, afin de déterminer si le modèle de performance visuelle mis en jeu était utilisable dans notre domaine. Ce modèle prédit des conditions de luminance et de contraste en dessous desquelles le système visuel humain n'est pas capable de distinguer ce qu'on lui demande, et des conditions au dessus desquelles la performance de détection de l'œil n'augmente plus, déterminant ainsi la plage dans lequel un

moyen immersif devrait se situer. Nos résultats théoriques semblent se comporter de manière radicalement différente par rapport aux prévisions du modèle, tout en montrant une corrélation statistique. On montre que le modèle de performance visuelle est utilisable, sous réserve de l'inclusion d'un facteur prenant en compte les spécificités de la Réalité Virtuelle. Pour l'application à notre critère de contraste et luminance, on revient toutefois à préconiser une technique utilisant les fonctions de sensibilité au contraste et la découpe en fréquence spatiale des tâches classique de vision.

Enfin, l'expérimentation sur la latence est quant à elle une comparaison entre deux systèmes et à différents niveaux de latence dans chaque système, de la performance de sujets à réaliser une tâche écologique. Les sujets devaient viser une série de cibles dans un véhicule modélisé en 3D, en maximisant leur précision par rapport au centre de la cible et leur vitesse de réalisation. On montre notamment que la performance se dégrade de manière non continue sous l'influence de la latence, tandis que le mal du simulateur est lui linéairement affecté. De même, le changement de système (de CAVE à casque) améliore la performance des sujets mais au détriment de leur expérience utilisateur : on met en cause la nature des mouvements impliqués pour la réalisation de la tâche demandée ainsi que le conflit visio-vestibulaire.

## Travaux futurs

Si on présente une modélisation cohérente, il reste néanmoins des axes de travail et des propositions à améliorer ; le temps imparti pour la thèse ne nous ayant pas permis de tout traiter. Tout d'abord, bien que l'on ait eu des avancées concrètes pour le critère de contraste et luminance, nous ne sommes pas encore en mesure de pouvoir déterminer pratiquement son score. Par conséquent, le critère d'uniformité est encore à l'état d'embryon, étant intrinsèquement lié au contraste et à la luminance. On présente et on initie la suite des travaux sur le critère de contraste et de luminance avec l'utilisation des fréquences spatiales et des fonctions de contraste.

D'autres critères cependant, tels que la quotité de couleur, la stéréoscopie et le tracking se sont vus attribuer une fonction de notation (linéaire ou binaire) que l'on juge insuffisante à terme ; ces notations sont fonctionnelles mais doivent être raffinées. On a présenté, par exemple dans le cas de la stéréoscopie, une piste d'amélioration de la fonction de notation, basée sur la technologie utilisée.

La plus grande partie de futurs travaux reste la pondération du modèle. Si on propose des ordres de grandeur, il sera nécessaire, dans une second temps, d'avancer des propositions chiffrées, amenant plus de finesse pour départager les critères. Ces pondérations chiffrés seront très dépendantes du cas d'utilisation et permettront de prendre en compte certains scénarios spécifiques (comme la haute vitesse par exemple).

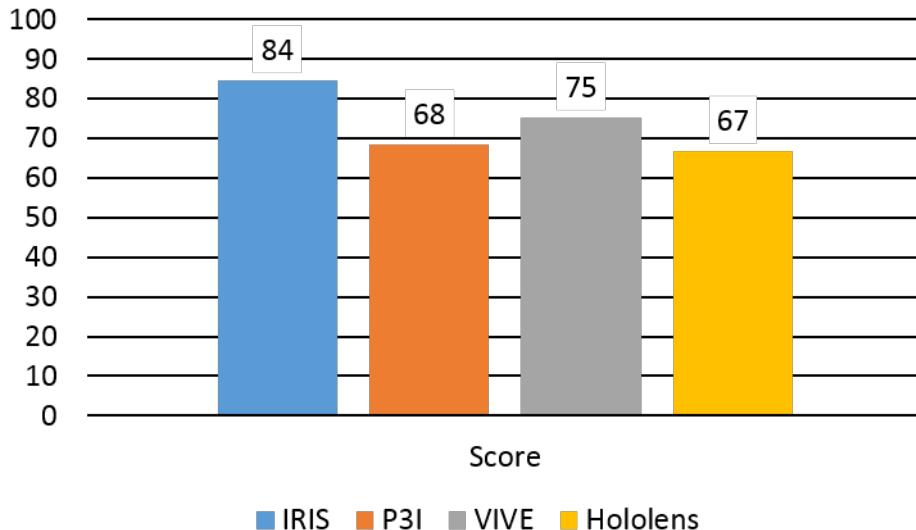


FIGURE II – Score de réalisme pour les quatre systèmes testés.

## Application du score de réalité

Si la thèse a pour l'instant eu une application soit théorique, soit pratique mais au service de la théorie, il manque encore une application purement pratique de notre modèle de score. Or, on est désormais globalement en mesure d'appliquer notre modèle de score de réalité visuel, à l'exception de deux critères, intrinsèquement liés, que sont le critère de contraste et luminance et le critère d'uniformité. On sélectionne des moyens immersifs présents chez Renault et on leur applique, sur chacun de leurs critères, les fonctions de notation déployées au cours de la thèse. Pour les critères non-aboutis, on utilise des estimations guidées par la littérature, à défaut de pouvoir les noter directement. De même, pour la pondération, on applique un coefficient uniforme de 2 pour les critères « forts » et de 1 pour les critères « faibles ». On sélectionne les systèmes suivants : deux CAVEs de chez Renault (nommés IRIS et P3I), un casque de Réalité Virtuelle grand public mais utilisé fréquemment dans l'industrie (HTC Vive) et un casque de Réalité Augmentée (Hololens). Techniquement, notre modèle de score n'est pas conçu pour une application en Réalité Augmentée, mais il peut être intéressant d'en tester les limites sur de tels dispositifs. Les résultats généraux de notation sont représentés en Fig. II.

IRIS III

P3I IV

VIVE V

Hololens VI

Conclusion et discussion

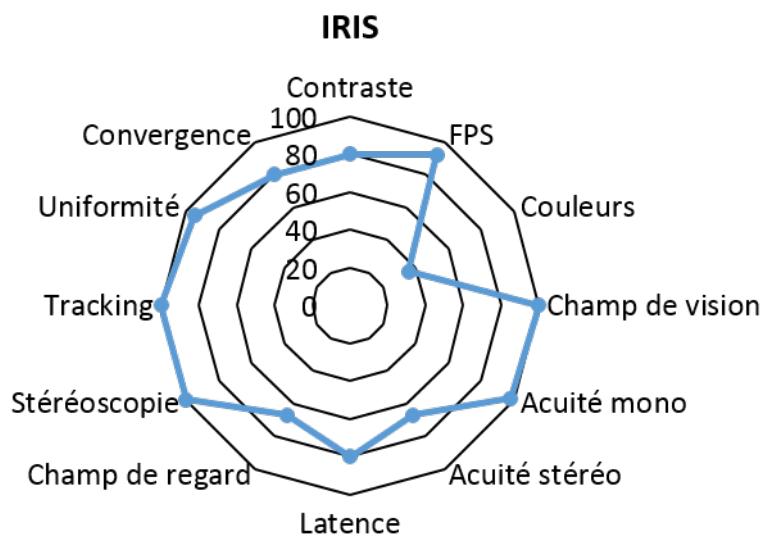


FIGURE III – Diagramme radar des critères du score pour IRIS.

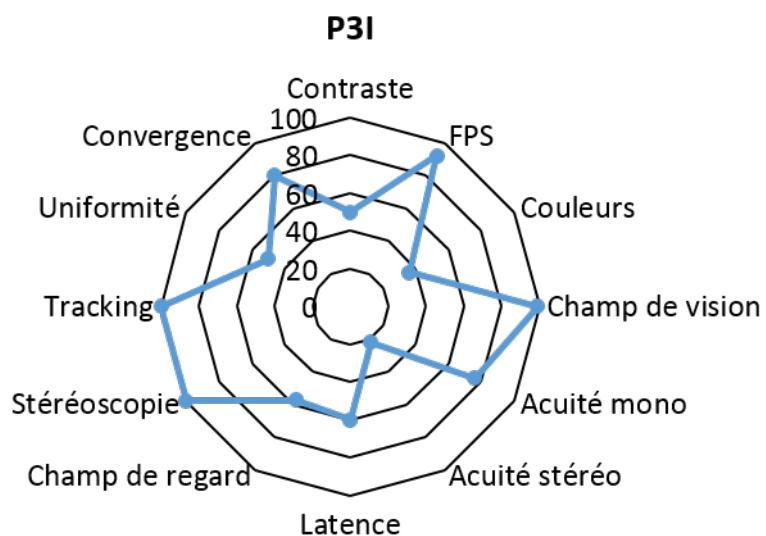


FIGURE IV – Diagramme radar des critères du score pour P3I.

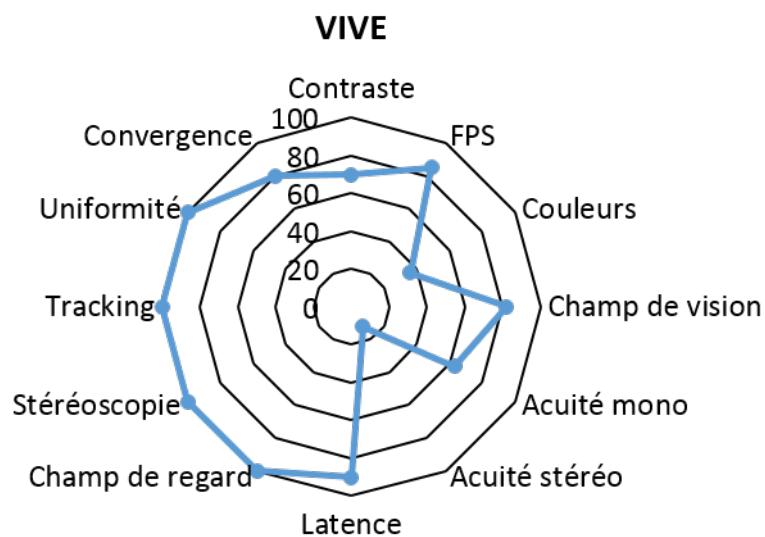


FIGURE V – Diagramme radar des critères du score pour le casque VIVE.

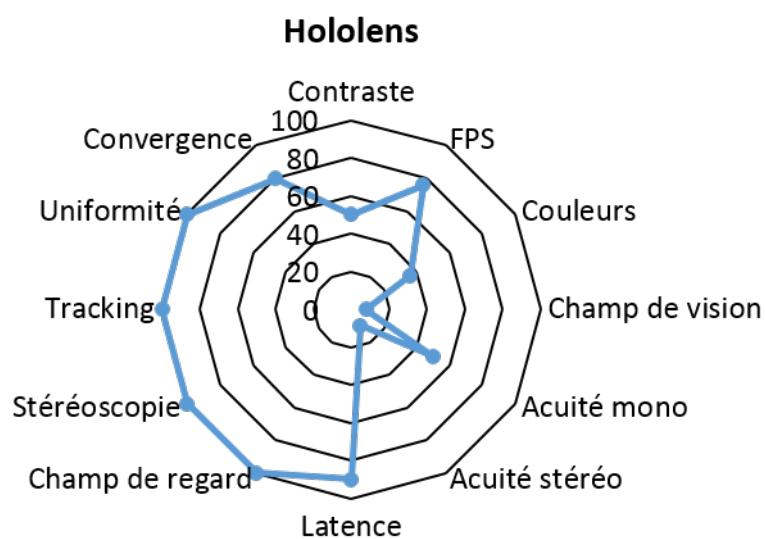


FIGURE VI – Diagramme radar des critères du score pour le casque Hololens.

## Perspectives d'utilisation

Pour terminer ce manuscrit, on propose de présenter des idées concrètes d'utilisation de notre score de réalisme. Le premier usage étant évidemment celui de l'aide à la conception ou à la mise à jour des systèmes immersifs. Lorsque que l'on doit par exemple améliorer les caractéristiques d'un système de RV, les critères disposant déjà d'une bonne note ne seront pas à traiter en premier. De même, une partie du système ayant atteint le score maximal, ne nécessite à priori pas d'amélioration, à moins que cette dernière ait un effet direct sur un autre critère.

On peut ensuite imaginer un usage visant à déterminer la lisibilité dans un simulateur : en pondérant tous les critères liés à cette tâche (acuités, contraste et luminance, fluidité en cas de haute vitesse) et en réglant la pondération des critères moins inutiles à 0. On pourra alors estimer quel moyen immersif est le plus adapté à des expérimentations impliquant de petits détails au loin comme la lecture en amont de panneaux d'autoroute (destination, limitations de vitesse, ...).

Enfin, il existe depuis 2016 une initiative permettant de mettre en location ses systèmes immersifs pour un usage quelconque : VR-BNB. Les entreprises (ou laboratoires) créent alors une page décrivant le système qu'elles mettent à disposition : on peut imaginer intégrer sur cette page, ou dans les critères de recherche, le score de réalisme dudit système, en pondération globale ou pondéré dans un cas d'utilisation spécifique. Les futurs utilisateurs auraient alors plus d'indications sur la capacité du système qu'ils envisagent de louer à répondre à leurs besoins expérimentaux.

# **Bibliographie**

- Aurat, David. 2016. *Immersion Visuelle Tridimensionnelle Haute Performance par Suivi du Mouvement des Yeux*. Informatique Temps Réel, Robotique & Automatique, Université de Recherche Paris Sciences 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.
- 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.
- 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.
- 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.
- 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.
- Kemeny, Andras. 2014. From driving simulation to virtual reality. ACM Press.
- Kuehni, Rolf G. 2015. How many object colors can we distinguish? *Color Research & Application*, July.

- 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.
- Linhares, João Manuel Maciel, Pinto, Paulo Daniel, & Nascimento, Sérgio Miguel Cardoso. 2008. The number of discernible colors in natural scenes. *Journal of the Optical Society of America A*, **25**(12), 2918.
- 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.
- Perales, Esther, Martínez-Verdú, Francisco, & Viqueira, Valentín. 2008. Calculation of number of distinguishable colours by real normal observers. *Conference on Colour in Graphics, Imaging, and Vision*, **2008**(1), 80–84.
- Pointer, M. R. 1980. The Gamut of Real Surface Colours. *Color Research & Application*, **5**(3), 145–155.
- Pointer, M. R., & Attridge, G. G. 1998. The number of discernible colours. *Color Research & Application*, **23**(1), 52–54.
- 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.
- 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.
- Vidal, Q., Michelin, S., Laborie, B., & Kemeny, A. 2016. Color-Difference Assessment and Enhancement for Driving Headlight Simulation. *SIMULATION*, **92**(5), 427–435.
- Wen, Senfar. 2006. Display gamut comparison with number of discernible colors. *Journal of Electronic Imaging*, **15**(4), 043001–043001–8.
- Woods, Andrew J., Docherty, Tom, & Koch, Rolf. 1993 (Sept.). Image distortions in stereoscopic video systems.