

THÈSE DE DOCTORAT  
**Immersion Visuelle Hyper-Réaliste et  
Multi-Sensorielle 3D**



# RÉSUMÉ

asd



# ABSTRACT

asd



# REMERCIEMENTS

qsd



# TABLE DES MATIÈRES

<b>Table des figures</b>	<b>xvii</b>
<b>Liste des tableaux</b>	<b>xxiii</b>
<b>Introduction générale</b>	<b>1</b>
<b>I État de l'Art : Partie Théorique</b>	<b>3</b>
<b>1 Définition du réalisme</b>	<b>5</b>
1.1 Enjeux et problématiques . . . . .	5
1.2 Définitions usuelles . . . . .	5
1.3 Définitions dans le domaine de la Réalité Virtuelle . . . . .	6
1.4 Cadre d'étude . . . . .	8
<b>2 Définition de la multi-sensorialité 3D</b>	<b>11</b>
2.1 Immersion . . . . .	11

2.2	Présence . . . . .	12
2.3	Proprioception . . . . .	12
2.4	Application dans la thèse . . . . .	12
<b>3</b>	<b>Fonctionnement de la vision</b>	<b>13</b>
3.1	Structure de l'œil . . . . .	13
3.2	Modélisation de l'œil . . . . .	15
3.3	Vision binoculaire . . . . .	19
3.4	Traitements post-rétiniens . . . . .	20
<b>4</b>	<b>Perception de la couleur</b>	<b>25</b>
4.1	Espaces colorimétriques . . . . .	25
4.2	Lois et effets . . . . .	27
4.3	Observateurs standards . . . . .	29
4.4	Illuminants . . . . .	30
4.5	Équations de différentiations des couleurs . . . . .	31
<b>5</b>	<b>Perception visuelle de la profondeur</b>	<b>35</b>
<b>6</b>	<b>Contraste</b>	<b>39</b>
6.1	Définitions mathématiques . . . . .	39
6.2	Lois de comportement . . . . .	40
<b>7</b>	<b>Influences Physiologiques</b>	<b>43</b>
7.1	Variations morphologiques . . . . .	43
7.2	Âge . . . . .	46

---

<b>II Première approche : modèles de vision</b>	<b>47</b>
<b>Introduction</b>	<b>49</b>
<b>1 État de l'Art</b>	<b>51</b>
1.1 Modèles principaux . . . . .	52
1.2 Pérennités des modèles principaux . . . . .	52
1.3 Modèles de Mantiuk <i>et al.</i> [2004, 2005, 2015] . . . . .	52
1.4 Reproduction de tons . . . . .	53
<b>2 Modèle de Pattanaik <i>et al.</i></b>	<b>57</b>
<b>3 Modèle de Mantiuk &amp; Ramponi</b>	<b>63</b>
<b>Conclusion</b>	<b>65</b>
<b>III Deuxième approche : score de réalisme</b>	<b>67</b>
<b>Introduction</b>	<b>69</b>
<b>1 Score de réalisme</b>	<b>73</b>
1.1 Ambition . . . . .	73
1.2 Méthodologie . . . . .	74
<b>2 Proposition de modèle</b>	<b>77</b>
2.1 Propositions préliminaires . . . . .	78
2.1.1 Première modélisation (Fig. 2.1) . . . . .	78
2.1.2 Seconde modélisation (Fig. 2.2) . . . . .	79
2.1.3 Troisième modélisation (Fig. 2.3) . . . . .	80
2.2 Modèle définitif . . . . .	82

<b>3 Indices de vision</b>	<b>85</b>
3.1 Contraste & luminosité . . . . .	85
3.1.1 Fonction de sensibilité au contraste (CSF) . . . . .	86
3.1.2 Performance visuelle relative . . . . .	87
3.2 Images par seconde . . . . .	88
3.2.1 Minimum de fonctionnement . . . . .	88
3.2.2 Phénomène de scintillement (flickering) . . . . .	88
3.2.3 Maximum de fonctionnement . . . . .	89
3.2.4 Fonction de notation du critère . . . . .	90
3.3 Couleurs . . . . .	90
3.3.1 Dénombrement des couleurs visibles . . . . .	90
3.3.2 Les espaces colorimétriques . . . . .	91
3.3.3 Indice de rendu des couleurs (IRC) . . . . .	92
3.3.4 Fonction de notation du critère . . . . .	93
3.4 Champ de vision . . . . .	93
3.4.1 Axe horizontal . . . . .	94
3.4.2 Axe vertical . . . . .	95
3.4.3 Pondération . . . . .	95
3.4.4 Fonction de notation du critère . . . . .	96
3.5 Acuité monoscopique . . . . .	96
3.5.1 Première approche . . . . .	96
3.5.2 Deuxième approche . . . . .	97
3.5.3 Fonction de notation du critère . . . . .	98
3.6 Acuité stéréoscopique . . . . .	98
3.6.1 Première modélisation . . . . .	98

3.6.2 Fonction de notation du critère . . . . .	101
<b>4 Indices d'immersion</b>	<b>103</b>
4.1 Latence . . . . .	103
4.2 Champ de regard . . . . .	104
4.2.1 Champ horizontal & champ vertical . . . . .	104
4.2.2 Pondération . . . . .	104
4.3 Stéréoscopie . . . . .	105
4.3.1 Fonctionnement . . . . .	105
4.3.2 Fonction de notation du critère . . . . .	106
4.4 Tracking . . . . .	107
4.5 Uniformité . . . . .	108
4.6 Orientation des caméras . . . . .	109
4.6.1 Mimétisme du fonctionnement oculaire . . . . .	109
4.6.2 Fonction de notation du critère . . . . .	111
<b>5 Pondération du modèle</b>	<b>113</b>
5.1 Contraste, luminance, taille et champs visuels . . . . .	113
5.2 Tracking, stéréoscopie, convergence des caméras et couleurs . . . . .	114
5.3 Conclusion . . . . .	115
<b>6 Etude expérimentale</b>	<b>117</b>
6.1 Objectif . . . . .	117
6.2 Apparatus . . . . .	117
6.3 Résultats . . . . .	119
6.4 Discussion . . . . .	121
6.4.1 Couleurs et acuité monoscopique . . . . .	121

---

6.4.2	Champ de vision & champ de regard . . . . .	121
<b>Conclusion</b>		<b>123</b>
<b>IV Partie expérimentale : contraste et luminance</b>		<b>125</b>
<b>Introduction</b>		<b>127</b>
<b>1 Modèles de Réaction</b>		<b>129</b>
1.1	Première modélisation de la performance visuelle relative . . . . .	129
1.1.1	Modèle précurseur . . . . .	129
1.1.2	Application à la performance visuelle . . . . .	130
1.2	Méthode des temps de réaction . . . . .	131
1.2.1	Protocole expérimental . . . . .	131
1.2.2	Calcul des temps de réaction théoriques . . . . .	133
<b>2 Mesures préliminaires</b>		<b>135</b>
2.1	Luminance globale (luminance de fond) . . . . .	135
2.2	Luminance de la cible (luminance de tache) . . . . .	137
2.3	Diamètre pupillaire . . . . .	138
2.4	Absorption des verres des lunettes 3D . . . . .	141
<b>3 Protocole expérimental</b>		<b>145</b>
3.1	Dispositif . . . . .	145
3.1.1	Hypothèses de travail . . . . .	145
3.1.2	Sujets & matériel . . . . .	146
3.1.3	Tâche à effectuer . . . . .	147
3.2	Choix des conditions expérimentales . . . . .	148

---

3.2.1 Luminance . . . . .	148
3.2.2 Contraste . . . . .	150
<b>4 Résultats</b>	<b>153</b>
4.1 Prédictions du modèle théorique . . . . .	153
4.2 Mesures réelles . . . . .	153
4.3 Analyse et discussion . . . . .	156
<b>Conclusion</b>	<b>159</b>
<b>V Expérimentation : latence</b>	<b>161</b>
<b>Introduction</b>	<b>163</b>
<b>1 Mesures préliminaires</b>	<b>165</b>
<b>2 Protocole expérimental</b>	<b>167</b>
2.1 Dispositif . . . . .	167
2.1.1 Hypothèses de travail . . . . .	167
<b>Conclusion</b>	<b>169</b>
<b>Conclusion générale</b>	<b>171</b>
<b>Bibliographie</b>	<b>173</b>
<b>ANNEXES</b>	<b>c</b>
<b>Calcul de la performance visuelle, méthode des temps de réaction</b>	<b>c</b>

<b>Luminance de fond</b>	<b>g</b>
<b>Questionnaires</b>	<b>i</b>
Questionnaire de Propension à l'Immersion - Version canadienne-française . . . . .	j
Questionnaire de Cybermalaise - Version canadienne-française . . . . .	n
Questionnaire de Présence - Version canadienne-française . . . . .	p
<b>Publications</b>	<b>w</b>
DSC 2017 EUROPE VR . . . . .	x
Transportation Research Part F : Traffic Behavior and Psychology . . . . .	z
IEEE Transactions on Computers . . . . .	an

# TABLE DES FIGURES

1.1	Illustration de bokeh . . . . .	7
1.2	Différentes acceptations du terme <i>réalisme</i> . . . . .	8
1.3	Cadre d'étude de la thèse . . . . .	9
3.1	Structure de l'œil . . . . .	15
3.2	Répartition (en <i>milliers/mm<sup>2</sup></i> ) des cônes/bâtonnets sur la rétine . . . . .	16
3.3	Points caractéristiques (cardinaux) d'un système optique plan . . . . .	17
3.4	Modèle et valeurs physiologiques de Gullstrand, image tirée de [Liou & Brennan, 1997] . . . . .	18
3.5	Modèle de Liou & Brennan, image tirée de [Liou & Brennan, 1997] . . . . .	18
3.6	Horoptères théorique et empirique, Aire de Panum et disparités rétinien-nes. Image tirée de [Neveu, 2012]. . . . .	20
3.7	Structure et cellules composant la rétine, image tirée de [ANSES, 2014]. . . . .	21
3.8	Direction des voies ventrale et dorsale dans le cerveau. . . . .	22
3.9	Construction des canaux chromatiques et achromatique dans la théorie des processus antagoniques . . . . .	23

3.10 Courbes moyennes de sensibilité au contraste en fonction de la fréquence spatiale du stimulus pour différentes valeurs d'acuité visuelle, tiré de [Owsley <i>et al.</i> , 1983] . . . . .	23
4.1 Composantes de l'espace CIELAB . . . . .	26
4.2 Répartition des domaines photopique, mésopique et scotopique . . . . .	29
4.3 Courbes d'efficacité lumineuse des Observateurs Standards de la CIE . . . . .	30
4.4 Spectre de l'illuminant D65 (CIE) . . . . .	31
5.1 Indices visuels de perception de la profondeur . . . . .	36
5.2 Illustration d'indices de profondeur : ombres (1, image tirée de [ANSES, 2014]), taille (2) et gradient de profondeur (3, images tirées de [Glassner, 1995]) . . . . .	37
7.1 Schématisation de la différence de profondeur stéréoscopique perçue en fonction de la distance interoculaire . . . . .	44
7.2 Abaques de la nouvelle distance perçue en fonction de la distance interoculaire . . . . .	45
7.3 Densité des cônes en fonction de l'âge (figure tirée de [Kilbride <i>et al.</i> , 1986]) .	46
1.1 Différentes étapes du modèle d'adaptation lumineuse. Images tirées de [Ferwerda <i>et al.</i> , 1996] . . . . .	55
2.1 Schéma synthétique du déroulé du modèle de [Pattanaik <i>et al.</i> , 1998] . . . . .	58
2.2 Schéma exhaustif du traitement de l'image par le modèle de [Pattanaik <i>et al.</i> , 1998] . . . . .	62
1 Fonctionnement du processus de mesure de qualité d'image, pour des séquences vidéo 2D. Traduction d'une figure reprise de Winkler [2000] . . . . .	70
2 Illustration de la théorie de détection du signal. Répartition des probabilités de détection autour d'un seuil de détection déterminé (critère de détection) .	71
1.1 Illustration de la différence qu'il existe entre les consignes données par la carte graphique (1) et les informations visuelles reçues par l'observateur (2). .	74

1.2 Résumé des étapes d'état de l'Art pour la construction du modèle et répartition des objectifs de représentation des notes intermédiaires dans le score de réalisme. . . . .	75
2.1 Première modélisation du score de réalisme. . . . .	79
2.2 Deuxième modélisation du score de réalisme. . . . .	80
2.3 Troisième modélisation du score de réalisme. . . . .	81
2.4 Dernière modélisation du score de réalisme. . . . .	83
3.1 Fonctions de sensibilité au contraste pour les domaines photopiques, scotopiques et mésopiques. . . . .	86
3.2 Exemple de performance visuelle pour deux tailles de cible différentes . . . . .	88
3.3 Illustration du « mouvement beta » et de l’« effet phi » . . . . .	89
3.4 Tracé de la fonction de notation du critère « images par seconde » . . . . .	91
3.5 Tracé de différents gamuts sur l'espace colorimétrique CIE RGB 1931. . . . .	92
3.6 Tracé de la fonction de notation du critère « couleurs » . . . . .	93
3.7 Répartition des zones visuelles sur l'axe horizontal du champ de vision. . . . .	94
3.8 Répartition des zones visuelles sur l'axe vertical du champ de vision. . . . .	95
3.9 Tracé de la sous-fonction horizontale de notation du critère « champ de vision » (ligne continue) et de la sous-fonction verticale (pointillés). . . . .	96
3.10 Classification ICD-9-CM . . . . .	97
3.11 Tracé de la fonction de notation du critère « acuité monoscopique » . . . . .	99
3.12 Variation de la constante du modèle d'acuité stéréoscopique en fonction de la distance interoculaire. . . . .	100
3.13 Variation de la distance interoculaire sur la population. Image tirée de [Dodgson, 2004] . . . . .	100
3.14 Tracé de la fonction de notation du critère « acuité stéréoscopique » ( $\Delta\nu_{min} = 8 \text{ arcsecs}$ ) . . . . .	102
4.1 Recouvrement des demi-champs de regard (temporal et nasal). . . . .	105
4.2 Lunettes de stéréoscopie. . . . .	106

4.3 Illustration d'une notation du critère « Stéréoscopie » en fonction de la technologie . . . . .	107
4.4 « Body » de tracking monté sur des lunettes de stéréoscopie & ensemble caméra-émetteur infrarouge pour le tracking. . . . .	108
4.5 Exemple de système de tracking de regard portatif. . . . .	110
4.6 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]). . . . .	110
 5.1 Modélisation avec ordre de grandeur de la pondération. . . . .	116
6.1 Echelle « aidée » pour noter subjectivement de 1 à 5 les différents critères du système . . . . .	118
6.2 Résultats des valeurs théoriques calculées et des valeurs subjectives relevées. 120	
 1.1 Illustration de l'effet de compression. Image tirée de [Naka & Rushton, 1966]	130
1.2 Installation de l'expérimentation de Rea et Ouellette. Image tirée de [Rea & Ouellette, 1988] . . . . .	132
 2.1 Chromameter CS100 . . . . .	136
2.2 Evolution de la luminance en fonction du niveau de gris sur les cinq points de mesure et leur moyenne. . . . .	137
2.3 Luminance de la cible en fonction de la luminance globale . . . . .	139
2.4 Oculomètre Dikablis Professional . . . . .	139
2.5 Exemple d'une mesure en pixel de largeur pupillaire au cours du temps. . . . 140	
2.6 Mesures pour la conversion pixel/mm. . . . .	141
2.7 Comparaison entre les mesures de diamètre pupillaire par rapport aux valeurs théoriques fonction de la luminance affichée. . . . .	142
2.8 Comparaison des mesures sans et avec filtre des lunettes stéréoscopiques. . . . 142	
2.9 Pourcentage d'absorption en fonction du niveau de gris affiché. . . . .	143

3.1 Répartitions des sujets pour les expérimentation de contraste/luminance et de comparaison objective/subjective. . . . .	146
3.2 Déroulement global de l'expérimentation contraste/luminance. . . . .	148
3.3 Déroulement d'une session de mesure de temps de réactions pour une luminance de fond donnée. . . . .	149
4.1 Performance théorique et réelle pour l'expérimentation de détection des stimuli contrastés, conditions 0 et 32. . . . .	155
4.2 Performance théorique et réelle pour l'expérimentation de détection des stimuli contrastés, conditions 80 et 128. . . . .	155
4.3 Performance théorique et réelle pour l'expérimentation de détection des stimuli contrastés, conditions 176 et 255. . . . .	155



# LISTE DES TABLEAUX

2.1 Valeurs de la puissance dans les fonctions de transducteurs de [Watson & Solomon, 1997] . . . . .	60
3.1 Gamuts Coverage of 1931 Color Space . . . . .	92
3.2 Acuity of the eye, [Gross <i>et al.</i> , 2008] . . . . .	98
5.1 Force relative des critères du modèle non liés à la perception de la profondeur	114
5.2 Force relative des indices de perception de la profondeur. Table adaptée depuis [Mehrabi <i>et al.</i> , 2013] . . . . .	115
6.1 Dimensions du CAVE « P3I » dans lequel a eu lieu l'étude expérimentale sur la comparaison entre les notes objectives et des notes subjectives. . . . .	118
6.2 Population et âge des sous-catégories de l'étude expérimentale sur la comparaison entre les notes objectives et des notes subjectives. . . . .	119
2.1 Extrait des mesures de luminance globale du simulateur en fonction de la nuance de gris affichée. . . . .	136
2.2 Extrait des mesures de luminance du disque en fonction de sa nuance de gris et de la nuance globale affichée. . . . .	138
2.3 Mesure pupillaires en fonction de la luminosité . . . . .	141

3.1 Valeurs des nuances de gris du stimulus nécessaire pour afficher un contraste par valeur supérieur désiré. . . . .	152
3.2 Valeurs des nuances de gris du stimulus nécessaire pour afficher un contraste par valeur supérieur désiré. . . . .	152
4.1 Temps de réaction théoriques (en ms) prédits par le modèle de Rea et Ouellette.	154
4.2 Temps de réaction moyens réels (en ms) mesurés pendant l'expérimentation et leurs écart-types moyens. . . . .	156
1 Luminance globale du simulateur en fonction de la nuance de gris affichée. . . . .	g
2 Luminance de la cible (en $cd/m^2$ ) en fonction de sa nuance de gris et de la nuance de gris du fond. . . . .	h

# INTRODUCTION GÉNÉRALE

**L**OREM ipsum dolor sit amet, consectetur adipiscing elit. Proin quis ipsum tellus. Cras placerat nulla sit amet urna iaculis sagittis. Curabitur ullamcorper leo est, at tempor nunc malesuada vitae. Sed ullamcorper tempus vehicula. Integer sit amet nulla ac lectus vulputate bibendum sed in arcu. Suspendisse potenti. Sed dapibus velit quis auctor pretium. Ut dictum, nisl ut tincidunt aliquam, odio orci dignissim libero, eu maximus dolor justo ut lorem. Duis accumsan sed dolor id elementum. Maecenas erat turpis, tincidunt quis sem vestibulum, tempor sagittis turpis. Vestibulum luctus congue nisl, non porttitor felis. Pellentesque eget nisl et nisl fringilla efficitur sed id nibh. Maecenas volutpat dolor ac lorem iaculis ultricies.

Nulla consectetur elit ac ante aliquam condimentum. Donec eu auctor turpis. Quisque purus ligula, viverra ut luctus quis, ultricies ac leo. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nam vel est enim. Suspendisse elementum fermentum blandit. Curabitur volutpat et turpis in condimentum. Proin dignissim nec nisl sit amet accumsan. Ut facilisis nulla at mi porttitor, nec venenatis mi pretium. Proin interdum aliquam odio, quis convallis nunc lobortis eget.

---

## INTRODUCTION GÉNÉRALE

---

## **Première partie**

# **État de l'Art : Partie Théorique**



## CHAPITRE

# 1

# DÉFINITION DU RÉALISME

## 1.1 Enjeux et problématiques

Qu'est-ce que le réalisme ? Comment le définir ? A quoi peut-il servir ? Peut-on faire *plus* de réalisme ? A quoi cela servirait-il ? Ces questions peuvent devenir des enjeux majeurs lorsqu'elles sont associées à des cas d'application réels et de portée économique importante. Un exemple flagrant se retrouve à l'endroit du constructeur français de voitures : Renault.

Les maquettes et prototypes numériques entrent de plus en plus dans le cycle de conception des véhicules. Les simulateurs et autres systèmes d'affichage immersif sont utilisés pour visualiser ces maquettes à l'échelle 1 :1, dans des conditions d'utilisation, d'éclairage et de ressemblance poussées à leur paroxysme. L'objectif est d'être le plus *réaliste* possible afin de légitimer au maximum et d'augmenter les domaines de prises de décisions dans les simulateurs.

Mais définir -et plus encore, quantifier- le *réalisme* n'est pas une tâche facile. Nous allons voir qu'il peut exister, voire coexister, plusieurs définitions applicatives du mot.

## 1.2 Définitions usuelles

Une fois l'évidence de la recherche de réalisme mise en avant, il devient nécessaire de définir ce qu'est le réalisme, sans quoi il serait impossible de le travailler. Tout d'abord, on peut s'intéresser à la définition que donne le Larousse<sup>2</sup> :

2. Réalisme. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à [www.larousse.fr/encyclopedie/divers/realisme/8600](http://www.larousse.fr/encyclopedie/divers/realisme/8600)

- Attitude qui tient compte de la réalité telle qu'elle est.
- Caractère de ce qui est une description objective de la réalité, qui ne masque rien de ses aspects les plus crus.
- Tendance littéraire et artistique du XIXe siècle qui privilégie la représentation exacte, tels qu'ils sont, de la nature, des hommes, de la société.
- Doctrine qui affirme que la connaissance du réel constitue le réel lui-même, que cette connaissance soit la seule réalité ou qu'à côté d'elle figure une autre réalité, l'objet auquel elle s'applique.

Mais ces définitions (notamment les n.3 et 4) semblent assez incomplète pour notre application car trop généraliste. Il faut s'intéresser à une définition qui soit spécifique au cadre de la réalité virtuelle (RV). Mais le *réalisme* est un terme très ambivalent et difficile à cerner [Burkhardt, 2003]. C'est ce que nous allons étudier dans la prochaine sous-partie.

### 1.3 Définitions dans le domaine de la Réalité Virtuelle

La première définition qui vient à l'esprit est celle de la ressemblance graphique avec le monde réel [Ferwerda *et al.*, 1996]. Est réaliste ce qui semble vrai, ce qui donne l'impression d'être vrai. On s'autorise alors l'utilisation d'un certain nombre d'artefacts graphiques (souvent issus de et popularisés par le cinéma), tel les bokehs<sup>1</sup> (cf. Figure 1.1) ou le tonemapping (cf. plus bas), afin de renforcer le sentiment d'apparence réelle. Ce type de réalisme est notamment développé par et mis en avant dans le monde du jeu vidéo avec des textures de plus en plus sophistiquées, des modélisations de plus en plus fines (et donc lourdes en polygones) mais surtout des jeux de lumière et d'optique saisissants.

Une deuxième approche, moins évidente au premier abord, est décrite chez Hoorn [Hoorn *et al.*, 2003] : une environnement virtuel sera assez réaliste s'il remplit sa fonction ; si, dans le cadre d'une application d'apprentissage par exemple, ce qui doit être appris est appris. Cette définition se détache complètement d'une volonté d'un réalisme d'apparence (comme on a pu voir précédemment) et peut même se détacher d'une construction logique et/ou anthropomorphe de l'environnement virtuel. Si une scène faite de volumes géométriques colorés flottants dans un espace monochrome suffit à transmettre le message voulu, la scène est jugée réaliste.

On trouve, dans [Fuchs *et al.*, 2003], cinq acceptations du terme *réalisme* dans le domaine de la RV :

- Évaluation subjective du degré de ressemblance d'une situation.
- Fidélité de construction de la simulation aux lois de la Nature.

---

1. Les bokehs (terme provenant du japonais pour flou) sont des artefacts graphiques ressemblant à des taches de flou et qui prennent la forme du diaphragme de l'objectif de l'appareil de capture (caméra ou appareil photo)

---

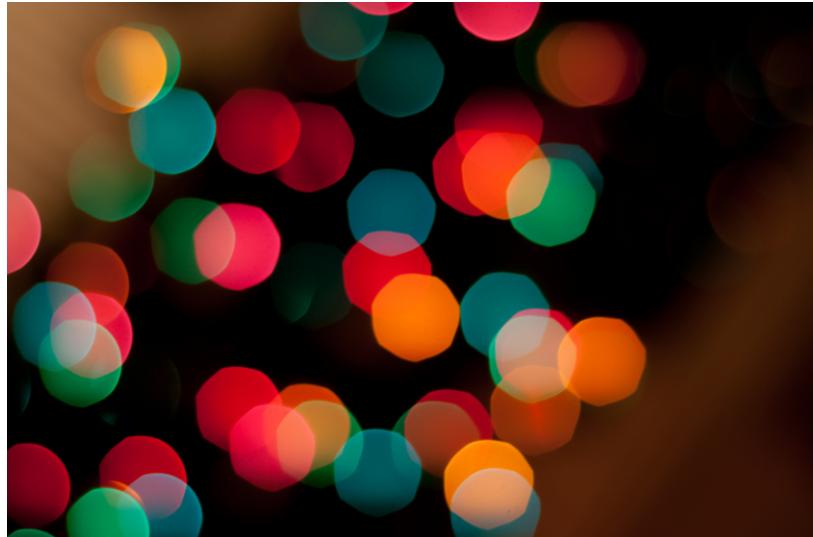


FIGURE 1.1 – Illustration de bokeh

- Évaluation subjective du degré de crédibilité d'une situation.
- Fidélité psychologique.
- Illusion du réel (connue aussi sous l'appellation *présence*).

La première acceptation, l'évaluation subjective du degré de ressemblance, a déjà été décrite dans le paragraphe précédent. Elle recouvre toute la partie esthétique et artistique de la conception de l'environnement virtuel.

La deuxième acceptation, la fidélité de la simulation aux lois de la Nature, accorde moins d'importance à la beauté de la simulation qu'à son comportement simulé ; en dépit de ce que peut juger l'utilisateur. Chaque fonctionnalité implémentée se comporte en suivant un modèle issu du monde physique réel. De par le biais de perception qui peut être induit dans un monde virtuel, les comportements (comme par exemple la gravité, l'adhérence sur un sol, etc ...) et/ou les apparences (rendu des couleurs, floutage des images en fonction de la vitesse, ...) peuvent apparaître comme non naturels à l'utilisateur. Son avis n'est pas pris en compte, on se contente d'afficher des modèles rigoureux.

La troisième acceptation, est la création d'une expérience perceptive qui serait crédible, tant sur le plan microscopique (chaque élément pris à part) que sur le plan macroscopique (l'environnement virtuel dans sa globalité). Même si les informations qui sont envoyées à l'utilisateur (via l'interface de l'environnement virtuel) ne suivent pas rigoureusement les lois de la physique et du monde *réel*, il faut que l'observateur les perçoive comme étant vraies.

On peut supposer un lien fort et systématique entre cette dernière acceptation (perception crédible) et la précédente (construction objective via des modèles déterminés) mais cette question est encore grande ouverte et sujette à recherche.

Ensuite, la quatrième acceptation du terme *réalisme* est liée à une fidélité d'ordre psychologique. L'utilisateur doit se comporter de la même manière dans l'environnement virtuel

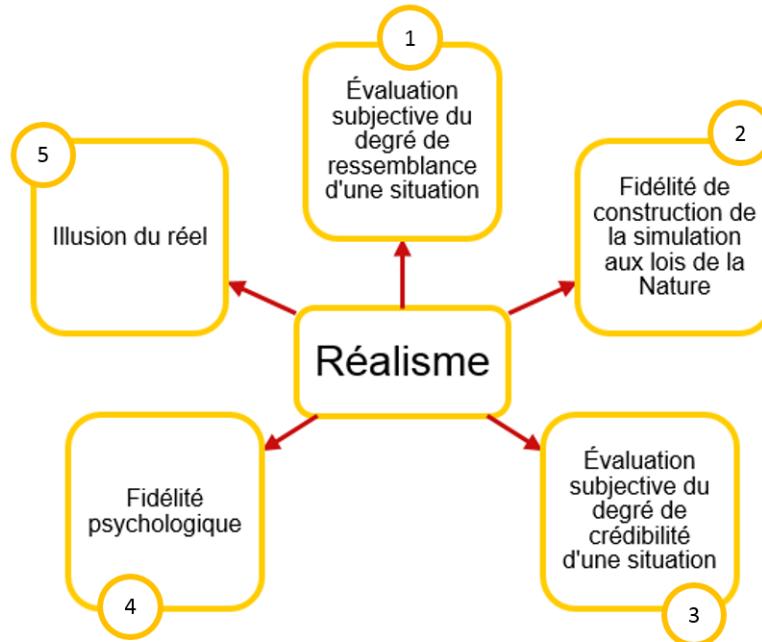


FIGURE 1.2 – Différentes acceptations du terme *réalisme*

que dans la même situation dans le monde physique. Les images affichées peuvent ne pas correspondre avec ce que l'utilisateur verrait dans la situation réelle (en fonction de la tâche à réaliser) tant que les performances et les résultats du sujet dans l'environnement virtuel sont similaires ou identiques à celles/ceux enregistré(e)s dans le monde réel. Cette acceptation est largement discutée dans la littérature [Burkhardt, 2003; Patrick, 1992; Stoffregen *et al.*, 2003].

Enfin, la cinquième et dernière acceptation fusionne les concepts de *réalisme* et de *présence*. Plus le sentiment de présence dans un environnement virtuel sera fort, plus le réalisme dudit environnement le sera aussi. La présence est "l'illusion d'une réalité qui n'existe pas" [Stoffregen *et al.*, 2003], c'est à dire le fait de faire croire au cerveau que les images/objets/personnes virtuelles que l'on voit sont en fait bien réels ou vivants [Burkhardt, 2003]. Plus la croyance en la réalité de la scène est forte, plus la présence est forte.

## 1.4 Cadre d'étude

Pour le cas de cette thèse, la définition du réalisme qui aura été retenue -et qui sera sous-entendue quand on utilisera le mot *réalisme* seul- est celle de la proximité physiologique avec le système visuel humain.

Le but de cette thèse n'est pas de construire de nouveaux modèles esthétiques et de travailler à l'amélioration graphiques des simulateurs (objectif qui incombe plutôt à un designer ou à un graphiste), mais plutôt de travailler sur des modèles de vision, c'est à dire

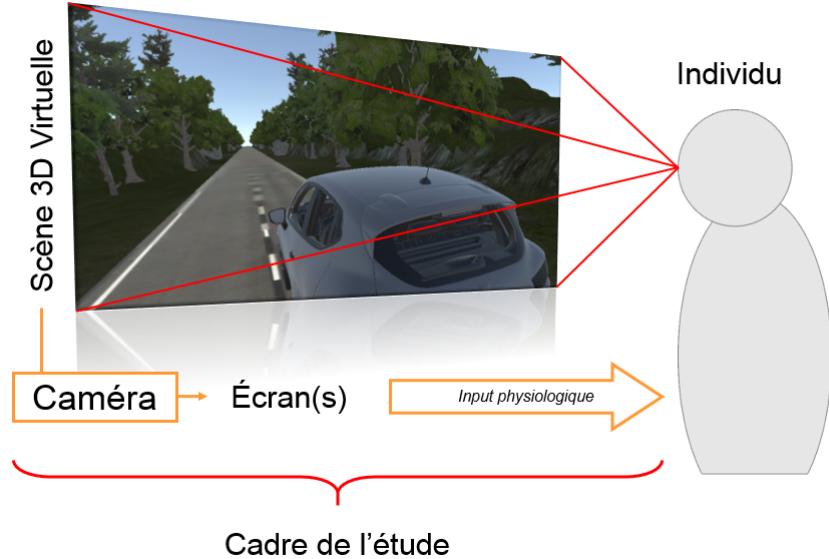


FIGURE 1.3 – Cadre d'étude de la thèse

la manière dont la caméra virtuelle va extraire les informations de la scène virtuelle -par opposition aux modèles d'affichage qui décrivent comment les informations capturées par la caméras doivent être affichées sur le(s) écran(s)- qui soient proche des facultés de la vision humaine.

Le réalisme physiologique doit, par construction et par définition, se baser sur le système visuel humain. Dans la suite de ce chapitre, nous passerons en revue un grand nombre de paradigme de la vision ; que ce soit dans la modélisation, dans le fonctionnement ou dans les différentes perceptions du monde qui nous entoure.

---

## CHAPITRE 1. DÉFINITION DU RÉALISME

---

## CHAPITRE

# 2

## DÉFINITION DE LA MULTI-SENSORIALITÉ 3D

### 2.1 Immersion

Deux sémantiques coexistent : l'immersion est à la fois l'action d'immerger un utilisateur dans un environnement complètement virtuel via des images de synthèse ; mais c'est aussi l'effet (avéré ou non) qu'a cette immersion sur ce même utilisateur. De manière plus formelle, on peut dire que l'immersion est le « degré et qualité, par l'interface [d'un] système, du contrôle des entrées sensorielles pour chaque modalité de perception et d'action » (depuis [Fuchs *et al.*, 2003]).

Pour Burkhardt, [Burkhardt *et al.*, 1999], le degré d'immersion est caractérisé par un ensemble de grandeurs :

- Le sous ensemble des modalités mises en œuvre dans l'interaction.
- Les propriétés des dispositifs d'interaction pour chacune des modalités visées (degré de complétude, qualité, paramètres du signal, ...)
- La cohérence interne et la latence globale de l'information et des réactions délivrées en temps réel par le système.
- Les propriétés physiques de l'environnement physique dans lequel se déroule l'expérience.

## 2.2 Présence

qsd

## 2.3 Proprioception

qsd

## 2.4 Application dans la thèse

Dans le cadre de la thèse, le sens de « multi-sensorialité » n'est donc pas l'usage des plusieurs des cinq sens traditionnels (ouïe, vue, odorat, toucher et goût) mais plutôt la prise en compte de la temporalité de l'expérience et de la capacité de l'observateur de bouger dans son environnement, au bénéfice de la vision. Par exemple, la parallaxe de mouvement.

# CHAPITRE

# 3

## FONCTIONNEMENT DE LA VISION

Première interface et premier organe de la chaîne, l'œil -à contrario de ce qui est généralement pensé- n'est pas la pièce maîtresse de la vision : il ne sert *qu'à* envoyer des signaux électriques au cerveau et pourrait être remplacé par une caméra par exemple, dans la mesure où celle-ci émet les bonnes informations à travers le nerf optique [Dobelle, 2000]. L'œil reste néanmoins un outil très puissant et versatile. Nous allons ici rappeler le fonctionnement de l'œil et de la vision, à travers les informations collectées et transmises par l'œil puis traitées par le cerveau. Le sujet est déjà tout à fait connu et largement documenté [Driscoll *et al.*, 1978; Fairchild, 2005; Gross *et al.*, 2008; Wandell, 1995] aussi nous donnerons plutôt des références sur les parties plus pointues.

### 3.1 Structure de l'œil

L'œil se divise en un certain nombre de parties dont les principales sont la cornée, l'iris, la pupille, le cristallin, la rétine et les deux chambres contenant les humeurs aqueuse et vitrée. Toute la structure de l'œil est résumée sur la Figure 3.1.

La cornée est la première interface entre l'extérieur et l'intérieur de l'œil.

Le cristallin est la lentille principale de l'œil et va s'occuper de la netteté du monde qui nous entoure via le processus d'accommodation. Le principe est simple : il faut placer le foyer image de l'œil sur la rétine (cf. Section suivante). Pour ce faire, l'œil accorde, c'est à dire qu'il va modifier sa géométrie dans le but de modifier ses propriétés optiques. Dans la pratique, les muscles ciliaires font déformer le cristallin (cf. Figure 3.1) en réduisant sa hauteur et donc en augmentant sa convexité ou inversement. Ce processus dure 1 seconde et n'est pas définitif : une fois l'œil accorde, celui-ci oscille légèrement autour de la valeur d'accommodation, à une fréquence de 5 Hz. Ce mécanisme permet de récupérer des feedbacks en cas d'ajustements à opérer [Gross *et al.*, 2008].

Le mécanisme d'accommodation ne fonctionne plus à partir d'un certain seuil de luminosité. Ce seuil est évalué à  $0.01\text{Cd}/\text{m}^2$ . En l'absence d'accommodation, l'œil prend une position intermédiaire entre la position relâchée (accommodation à l'infini) et une position presque-accommodée [Gross *et al.*, 2008].

La pupille et l'iris sont intimement liées car la première est l'espace laissé par la seconde au centre de l'œil. L'iris est une composante contrôlable de l'œil. L'iris contrôle la quantité de lumière qui arrive dans l'œil et donc sur la rétine. Plus la pupille sera grande, plus la lumière pourra rentrer dans l'œil. Un œil humain peut supporter des luminosité (luminance) allant de  $10^{-6}$  à  $10^5\text{Cd}/\text{m}^2$ . Le diamètre de la pupille est régit par la l'équation Eq. 3.1. De la même manière que l'accommodation, le diamètre de la pupille oscille autour de sa valeur moyenne à l'instant t. Ce phénomène est appelé *hippus* [Gross *et al.*, 2008].

$$\log_{10}(D_{iris}) = 0.8558 - 0.000401 \cdot [8.4 + \log_{10}(L)]^3 \quad (3.1)$$

De la même manière, la vergence est la capacité des yeux de s'orienter vers le point d'accommodation lorsque celui est proche. On appelle convergence le phénomène qui consiste à augmenter l'angle formé par l'intersection des lignes du regard de chaque œil (quand la cible se rapproche) et divergence le cas inverse, lorsque cet angle diminue (la cible s'éloigne). Le processus possède une latence estimée à 150 ms et une durée approximative évaluée à 0.2 - 0.6 secondes [Devisme, 2004; Gross *et al.*, 2008].

L'œil humain a une taille globalement constante entre les individus : autour de 24 mm de diamètre [Glassner, 1995]. Le pouvoir optique de l'œil, c'est à dire sa capacité à adapter ses propriétés optiques, se mesure en dioptries. Les dioptries ( $\delta$ ) sont l'inverse de la distance focale d'un système optique, elles sont homogènes à des  $\text{m}^{-1}$ . Faire varier ses dioptries (et donc sa distance focale) permet l'accommodation et l'adaptation de l'œil. L'œil humain possède environ  $42\delta$  en fonctionnement et peut monter jusqu'à une puissance de 60 à  $80\delta$  pour compenser les défauts de l'œil [Glassner, 1995].

La rétine tapisse le fond de l'œil et est composée de millions de récepteurs photosensibles qui vont être responsables de la captation de l'image. Ces photo-récepteurs sont les cônes, sensibles à la couleur, et les bâtonnets sensibles à la luminosité. Il existe 3 types de cônes, qui sont chacun sensibles à différentes longueurs d'onde : on retrouve les cônes de type S (pour *small*, petite longueur d'onde) avec un maximum de sensibilité à 420 nm ; ensuite, on trouve les cônes de type M (pour *medium*, moyenne longueur d'onde) avec un maximum de sensibilité à 530 nm ; et enfin, on a les cônes de type L (pour *long*, grande longueur d'onde) avec un maximum de sensibilité à 560 nm. Le seul message émanant d'un cône ou d'un bâtonnet est celui de son activation. La reconnaissance des couleurs et de l'intensité de la lumière se fait par la combinaison des résultats d'activation des différents types de cône ; ce procédé est analogue à celui du tramage<sup>1</sup> en informatique. L'œil humain ne possède que trois types de cône, alors qu'on pourrait penser qu'il y aurait un cône dédié à chaque couleur, et cela résulte d'une optimisation du nombre de cônes c'est à dire de la capacité de l'œil à extrapoler la couleur à partir du minimum de données, et la densité de ces capteurs.

1. Tramage : Procédé permettant de générer de nouvelles couleurs à partir d'une base limitée de couleurs. Typiquement, en informatique, les pixels des écrans sont en fait composés de 3 sous pixels : blanc, rouge et vert. Une fois combinés et vu d'assez loin, ils semblent ne créer qu'une seule couleur différente.

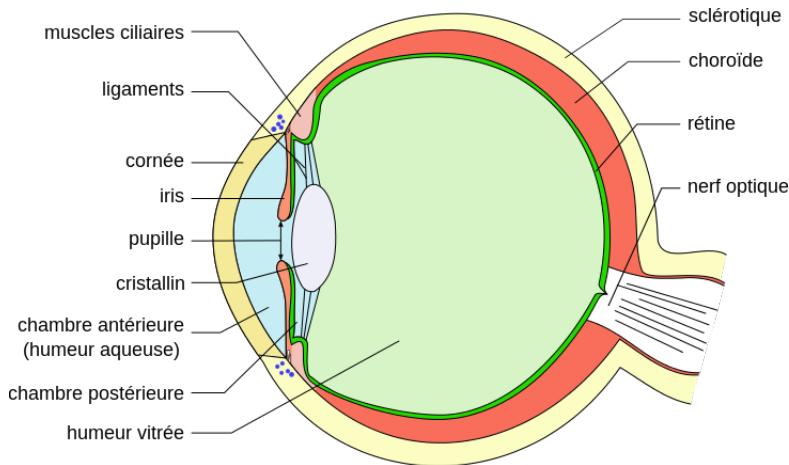


FIGURE 3.1 – Structure de l’œil

En effet, si l’œil possédait de nombreux autres types de cônes, la densité par cône serait grandement réduite et donc la finesse de vision aussi [Glassner, 1995].

Les cônes et les bâtonnets ont chacun leur plage de fonctionnement optimale. Dans le domaine photopique (lumière du jour), les cônes sont saturés en luminosité et sont beaucoup moins efficaces que dans le domaine scotopique (de nuit). De leur côté, les cônes adaptent leur maximum de saturation par rapport au niveau global d’illumination [Glassner, 1995].

Enfin, la répartition des cônes et des bâtonnets sur la rétine n’est pas du tout homogène (cf. Fig. 3.2). Une zone de la rétine présente une extrême concentration de cônes. Cette zone est appelée *fovée*. De par l’utilité des cônes et par construction du système optique de l’œil, c’est l’endroit où les rayons lumineux issus de la cible regardée convergent. En périphérie de la *fovée* on retrouve les bâtonnets. Ce qui est vu dans la *fovée* est net, tandis que ce qui est vu par les bâtonnets ne l’est pas. L’axe de vision est d’ailleurs défini par le rayon issu du point nodal image jusqu’à la *fovée* (cf. section suivante). Cet axe, représenté sur la Figure 3.5 est donc fixe et l’angle qu’il forme avec l’axe optique de l’œil (ici appelé *alpha*) vaut entre 3 et 8 degrés suivant les individus [Gross *et al.*, 2008].

La rétine n’est que la première étape du système d’acquisition et la totalité de celui-ci ainsi que le traitement par le cerveau sera traité plus loin dans le chapitre.

## 3.2 Modélisation de l’œil

De nombreux scientifiques se sont intéressés à la compréhension et la modélisation de l’œil, et ce, dès la moitié du XIX<sup>e</sup> siècle avec Carl Friedrich Gauss qui en a définit les grands principes optiques. Cependant, avant de discuter des différents modèles de l’œil qui ont pu émerger au cours (notamment) du XX<sup>e</sup> siècle, il paraît nécessaire de reposer certaines bases élémentaires de l’optique géométrique.

On note quelques points particuliers dans un système optique simple (type lentille concave ou convexe). De manière générale, on rappelle que les points caractéristiques d’un lentille

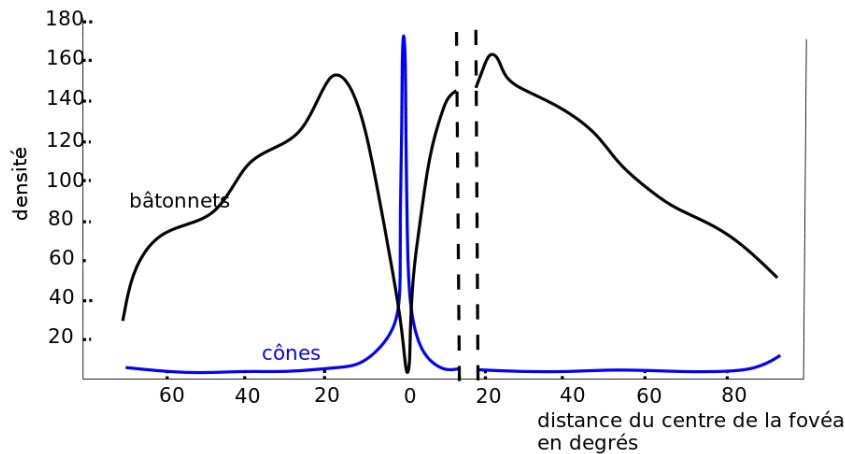


FIGURE 3.2 – Répartition (en *milliers/mm<sup>2</sup>*) des cônes/bâtonnets sur la rétine

sont qualifiés de *objet* quand ils sont avant la lentille (en terme de trajet du rayon lumineux) et de *image* lorsqu'ils sont après la lentille. Par convention, un point caractéristique est noté avec une lettre majuscule tandis que son point conjugué image est noté avec la même lettre agrémentée d'un *prime*.

- Le foyer objet est le point à partir duquel les rayons issus sont ensuite dirigés vers l'infini après passage par la lentille. Le foyer objet est noté F.
- Le foyer image est le point vers lequel convergent tous les rayons issus de l'infini avant la lentille. Dans le cas d'un œil sain (sans problèmes ophtalmologiques de type myopie ou hypermétropie), ce point doit être confondu avec la rétine pour avoir une image nette. Comme on a pu voir précédemment, c'est là le principal mécanisme de l'accommodation qui va déformer le système optique pour déplacer les foyers afin de faire la netteté sur la rétine. Le foyer image est noté F'.
- Le point central, par lequel un rayon passe sans être dévié, qui correspond au centre de rotation de l'œil. Il est noté C.
- Le point nodal objet est un point très particulier parce qu'il peut être considéré comme l'équivalent du *point de vision*, c'est à dire le point où on pourrait ramener l'intégralité de l'œil à un équivalent ponctuel. Dans un système optique, il est défini comme le point par lequel passe un rayon lumineux avec une incidence donnée et ressort au point conjugué avec la même incidence. Il n'existe que dans un système optique complexe (fait de plusieurs lentilles ou interfaces optiques), dans le cas d'une lentille simple seule il est confondu avec le centre. Ce point est globalement situé 6 mm à l'avant du centre de l'œil, sur l'axe optique [Gross *et al.*, 2008; Ogle, 1968]. Les points noraux sont notés N et N'.

L'œil possède deux axes principaux : l'axe optique, qui passe par les centres de toutes les surfaces optiques ou composantes de l'œil (cristallin, pupille, ...) et l'axe de vision qui est composé en fait de deux demi-axes que sont les rayons qui passent par les points noraux

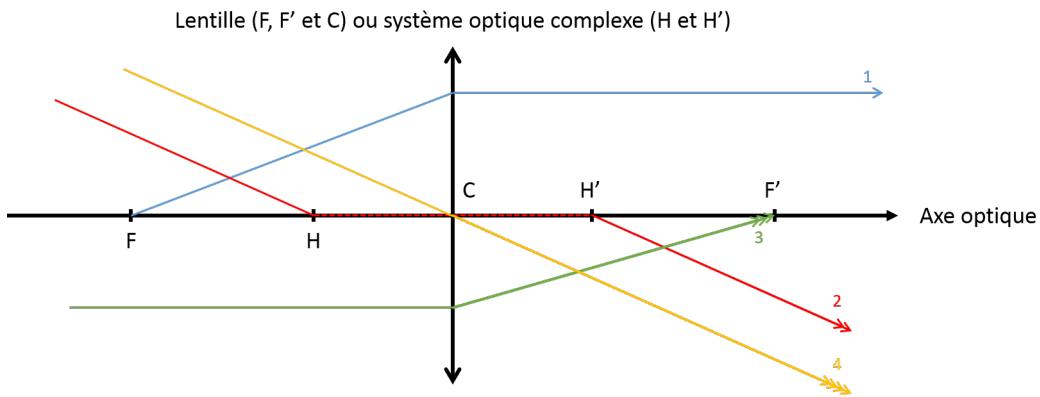


FIGURE 3.3 – Points caractéristiques (cardinaux) d'un système optique plan

et jusqu'à la fovéa. L'œil est orienté selon l'axe optique mais la vision se fait le long de l'axe optique. La vision hors axe optique est appelée *vision périphérique*.

Enfin, un grand nombre de modèle de l'œil ont été proposés, plus ou moins complexes, avec notamment des variations sur le nombre de surfaces optiques, jusqu'à l'adoption du modèle de [Liou & Brennan, 1997]. Malgré tout ce modèle est tout sauf définitif et est toujours susceptible d'être amélioré ou affiné dans le futur. Voici une liste pas tout à fait exhaustive des différents modèles de l'œil qui ont pu être élaborés au cours du XXème siècle [Gross *et al.*, 2008; Liou & Brennan, 1997] :

- Modèle de Helmholtz - Laurence (1909)
- Modèle de Gullstrand (1911) : modèle le plus utilisé, notamment car un grand nombre de distances (taille de l'œil, distances focales, distance de la pupille, ...) et tous les indices de réfractions des différents milieux de l'œil y sont reportés (cf. Figure 3.4). Le modèle optique théorique pour la propagation de la lumière est composé de 3 surfaces.
- Modèle de Emsley (1946) : modèle simplifié réduit à 1 seule surface.
- Modèle de Lotmar (1971) : modélisation de la cornée et de la face arrière du cristallin par des surfaces polynomiales (plutôt que des sections de sphère).
- Modèle de Kooijman (1983) : modèle en 4 surfaces, ajout d'asphéricités sur les surfaces sphériques.
- Modèle de Navarro (1985) : idem que Kooijman avec des effets chromatique type dispersion de la lumière en plus.
- Modèle de Schwiegerling (1995)
- Modèle de Liou & Brennan (1997) : modèle à 5 surfaces dont 1 purement théorique (cf. Figure 3.5). C'est le modèle le plus utilisé dans le domaine du calcul et de la simulation de rayons.

Jusqu'à présent, nous avons détaillé le fonctionnement d'un œil seul, nous allons maintenant nous intéresser à la vision binoculaire, c'est à dire de deux yeux en même temps.

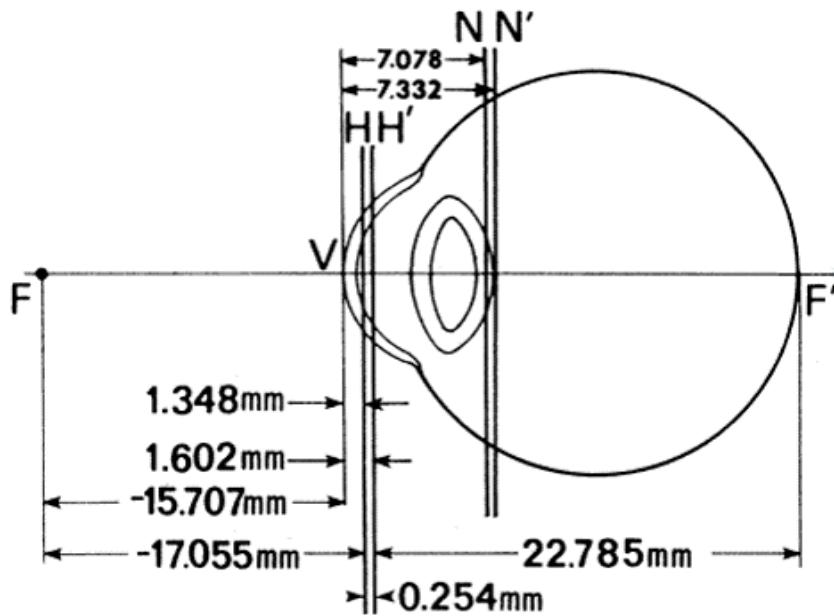


FIGURE 3.4 – Modèle et valeurs physiologiques de Gullstrand, image tirée de [Liou & Brennan, 1997]

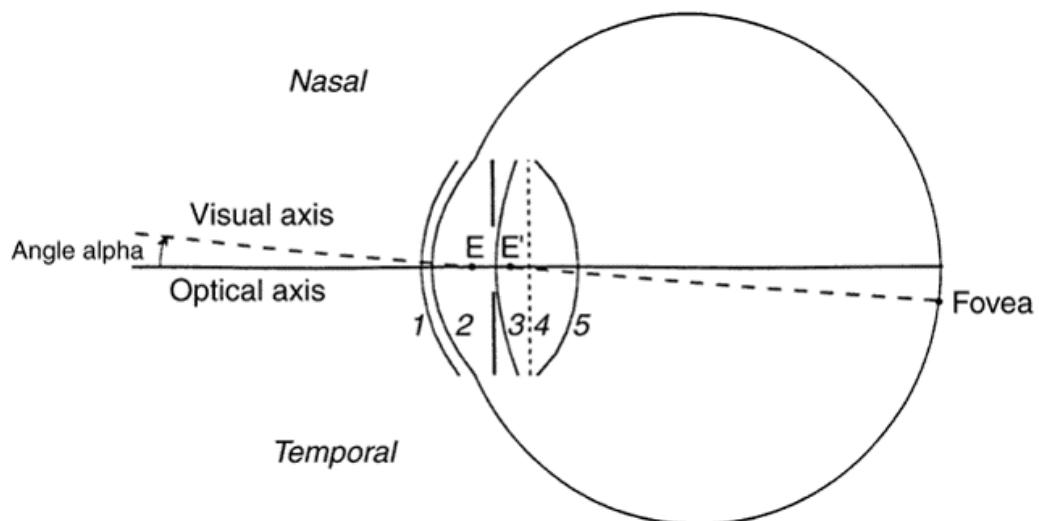


FIGURE 3.5 – Modèle de Liou & Brennan, image tirée de [Liou & Brennan, 1997]

### 3.3 Vision binoculaire

La vision binoculaire est avant tout un excellent outils de compréhension de l'environnement qui nous entoure, et notamment de la profondeur : le cerveau peut comparer en temps réel deux points de vue légèrement différents et en tirer des conclusions.

L'élément de fonctionnement de la vision binoculaire est les disparités, c'est à dire la différence de position relative entre les deux images perçues (une par œil) d'un même objet. Si un objet est pleinement centré sur l'image perçue par l'œil gauche, il sera vu légèrement décalé du centre par l'œil droit. Il existe deux types de disparités ; les disparités verticales et les disparités horizontales [Devisme, 2004].

Les disparités horizontales sont générées par une différence angulaire sur le plan de l'azimut. Elles sont en charge de la perception relative de la distance, c'est à dire à la perception de la profondeur (relief).

Les disparités verticales sont issues de la perception par les yeux d'une différence d'élévation pour le même point. Elles peuvent aussi apparaître si un des yeux est plus proche de l'objet que l'autre (on parle alors de grossissement différentiel). Les disparités verticales permettent d'estimer la distance absolue d'un objet ainsi que l'excentricité d'une surface, indépendamment de son orientation.

Les disparités peuvent être source de fatigues dans le système visuel humain. Ce problème a été adressé notamment en ajoutant du flou sur les disparités les plus grandes, en ajoutant un flou périphérique pour reproduire le flou rétinien et en proposant un modèle de caméras convergentes qui génère des disparités plus conformes à la réalité [Aurat, 2016].

L'horoptère est le lieu des points de disparité horizontale nulle pour un point de fixation donné, c'est à dire, à un instant  $t$  donné, lorsqu'on regarde à un endroit donné, il existe une infinité de points pour lesquels aucun disparité n'est perçue : ils sont au même endroit sur l'image perçue par l'œil gauche et par celle perçue par l'œil droit. Théoriquement, le lieu de ces points est un cercle passant par le point de fixation et par le premier point nodal (point nodal objet,  $H$ ) de chaque œil ; ce cercle s'appelle le Cercle de Vieth-Müller. Dans la pratique, le lieu de l'horoptère n'est pas tout à fait un cercle et présente une légère déviation, nommée Déviation de Hering-Hillebrand (cf. Fig. 3.6) [Neveu, 2012].

Si on prend en compte les disparités verticales, le lieu des points de disparité horizontale-verticale nulle devient un cylindre (Extension du Cercle de Vieth-Müller).

L'œil possède une certaine capacité à adapter sa puissance (comme vu précédemment). On retrouve cette capacité au niveau du cerveau pour la fusion stéréoscopique. Dans une zone proche autour de l'horoptère (Aire de Panum) le cerveau écrase les disparités et fusionne quand même les images perçues par les yeux. L'étendue et la forme de l'aire de Panum est grandement dépendante des caractéristiques locales du stimulus [Devisme, 2004].

Enfin, la vision binoculaire est tributaire d'une certaine acuité stéréoscopique, c'est à dire la capacité à percevoir un écart de profondeur entre deux plans à une distance donnée. Celle-ci est présentée et son modèle démontré à plusieurs endroits [Fuchs *et al.*, 2003; Gross *et al.*, 2008] et sera abordée dans un chapitre suivant. Toutefois, si l'âge et le niveau

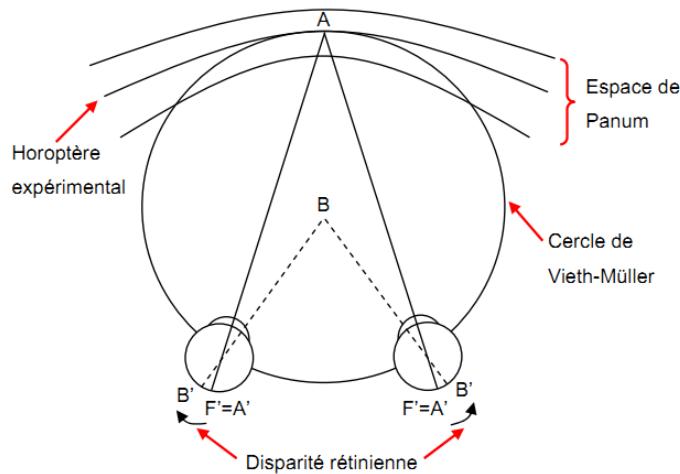


FIGURE 3.6 – Horoptères théorique et empirique, Aire de Panum et disparités rétiniennes. Image tirée de [Neveu, 2012].

d'accommodation n'ont pas l'air d'affecter l'acuité stéréoscopique, cette dernière semble dépendante du contraste et de la quantité d'éclairement rétinien [Devisme, 2004].

### 3.4 Traitement post-rétinien

Nous allons maintenant aborder toute la partie transmission et traitement par le cerveau du signal émis par les rétines. Le comportement précis dans le cerveau est encore à ce jour une question ouverte que nous traiterons donc dans les limites du possible et de l'utile dans le cadre de cette thèse. On ne peut ici que conseiller l'excellent *Neurosciences : A la découverte du Cerveau* de [Bear *et al.*, 2007].

Tout d'abord, il convient de développer quant à la structure même des cônes, de quelle manière ceux-ci jouent leur rôle de transducteur<sup>1</sup> : transformant un signal lumineux entrant en signal électrique sortant.

La rétine est en fait composée de plusieurs types de cellules intermédiaires, les cônes et les bâtonnets n'étant qu'en bout de chaîne : cellules ganglionnaires, cellules amacrines, cellules horizontales et cellules bipolaires (cf. Fig. 3.7) [Bear *et al.*, 2007]. Ces cellules se réunissent ensuite pour former le nerf optique et sortir de l'œil. Parmi les cellules ganglionnaires il existe plusieurs types, chaque type véhiculant des informations différentes [ANSES, 2014].

La voie magnocellulaire (M) représente 5% de la population de cellules ganglionnaires. Ces cellules sont sensibles aux contrastes de luminance et ont une vitesse de conduction plutôt rapide. Elles sont reliées à la voie dorsale (cf. ci-après).

La voie parvocellulaire (P) représente 90% de la population de cellules ganglionnaires. Ces

1. Transducteur : Dispositif assurant une conversion ou un transfert de signaux et dans lequel un signal au moins est de nature électrique. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/transducteur/79088>

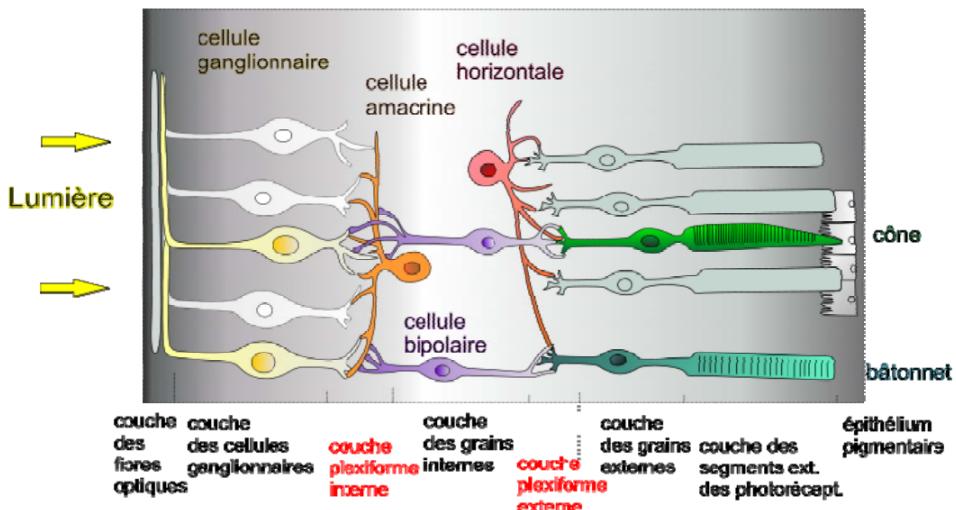


FIGURE 3.7 – Structure et cellules composant la rétine, image tirée de [ANSES, 2014].

cellules sont sensibles aux couleurs et, malgré une vitesse de conduction plus lente que les cellules de la voie M, elles répondent de manière tonique aux stimulations. Ces cellules sont à rapprocher de la voie ventrale.

Enfin, il existe une voie M-non P qui représente les 5% de population restants et participent à d'autres tâches que la vision pure (cf. ci-après).

Une fois le signal lumineux transformé dans une des trois voies (M, P ou M-non P), l'information circule dans le nerf optique en direction du cerveau. Les nerfs optiques de chaque œil se croisent au niveau du chiasma optique avant de continuer leur chemin dans le tractus optique et d'arriver au cerveau. Au niveau du chiasma, les fibres porteuses des informations visuelles sont redirigées : les fibres des deux moitiés gauches des yeux sont envoyées à la partie droite du cerveau, tandis que les fibres des deux moitiés droites des yeux sont envoyées à la partie gauche.

L'interface entre le cerveau à proprement parler et le tractus optique s'appelle le CGL (Corps Genouillé Latéral). Le CGL est composé de 6 couches qui vont permettre la distribution de l'information dans les différents cortex visuels. Les informations issues de la voie M sont attribuées aux couches 1 et 2 du CGL, tandis que les informations de la voie P sont attribuées aux couches 3, 4, 5 et 6 du CGL. Les 5% d'informations restantes (la voie M non P) ne va pas dans le CGL (et donc dans le cortex d'analyse de l'image) mais se dirige dans la mésencéphale et l'hypothalamus. Ces deux dernières aires du cerveau sont dédiées à la gestion du reste du corps humain : régulation du bio-rythme, sécrétion des hormones du sommeil, gestion de l'attention, ...

La radiation optique permet de faire le lien, via des neurones, entre les 6 couches du CGL et le Cortex Visuel Primaire (V1) qui marque l'entrée dans le cerveau à proprement parler.

Le traitement et la compréhension véritable des informations qui ont transité de l'œil jusqu'au CGL se fait dans les cortex visuels. La cartographie et la compréhension du fonctionnement en cortex dans le cerveau fait l'objet de plusieurs théories. Aux côtés des modèles

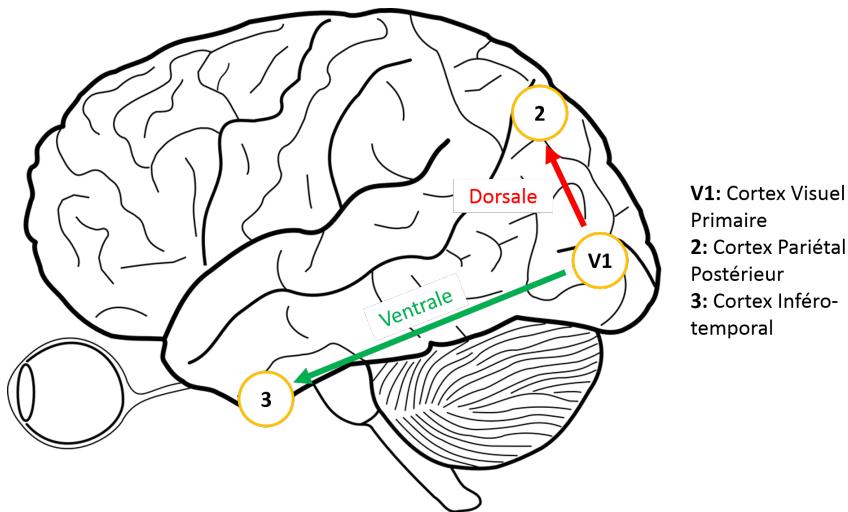


FIGURE 3.8 – Direction des voies ventrale et dorsale dans le cerveau.

hiérarchique et des agrégats, c'est l'hypothèse des deux voies [Goodale *et al.*, 1991; Mishkin *et al.*, 1983; Ungerleider & Mishkin, 1982] qui prédomine. Cette hypothèse est aussi étendue au fonctionnement de l'audition.

Le traitement de l'image dans les cortex visuels est décomposé en deux voies, à l'image des voies P et M. A l'arrivée dans le Cortex Visuel Primaire (V1), l'information est divisée en deux boucles de traitement indépendantes : la voie dorsale (appelée aussi pariétale) et la voie ventrale (appelée aussi temporelle). La voie dorsale correspond au traitement du mouvement et de la position (« Where »), elle est constituée des cortex V1, V2, V3, V3A, MT et MST avant d'arriver dans le Cortex Pariétal Supérieur. Cette fois fonctionnement à une vitesse plutôt lente. De l'autre côté, la voie ventrale s'occupe de la gestion des formes et des couleurs (« What ») et passe par les cortex V1, V2, VP, V4 et V8 avant d'arriver dans le lobe temporal. Cette voie fonctionne à grande vitesse [D'hondt, 2011; James & Kim, 2010].

L'information visuelle n'est pas transmise au cerveau telle qu'elle sous la forme de quatre informations (bâtonnets, cônes S, cônes M, cônes L). Ces signaux de sortie des photorécepteurs sont additionnés ou soustraits les uns entre les autres pour donner naissance à trois canaux d'information qui eux seront traités par le cerveau. Cette hypothèse de fonctionnement -la *Théorie des processus antagoniques* proposée pour la première fois par Ewald Hering en 1872- est compatible avec la théorie de la vision trichromatique (*Théorie de Young-Helmholtz*, 1802 puis prouvée en 1861) puisqu'elle intervient immédiatement après dans le traitement de la lumière.

Après les photo-récepteurs, le signal lumineux est donc découpé en trois canaux [Glassner, 1995; Winkler, 1999] un canal achromatique (A) et deux canaux chromatiques (R/G pour le canal rouge-vert et B/Y pour le canal bleu-jaune). On retrouvera ces canaux plus tard dans la description des espaces colorimétriques en informatique. Le canal achromatique sert à coder la valeur de luminance et est issu de la somme des signaux émis par les cônes de type M et L,  $A \equiv M + L$ . Le canal R/G est quand à lui la différence entre les deux précédents canaux M et L ( $R/G \equiv M - L$ ) tandis que le canal B/Y est la différence entre les signaux émis par les cônes de type S et le canal achromatique :  $B/Y \equiv S - A$ . Ces mécanismes sont résumés dans la Fig. 3.9.

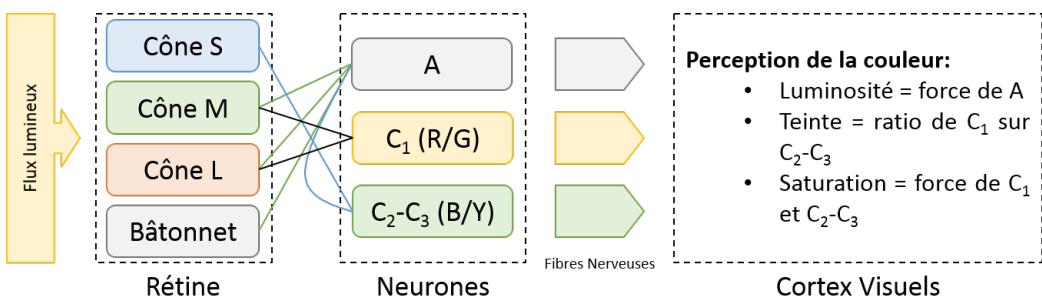


FIGURE 3.9 – Construction des canaux chromatiques et achromatique dans la théorie des processus antagoniques

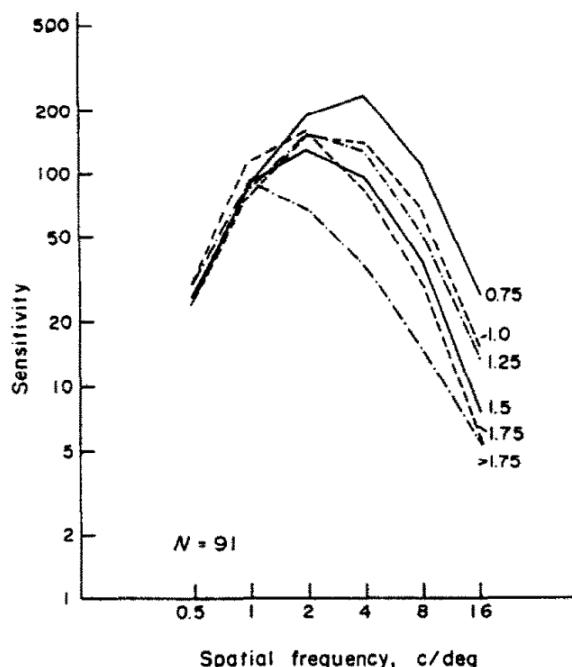


FIGURE 3.10 – Courbes moyennes de sensibilité au contraste en fonction de la fréquence spatiale du stimulus pour différentes valeurs d'acuité visuelle, tiré de [Owsley *et al.*, 1983]

Enfin, le cerveau dénote une certaine sensibilité au contraste (entendu ici comme une variation relative de luminance entre deux points de l'image perçue). Cette sensibilité dépend à la fois de la couleur (on a vu que le canal achromatique était en fait composé des retours de deux types de cônes) et à la fois des fréquences spatiales et temporelles du stimulus lumineux reçu. On peut alors établir des fonctions de sensibilité au contraste (Contrast Sensibility Functions - CSF) [Bezzubik & Belashenkov, 2015; Driscoll *et al.*, 1978]. La sensibilité au contraste est définie comme l'inverse du seuil de contraste, c'est à dire le minimum nécessaire de contraste suffisant à un observateur pour détecter une variation. Les variables *temps* et *espace* des CSF ne sont pas séparables. On peut trouver des exemples de CSF chez une grand nombre d'auteurs, comme ici avec les CSF moyennes en fonction de la fréquence spatiale du stimulus, pour différentes valeurs d'acuité visuelle (cf. Fig. 3.10) [Owsley *et al.*, 1983].



# CHAPITRE

## 4

# PERCEPTION DE LA COULEUR

La perception de la couleur a été théorisée via un certain nombre de concepts, lois ou effets ; on présente ici les principales contributions [Judd & Wyszecki, 1975; Le Grand, 1972; Wyszecki & Stiles, 2000].

### 4.1 Espaces colorimétriques

Il existe de très nombreux espaces colorimétriques (sRGB, YIQ, YES, YCC, LCh, TekHVC, NCS, Munsell, Coloroid, Pantone, CIECAM97, ... [Beretta, 2000]) mais une étude globale poussée sort très largement du cadre de l'étude. Nous nous concentrerons donc à évoquer rapidement les différents espaces principaux adoubés par la CIE.

La première tentative à avoir été déposée par la CIE, en 1931, est l'espace colorimétrique CIE RGB. C'est une première tentative de quantification de la couleur avec des primaires proches des maxima de réponse des cônes. La démarche est donc anthropomorphique et cherche à construire un espace linéaire. L'inconvénient (majeur) de ce système est l'obligation de passer par des composantes négatives pour certaines couleurs très saturées ; on s'éloigne alors du fonctionnement de la vision humaine qui est basée sur l'additivité des couleurs.

Les défauts de ce premier espace ont amené la CIE à proposer dans le même temps un autre espace, le CIE XYZ, lui aussi linéaire<sup>2</sup>. Dans ce modèle, Y donne la luminance tandis que X et Z donnent la chrominance et sont définis positifs par construction.

Mais la vision humaine des couleurs n'étant pas linéaire, un nouvel espace est adoubé

2. Linéaire : les valeurs des primaires varient linéairement, pas les couleurs perçues : une variation de primaire ne donne pas la même variation de couleur

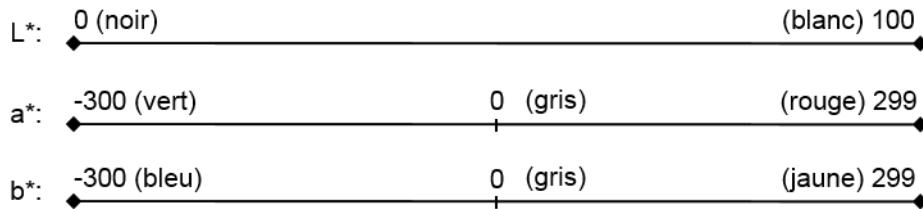


FIGURE 4.1 – Composantes de l'espace CIELAB

par la CIE, en 1960, c'est le CIE UVW. Cet espace est non-linéaire et vise à améliorer l'uniformité de répartition des couleurs<sup>1</sup>. Dans ce modèle, la composante V correspond en tout point à la composante Y de CIE XYZ (luminance). UVW est supplanté 16 ans plus tard (1976) par CIE U'V'W' dont la variation principale est le changement de la matrice de passage vers XYZ.

Enfin, en 1976 également, apparaissent les deux espaces les plus complets et utilisés à l'heure d'aujourd'hui : CIELAB et CIELUV. CIELUV est utilisé pour caractériser les couleurs de lumières tandis que CIELAB sert dans le domaine des couleurs de surface. Ils sont basés sur U'V'W' et appartiennent à la famille des espaces uniformes mais non-linéaires. Nous détaillerons ici CIELAB car c'est celui qui est utilisé dans le domaine de l'informatique. Pour de plus amples informations, on conseillera [Schanda & International Commission on Illumination, 2007] au lecteur.

L'espace CIELAB, de sa vraie dénomination CIE  $L^*a^*b^*$ , est caractérisé par une composante de clarté  $L^*$  et deux paramètres  $a^*$  et  $b^*$  qui expriment l'écart de la couleur par rapport à une surface grise de même clarté.  $a^*$  et  $b^*$  prennent leurs valeurs entre -300 et +299, 0 étant le gris de référence, mais sont en fait en général restreints entre -128 et +127 de manière à avoir 256 valeurs et être codé sur 8 bits. Le passage en couleurs codées sur 10 bits permet une nuance d'autant plus importante.

La composante  $a^*$  varie du vert (-300) vers le rouge (+299) tandis que la composante  $b^*$  varie du bleu (-300) au jaune (+299) (cf. Fig. 4.1) ; ce schéma est analogue au fonctionnement des canaux R-G et B-Y pour le traitement de l'image dans le cerveau.

Le gris achromatique de référence est calculé en fonction de la lumière d'éclairage, l'illuminant choisi est en général D65, par emprunt à l'esthétique du cinéma.

La conversion de CIEXYZ vers CIELAB se fait avec les équations suivantes [Robertson, 1990] :

$$f(x) = \begin{cases} x^{1/3}, & x > (\frac{6}{29})^3 \\ \frac{1}{3} (\frac{6}{29})^2 x + \frac{4}{29}, & \text{sinon} \end{cases} \quad (4.1)$$

Avec  $X, Y, Z$  les composantes CIEXYZ de la couleur,  $X_N, Y_N, Z_N$  les composantes CIEXYZ

1. Uniforme : les couleurs perçues varient de manière linéaire, au contraire de la valeur des primaires dans l'espace colorimétrique

du point blanc de référence, et :

$$\begin{cases} L^* = 116 \cdot f\left(\frac{Y}{Y_N}\right) - 16 \\ a^* = 500 \cdot \left[ f\left(\frac{X}{X_N}\right) - f\left(\frac{Y}{Y_N}\right) \right] \\ b^* = 200 \cdot \left[ f\left(\frac{Y}{Y_N}\right) - f\left(\frac{Z}{Z_N}\right) \right] \end{cases} \quad (4.2)$$

Les valeurs Des coefficients des équations de L, a et b sont en fait des valeurs approchées. Les valeurs exactes sont :

$$L^* : 117,16 \text{ et } 17,16$$

$$b^* : 509,39$$

$$b^* : 203,75$$

CIELAB a aussi été traduit en coordonnées cylindriques : LCh. Les composantes C et h sont les coordonnées polaires de  $a^*$  et  $b^*$ . Ces espaces permettent de qualifier n'importe quelle couleur indépendamment de la luminosité.

## 4.2 Lois et effets

### Lois de Grassmann

Hermann Grassmann (1809 - 1877) énonce trois lois fondamentales du comportement des couleurs en 1953. Ces lois linéaires ne sont toutefois pas universelles et présentent encore des défauts dans la zone des courtes longueurs d'onde tandis que la primaire bleue est jugée trop inhibitrice [Le Grand, 1972].

La première loi stipule que toute sensation colorée (couleur) peut être reproduite par le mélange de trois primaires fixées (Eq. 4.3).

$$\{C\} \equiv R \cdot \{R\} + V \cdot \{V\} + B \cdot \{B\} \equiv \begin{pmatrix} R \\ V \\ B \end{pmatrix} \quad (4.3)$$

La seconde loi stipule que toute couleur produite par la somme de deux couleurs est égale à la somme de ces deux couleurs (Eq. 4.4).

$$\{C\} \equiv \{C_1\} + \{C_2\} \equiv \begin{pmatrix} R_1 + R_2 \\ V_1 + V_2 \\ B_1 + B_2 \end{pmatrix} \quad (4.4)$$

Enfin, la troisième loi de Grassmann indique que si on augmente ou diminue une couleur dans une certaine proportion, cela revient à modifier chaque composante de la couleur par cette même proportion (Eq. 4.5).

$$\{C'\} \equiv k \cdot \{C\} \equiv \begin{pmatrix} k \cdot R \\ k \cdot V \\ k \cdot B \end{pmatrix} \quad (4.5)$$

## Loi d'Abney

Si quatre couleurs A, B et X, Y ont la même luminosité perçue deux à deux alors les mélanges additif de A et X et de B et Y auront la même luminosité perçue.

Cette loi n'est valable que dans certaines conditions très précises : en vision scotopique (vision de nuit) [Le Grand, 1972], lors de comparaison de couleurs par papillotement (alternance courte et rapide de couleurs) [Sèvre, 2009] ou en vision photopique (vision de jour) pour des angles de vision très restreints et des lumières très peu saturées.

## Effet Purkinje

Le spectre d'efficacité de vision des couleurs dépend de la luminosité, il se décale en fonction de la valeur de la luminosité : lorsqu'elle est forte, l'œil humain distingue mieux les couleurs rouges tandis que lorsqu'elle est faible, les couleurs bleues sont mieux perçues [Le Grand, 1972].

## Effet Helmholtz-Kohlrausch

La luminance perçue des couleurs varie, à luminance constante, en fonction de la teinte et de la saturation. Plus une couleur est saturée/teintée, plus elle est lumineuse [Le Grand, 1972].

Il en découle que le calcul de la luminosité ne devrait pas se faire sur la base seule de la composante Luminance.

## Effet Stiles-Crawford

On note une perte de luminosité en bordure de champ visuel, ce qui induit des couleurs altérées en périphérie [Damelincourt *et al.*, 2010].



FIGURE 4.2 – Répartition des domaines photopique, mésopique et scotopique

## Loi de Rico

Pour les petits angles de vision (inférieurs à 1,72 degrés - 30 millirad), l'acuité visuelle dépendrait du contraste [Damelincourt *et al.*, 2010] :

$$\begin{cases} C_s \cdot \alpha^2 = \text{cste}, & \alpha \leq 0.17^\circ \\ C_s \cdot \alpha = \text{cste}, & 0.17^\circ < \alpha \leq 1.72^\circ \end{cases}$$

## 4.3 Observateurs standards

Tout d'abord, il convient de (re)définir les trois domaines de vision qui correspondent à des quantités d'illumination différentes (cf. Fig. 4.3) 4.2) [Damelincourt *et al.*, 2010] :

- La vision *photopique* décrit la vision diurne (ou à haute luminosité). On entre dans le domaine *photopique* à partir d'une illumination de  $5 \text{ Cd}/\text{m}^2$ .
- La vision *scotopique* décrit la vision nocturne (ou à très basse luminosité). On entre dans le domaine *scotopique* en dessous d'une illumination de  $0.005 \text{ Cd}/\text{m}^2$ .
- La vision *mésopique* décrit la vision intermédiaire entre diurne et nocturne (luminosité basse). Le domaine *mésopique* se situe pour une illumination entre  $0.005 \text{ Cd}/\text{m}^2$  et  $5 \text{ Cd}/\text{m}^2$ .

La CIE a défini le concept d'*Observateur Standard* comme un profil d'efficacité lumineuse en fonction de la longueur d'onde (cf. Fig. 4.3). Chaque domaine de luminosité possède son *Observateur Standard*.

Soit : 
$$\begin{cases} \Phi_e(\lambda) & \text{le flux énergétique en W (watts)} \\ \Phi_v(\lambda) & \text{le flux visuel en lm (lumen)} \\ \lambda & \text{la longueur d'onde en nm (nanomètres)} \end{cases}$$

En 1931, un premier *Observateur Standard Photopique* pour le domaine photopique est mis au point. Il est noté  $V(\lambda)$  et est valable pour un cône de vision de  $2^\circ$ , c'est à dire globalement la zone d'acuité de la fovea. Cet observateur est révisé en 1964, en prenant en compte les résultats expérimentaux de plus de sujets pour calculer les coefficients. L'observateur est noté  $V_{10}(\lambda)$  et couvre cette fois un cône de vision de  $10^\circ$  [Damelincourt *et al.*, 2010; Le Grand, 1972] :

$$\begin{cases} \Phi_v(\lambda) = K_M \cdot V(\lambda) \cdot \Phi_e(\lambda) \\ K_M = 683 \text{ lm} \cdot \text{W}^{-1} \end{cases} \quad (4.6)$$

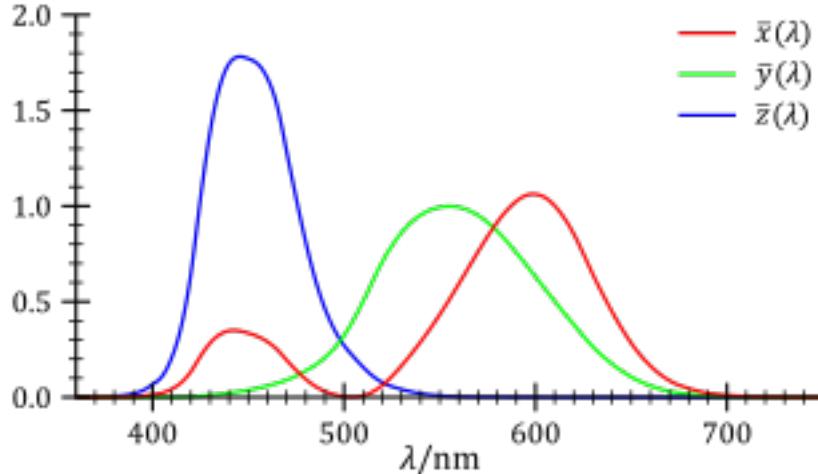


FIGURE 4.3 – Courbes d'efficacité lumineuse des Observateurs Standards de la CIE

L'*Observateur Standard Scotopique* est quant à lui défini en 1951. Il est noté  $V'(\lambda)$  :

$$\begin{cases} \Phi_v(\lambda) = K'_M \cdot V'(\lambda) \cdot \Phi_e(\lambda) \\ K'_M = 1700 \text{ lm.W}^{-1} \end{cases} \quad (4.7)$$

Enfin, il faut attendre 2010 pour voir apparaître l'*Observateur Standard Mésopique*,  $V_{mes}(\lambda)$ , qui n'est rien d'autre qu'une somme pondérée des deux autres observateurs en fonction de la luminance [Halonen & Puolakka, 2011] :

$$\begin{cases} \Phi_v(\lambda) = K_{mes} \cdot V_{mes}(\lambda) \cdot \Phi_e(\lambda) \\ V_{mes}(\lambda) = (1 - m) \cdot V'(\lambda) + m \cdot V(\lambda) K_{mes} = \frac{683}{V_{mes}[555]} \text{ lm.W}^{-1} \end{cases} \quad (4.8)$$

Les valeurs de  $m$  sont fixées en dehors des bornes du domaine mésopique ( $m = 1$  pour  $L_{mes} \geq 5 \text{ Cd/m}^2$  et  $m = 0$  pour  $L_{mes} \leq 0.005 \text{ Cd/m}^2$ ), sinon elles sont calculées par récurrence [Halonen & Puolakka, 2011] :

$$\begin{cases} m_0 = 0.5 & a = 0.7670 \\ m_n = a + b \cdot \log_{10}(L_{mes,n}) & b = 0.3334 \\ L_{mes,n} = \frac{m_{n-1} \cdot L_{photopic} + (1 - m_{n-1}) \cdot L_{scotopic} \cdot V'(\lambda_0)}{m_{n-1} + (1 - m_{n-1}) \cdot V'(\lambda_0)} & \lambda_0 = 555 \text{ nm} \end{cases} \quad (4.9)$$

## 4.4 Illuminants

Les illuminants sont des standards de lumière déposés par la CIE (Commission Internationale de l'Eclairage) sous la forme de spectre de lumière visible (cf. Figure 4.4). Chaque illuminant représente un certain type et une certaine couleur de lumière. Si l'illuminant D65 est le plus connu et le plus utilisé car il correspond à une lumière de midi en Europe occidentale, il existe plusieurs catégories.



FIGURE 4.4 – Spectre de l'illuminant D65 (CIE)

- La classe A, déposée en 1931, représente une lumière moyenne d'une lampe à incandescence (filament en tungstène).
- La classe B (1931) représente la lumière directe émanant du soleil.
- La classe C (1931) représente la lumière du jour (après passage dans l'atmosphère).

Une deuxième vague d'illuminants ont été adoptés plusieurs années après, à l'occasion de l'adoption du second observateur standard :

- La classe D (1964), qui se subdivise en plusieurs illuminants qui représentent les différentes phases de la lumière du jour, D65 étant midi.
- La classe E associée aux illuminants d'énergie égale (iso-énergie).
- La classe F représentant différentes lampes fluorescentes.

## 4.5 Equations de différentiations des couleurs

Une fois l'espace colorimétrique choisi, on peut dénombrer le nombre de couleurs théoriques discernables. Par exemple, dans le cas de l'espace RGB codé sur 8 bits on a 256 valeurs possibles pour chaque primaire. Lorsque l'on combine les trois primaires, on obtient un nombre théorique de 16.8 millions de couleurs. Si chacune de ces couleurs est à priori discernable, elles ne le sont pas forcément deux à deux. La CIE a donc mis en place une métrique qui indiquerait si la différence entre deux couleurs était, ou non, perceptible :  $\Delta E$ . Par construction, pour  $\Delta E < 1$  la différence de sensation colorée n'est pas perceptible et l'œil humain ne voit aucune différence. Empiriquement la valeur de seuil est bien plus élevée et dépend même du type de population (experte ou néophyte dans la reconnaissance de couleurs) [Sharma & Bala, 2013; Vidal *et al.*, 2016].

Il existe plusieurs équations pour calculer l'écart colorimétrique  $\Delta E$ , toutes ont été officialisées par la CIE en attendant des versions plus pertinentes [Beretta, 2000; Habekost, 2013; Robertson, 1990; Sharma *et al.*, 2005] :

- CIE76
- CIE94
- CIEDE2000
- CMC(l,c)

On exprimera ici les équations dans l'espace CIELAB dans sa version originale ( $L^*a^*b^*$ ) ou dans sa version en coordonnées cylindriques ( $L^*Ch_{ab}$ ).

La première équation, CIE76, est la plus simple et consiste en une simple distance euclidienne dans l'espace colorimétrique [Sharma & Bala, 2013] :

$$\Delta E_{76} = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2} \quad (4.10)$$

Ensuite, CIE76 a été modifiée pour s'adapter à la non linéarité de perception des couleurs [Beretta, 2000], elle est adoptée en 1994 et devient CIE94 :

$$\Delta E_{94}^* = \sqrt{\left(\frac{\Delta L^*}{k_L \cdot S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C \cdot S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H \cdot S_H}\right)^2} \quad (4.11)$$

Avec :

$$\begin{cases} C_{ab}^* = \sqrt{a^{*2} + b^{*2}} \\ \Delta H_{ab}^* = \sqrt{\Delta E_{ab}^{*2} - \Delta L^{*2} - \Delta C_{ab}^{*2}} = \sqrt{\Delta a^{*2} + \Delta b^{*2} - \Delta C_{ab}^{*2}} \\ S_C = 1 + 0.0045 \cdot C_{ab}^* \\ S_H = 1 + 0.0015 \cdot C_{ab}^* \\ S_L = k_L = k_C = k_H = 1 \end{cases}$$

L'équation de 1994 résout un grand nombre de problèmes mais éprouve toujours des difficultés à décrire parfaitement le comportement de l'oeil, notamment dans le bleu. Une troisième équation est donc officialisée en 2000, CIEDE2000 [Schanda & International Commission on Illumination, 2007; Sharma *et al.*, 2005] :

$$\Delta E_{00}^* = \sqrt{\left(\frac{\Delta L'}{k_L \cdot S_L}\right)^2 + \left(\frac{\Delta C'}{k_C \cdot S_C}\right)^2 + \left(\frac{\Delta H'}{k_H \cdot S_H}\right)^2 + R_T \cdot \frac{\Delta C'}{k_C \cdot S_C} \cdot \frac{\Delta H'}{k_H \cdot S_H}} \quad (4.12)$$

Le détail des coefficients de  $\Delta E_{00}^*$  (parce que extrêmement complexes et hors de propos ici) sont visibles sur l'excellent site de Bruce Lindbloom<sup>1</sup>.

1. Bruce Lindbloom, *Delta E (CIE 2000)*, [http://www.brucelindbloom.com/index.html?Eqn\\_DeltaE\\_CIE2000.html](http://www.brucelindbloom.com/index.html?Eqn_DeltaE_CIE2000.html) (Juillet 2016)

Enfin, si CIEDE2000 donne de très bon résultats lorsqu'il faut discriminer des couleurs de manière objective, c'est une autre équation, DECMC qui semble plus appropriée pour une discrimination subjective/perceptuelle [Habekost, 2013] :

$$\Delta E(l, c) = \sqrt{\left(\frac{\Delta L}{l \cdot S_L}\right)^2 + \left(\frac{\Delta C}{c \cdot S_C}\right)^2 + \left(\frac{\Delta H}{S_H}\right)^2} \quad (4.13)$$

Avec :  $\begin{cases} (l, c) = (2, 1) & \text{pour l'acceptabilité} \\ (l, c) = (1, 1) & \text{pour la perceptibilité} \end{cases}$

De même que pour CIEDE2000, le détail des coefficients de  $\Delta E(l, c)$  sont disponibles sur le site de Bruce Lindbloom<sup>1</sup>.

---

1. Bruce Lindbloom, *Delta E (CMC)*, [http://www.brucelindbloom.com/index.html?Eqn\\_DeltaE\\_CMC.html](http://www.brucelindbloom.com/index.html?Eqn_DeltaE_CMC.html) (Juillet 2016)



## CHAPITRE

# 5

# PERCEPTION VISUELLE DE LA PROFONDEUR

La perception de la profondeur est une des composantes fondamentales de la vision et doit faire partie des critères d'une immersion réussie. Cette perception passe par un certain nombre de principes et de paradigmes à la fois monoculaires (fonctionnement grâce à un seul œil) ou binoculaire (fonctionnement nécessairement avec les deux yeux). Les indices de profondeurs utilisés par le cerveau sont les suivants, et sont résumés et ordonnés dans la Fig. 5.1 :

- Accommodation
- Vergence
- Disparités
- Parallaxe
- Indices statiques : interposition, taille, perspective linéaire, gradient de texture, perspective aérienne, ombre, ...

Premièrement, le cerveau peut récupérer assez facilement les informations d'*accommodation* et de *vergence* via le principe de proprioception évoqué plus haut. En récupérant des informations sur les muscles oculomoteurs le cerveau en déduit l'état de vergence et de l'accommodation puis des informations sur la proximité de l'objet sur lequel est concentré le regard. En effet, une accommodation et une vergence à l'infini signifieront un objet éloigné et inversement. Ces indices ne sont pas forcément les plus forts et les plus robustes mais ils participent à un tout.

Les *disparités* ont déjà été présentées précédemment, aussi on renverra le lecteur à la section correspondante pour une description plus détaillée. La différence de disparité entre

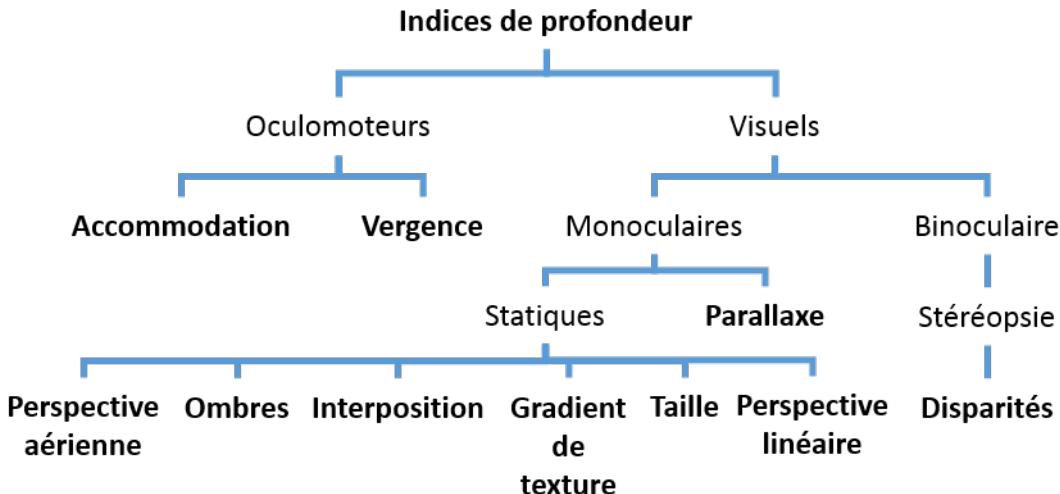


FIGURE 5.1 – Indices visuels de perception de la profondeur

deux points permet au cerveau de comparer leurs positions respectives. Au niveau technologique cet indice est apporté par la stéréoscopie. Pour plus d'informations sur l'indice binoculaire de profondeur, on pourra se reporter à [Glassner, 1995] (en anglais).

La *parallaxe*, ou parallaxe de mouvement, est un indice temporel, c'est à dire qu'il ne fonctionne pas à un instant  $t$  donné mais plutôt sur une période de temps  $\Delta t$ . En effet, la parallaxe est l'effet sur le cerveau d'un changement de point de vue : le cerveau analyse les différences entre les images  $n$  et  $n + 1$  et en fonction des quantités et des vitesses de déplacement en déduit les profondeurs des objets. Lorsque qu'un observateur bouge la tête en gardant le même point de fixation, tous les objets entre le point de fixation et l'observateur vont bouger avec un gradient de vitesse basé sur la distance au point de fixation (plus l'objet est éloigné, plus il va vite). Tous les points derrière le point de fixation bougeront aussi mais avec le gradient de vitesse inverse. La parallaxe est assurée par la capture et l'intégration du mouvement de la tête et/ou des yeux de l'observateur dans le calcul de l'image.

Parmi les indices statiques, l'*interposition* (ou occultation) est assez simple et révélateur : si un objet masque (partiellement) un autre, alors l'objet masqué à une profondeur supérieure à l'objet masquant. Cet indice est notamment très utilisé en informatique graphique dans le pipeline de rendu<sup>1</sup> au moment de calculer une image : tous les objets masqués sont ignorés pour le rendu de l'image (ce qui permet un gain de temps énorme, notamment sur des scènes complexes) et les objets partiellement masqués voient leur géométrie modifiée afin de ne garder que la (les) partie(s) visible(s).

La *taille* est un autre indice, basé cette fois sur notre expérience d'objets familiers dont on connaît ou dont on visualise la taille. Si une voiture est vue minuscule, par expérience, on saura que c'est parce que la voiture est loin. Cet indice ne marche qu'avec des objets que l'on connaît déjà et peut être berné assez facilement (comme par exemple avec l'illusion d'optique de la chambre d'Ames).

1. Pour plus d'informations sur le pipeline de rendu : *Rendering Pipeline Overview*, [https://www.opengl.org/wiki/Rendering\\_Pipeline\\_Overview](https://www.opengl.org/wiki/Rendering_Pipeline_Overview) (Juillet 2016, en anglais).

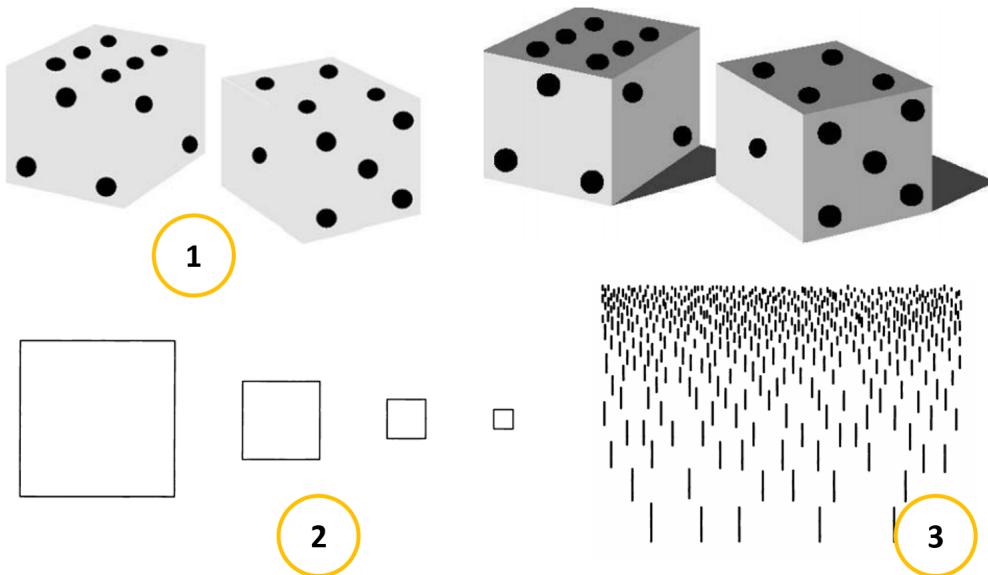


FIGURE 5.2 – Illustration d'indices de profondeur : ombres (1, image tirée de [ANSES, 2014]), taille (2) et gradient de profondeur (3, images tirées de [Glassner, 1995]).

*L'ombre* permet localement de donner des indices flagrants sur la profondeur de l'objet (cf. Fig. 5.1).

Enfin il existe un certain nombre d'indices statiques monoculaires [Fuchs *et al.*, 2003; Glassner, 1995] que l'on peut regrouper sous la dénomination « Perspective ». Certains de ces différents types de perspective sont illustrés Fig. 5.2.

Le *gradient de texture* consiste à un rapprochement apparent (un écrasement des distances relatives) entre les objets de plus en plus lointains.

La *perspective aérienne* fonctionne pour les grandes distances uniquement : plus un objet va s'éloigner de l'observateur, plus sa couleur va perdre en intensité et fondre vers le bleu-gris (couleur de l'atmosphère), typiquement le cas des montagnes ou d'un paysage que l'on voit à l'horizon.

La *perspective linéaire* est la perspective bien connue avec lignes de fuite : plus un objet s'éloigne, plus il semble rapetisser.



## CHAPITRE

# 6

## CONTRASTE

### 6.1 Définitions mathématiques

Tout d'abord, il convient de définir le contraste comme une « opposition de deux choses, l'une faisant ressortir l'autre »<sup>2</sup>. Deux définitions plus spécifiques sont « effet subjectif d'une apposition quantitative de couleurs, par exemple des stimulations sensorielles juxtaposées dans l'espace (contraste simultané) ou dans le temps (contraste successif) »<sup>3</sup> et « variation relative de l'éclairement d'une image lorsqu'on se déplace à l'intérieur de cette image »<sup>4</sup>. L'opposition se ferait ici entre le blanc et le noir pour avoir la valeur maximale du contraste.

Ces définitions sont toutefois insuffisantes ou tout du moins inutilisables dans le monde de la vision et de l'imagerie informatique. Il faut alors théoriser plus précisément le concept de contraste et le représenter mathématiquement pour à la fois pouvoir le mesurer mais aussi avoir plus de contrôle sur sa caractérisation et son obtention. Il existe pour ce faire au moins 3 méthodes analytiques, résumées ci-après :

- Le contraste de Michelson [Fuchs *et al.*, 2003; Michelson, 1995; Winkler, 1999].
- Le contraste de Weber [Winkler, 1999; ?].
- Le contraste de Peli [Peli, 1990; Winkler, 1999; ?].

2. Contraste. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/contraste/18688>

3. Contraste des couleurs. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/contraste/18688/locution>

4. Contraste d'une image optique. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/contraste/18688/locution>

Premièrement, le contraste de Michelson est défini par l'Eq. 6.1, avec  $L_{max}$  la valeur maximale de la luminance de l'image, et  $L_{min}$  la valeur minimum. Le contraste de Michelson prend sa valeur dans l'intervalle  $[0; 1]$ .

$$C_{Michelson} = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (6.1)$$

Le contraste de Weber est quant à lui défini tel que dans l'Eq. 6.2.  $\Delta L$  étant la plus petite différence de luminance visible, pour une luminance  $L$  donnée. A la différence du contraste de Michelson, le contraste de Weber prend ses valeurs dans l'intervalle  $[-1; +\infty[$ .

$$C_{Weber} = \frac{\Delta L}{L} \quad (6.2)$$

Enfin, Peli a proposé un modèle de contraste pour les images complexes basé sur une analyse fréquentielle de l'image (cf. Eq. 6.3). Avec  $BP_i(x, y)$  la bande passante de l'image  $i$  et  $LP_i(x, y)$  l'énergie sous la fréquence de coupure.

$$C_i(x, y) = \frac{BP_i(x, y)}{LP_i(x, y)} \quad (6.3)$$

Malgré tout, il n'existe pas (encore) de méthode parfaite et définitive pour mesurer le contraste dans les images complexes de type image de simulateur [?]. Les définitions analytiques vues à l'instant sont valables pour des images très simples et statiques.

## 6.2 Lois de comportement

Le calcul du contraste en lui même ne suffit pas, il faut aussi savoir comment celui évolue en fonction de la luminance de l'écran : il semble assez naturel qu'à très forte luminance la capacité à distinguer un certain niveau de contraste ne soit pas la même qu'à très faible luminance. On recense ici trois modèles de comportement pour le contraste issus de la littérature :

- La fraction de Weber<sup>1</sup>.
- Le modèle de Blackwell [Blackwell, 1981].
- Le modèle de Ward [Ward, 1994].

Ces modèles, notamment la fraction de Weber, ne sont valables que dans le domaine photopique (lumière du jour) et sont à utiliser avec prudence. A notre connaissance, il n'existe pas de modèle de comportement de la vision du contraste dans les domaines spécifiques du mésopique et du scotopique.

---

1. Serge Bertorello, *Notions d'Optique - La Vision*, <http://serge.bertorello.free.fr/optique/vision/vision.html> (Juillet 2016)

Surement le modèle le plus connu et le plus utilisé, la fraction de Weber est dans la continuité du contraste de Weber définit précédemment. Si il s'applique notamment au contraste, ce modèle est en fait utilisable pour n'importe quel type de stimulus et est assez générale. La fraction est simplement posée telle que dans Eq. 6.4, avec  $k$  une constante,  $\Delta I$  la plus petite différence d'intensité perceptible et  $I$  l'intensité du stimulus. La valeur de  $k$  est généralement admise à 0.02.

$$\frac{\Delta I}{I} = k = 0.02 \quad (6.4)$$

De cette fraction de Weber est issue une autre loi à propos du comportement de la rétine face à une intensité lumineuse, la loi de Weber-Fechner (Eq. 6.5). Avec  $S$  la sensation perçue,  $I$  l'intensité de la stimulation et  $k$  une constante.

$$S = k \cdot \log(I) \quad (6.5)$$

En 1981, la CIE a adopté un modèle établi par Blackwell et destiné au comportement du contraste en fonction de la luminance uniquement. Ce modèle donne la plus petite différence de luminance perceptible  $\Delta L$  pour une luminance moyenne donnée  $L_a$  (Eq. 6.6). Ici,  $L_a \pm \Delta L$  est perceptible tant que  $L_a \pm \epsilon$  alors  $\epsilon < \Delta L$  ne l'est pas.

$$\Delta L(L_a) = 0.0594(1.219 + L_a^{0.4})^{2.5} \quad (6.6)$$

Enfin, le modèle de Ward donne un lien direct (linéaire) entre la différence minimale perceptible de luminance qui serait perçue dans le monde réel ( $L_{wa}$ ) et sont équivalent sur un dispositif d'affichage  $L_{da}$  (Eq. 6.7). Le coefficient de linéarité  $m$  est donné par l'Eq. 6.8.

$$\Delta L(L_{da}) = m \cdot \Delta L(L_{wa}) \quad (6.7)$$

$$m = \left[ \frac{1.219 + (L_{da})^{0.4}}{1.219 + (L_{wa})^{0.4}} \right]^{2.5} \quad (6.8)$$

Une fois ces modèles analytiques mis en place, on se rend compte assez facilement de leur lourdeur et de incapacité à s'adapter à des situations concrètes, notamment dans le domaine le l'informatique temps réel. Nous verrons donc dans un prochain chapitre d'autres pistes et d'autres moyens plus pratiques de quantifier le contraste.



## CHAPITRE

# 7

# INFLUENCES PHYSIOLOGIQUES

Outre les maladies ou autres défauts oculaires qui ne seront pas abordés ici, il existe un certain nombre de paramètres liés au corps humain qui influent sur la qualité de perception des images projetées par le dispositif immersif d'affichage. On retrouve, au niveau de précision qui nous intéresse, principalement des influences morphologiques (au niveau du visage) ou en terme d'âge de l'observateur. Mais on peut par exemple noter que la sensibilité au contraste est dépendante des conditions d'expérimentations évidemment, mais aussi de facteurs ophtalmologiques (myopie, éblouissement rétinien, cataracte, amblyopie, dégénération de la macula avec l'âge, hyper-tension oculaire, glaucome, sécheresse oculaire) mais aussi de facteurs neurologiques (lésions cérébrales, scléroses, maladie de Parkinson et schizophrénie), ou encore de facteurs liés à la prise de médicaments [Pelli & Bex, 2013]. Nous nous plaçons évidemment dans le cas où le sujet est sain.

## 7.1 Variations morphologiques

Les influences morphologiques sont très nombreuses au niveau du visage et plus ou moins faciles à caractériser et à traiter. On retrouve par exemple la distance inter-oculaire (distance entre le centre des pupilles), profondeur des yeux, non-horizontalité de la droite passant par le centre des pupilles, glissement des lunettes pendant l'utilisation, position des oreilles par rapport au nez pour la poste des lunettes 3D, écart de position entre une première et une seconde mise de lunettes, la position du casque immersif sur le visage ... tout ce qui fait de l'observateur humain un observateur symétriquement et géométriquement imparfait.

On développera ici le cas simple de la distance inter-oculaire. Il apparaît que lorsque l'on présente les mêmes images stéréo à deux personnes différentes, la profondeur stéréoscopique des objets proches n'est pas perçue de manière égale. On prend l'exemple l'image

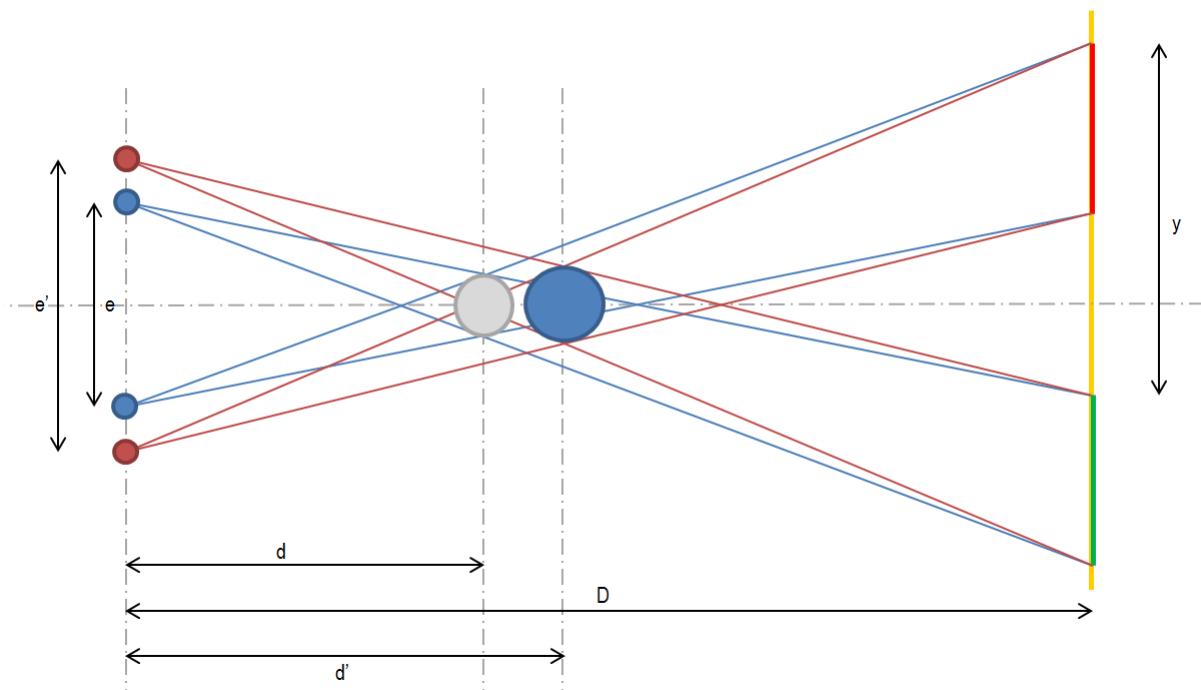


FIGURE 7.1 – Schématisation de la différence de profondeur stéréoscopique perçue en fonction de la distance interoculaire

d'un volant de virtuel normalement paramétré pour apparaître confondu avec le volant réel du simulateur de conduite. D'aucuns verront effectivement le volant là où il est sensé être, tandis que d'autres verront l'image légèrement en avant du volant et d'autres la verront légèrement en retrait. Cet écart peut être extrêmement gênant et s'explique notamment par l'écart de distance inter-pupillaire entre les différents observateurs. Il serait cependant intéressant de comparer l'influence des différentes influences physiologiques vues précédemment afin de les hiérarchiser et éventuellement d'en négliger certaines.

Les longueurs ici évoquées font référence à la Fig. 7.1. Pour un individu A qui a une distance interoculaire  $e$ , on projette sur les murs du CAVE, où dans un visio-casque, à distance  $D$ , un couple d'images stéréoscopiques (ici représentées par les segments rouge et bleu). L'image reconstruite par le cerveau est vue à une distance  $d$  de l'individu. Les images œil gauche et œil droit sont dissociées d'une distance  $y$ . En gardant les mêmes images, un individu B avec une distance interoculaire  $e'$  se positionne au même endroit dans le CAVE. C'est à ce niveau là qu'intervient la différence de perception, et le second individu verra l'objet virtuel non pas à une distance  $d$  mais à une distance  $d'$  qui peut être supérieure ou inférieure à  $d$ .

En utilisant les règles élémentaires de trigonométrie (et après simplification), on trouve que la différence de distance de l'objet virtuel entre les deux individus se note de la manière suivante (Eq. 7.1) :

$$\Delta d = \frac{(D - d)(e + \Delta e) - y \cdot d}{y + e + \Delta e} \quad (7.1)$$

La Fig. 7.2 est un abaque partiel pour une configuration donnée et pour le spectre des

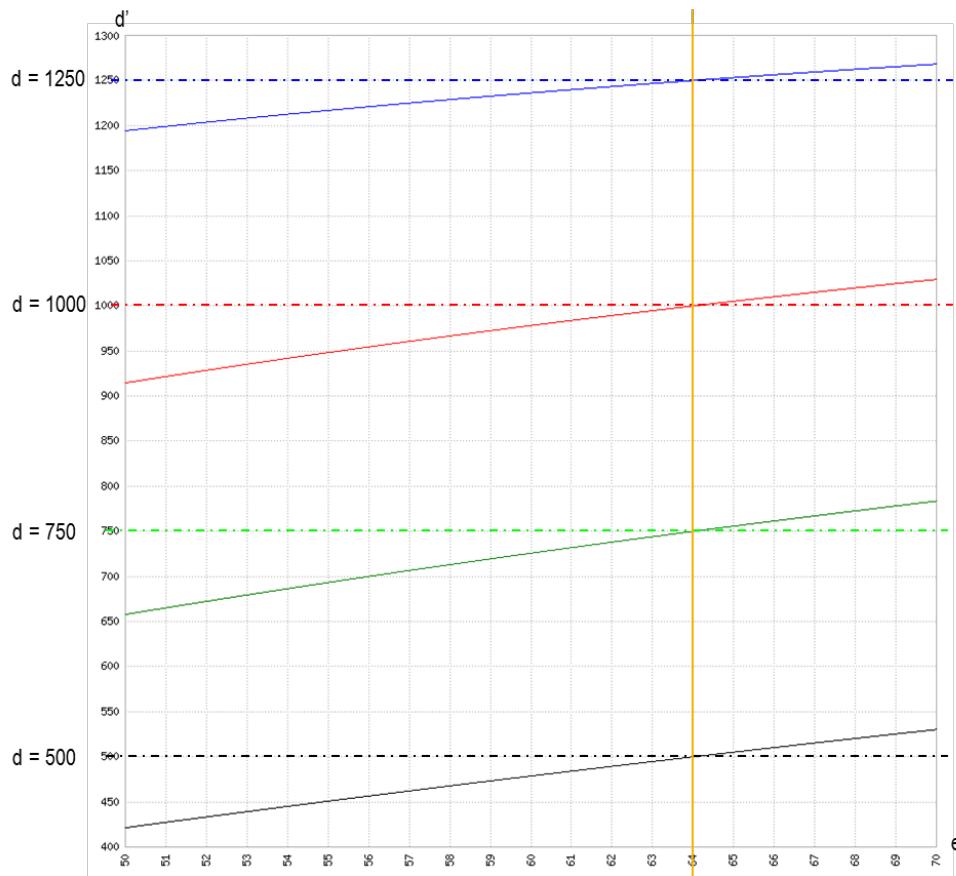


FIGURE 7.2 – Abaques de la nouvelles distance perçue en fonction de la distance inter-oculaire

distances inter-oculaires, de l'écart de profondeur perçue entre l'observateur A et l'observateur B pour un même jeu d'images stéréoscopiques.

L'abaque est tracé dans la configuration spécifique avec des images projetées à la distance  $D = 1.5m$  et pour des distances de perception de l'objet en 3D de  $d = 50cm, 75cm, 1met1m25$ . De même, on se place dans le cas général, et dans les pratiques usuelles en 3D échelle 1 : 1, la distance interoculaire initiale (individu A) utilisée pour calculer le couple d'images stéréoscopiques est de 64 mm, qui est la distance interoculaire moyenne chez les hommes [Dodgson, 2004].

Dans le cas idéal, celui où la distance inter-oculaire n'influerait pas sur la perception de la profondeur, les images calculées pour 64mm (trait orange vertical) seraient vues toujours à la même distance quelque soit la distance interoculaire (traits horizontaux en pointillés). Mais la géométrie montre bien que la réalité est tout autre (traits pleins).

Les résultats qui émergent des courbes sont les suivants. Si l'individu A perçoit l'objet virtuel à une distance de 1m (courbes rouges, Cf. 7.2), un individu B avec une distance oculaire de 60mm (distance inter-oculaire moyenne chez les femmes [Dodgson, 2004]) verra le même objet virtuel à 980mm (donc légèrement plus proche), soit avec un écart de 2cm par rapport à l'observateur calculé.

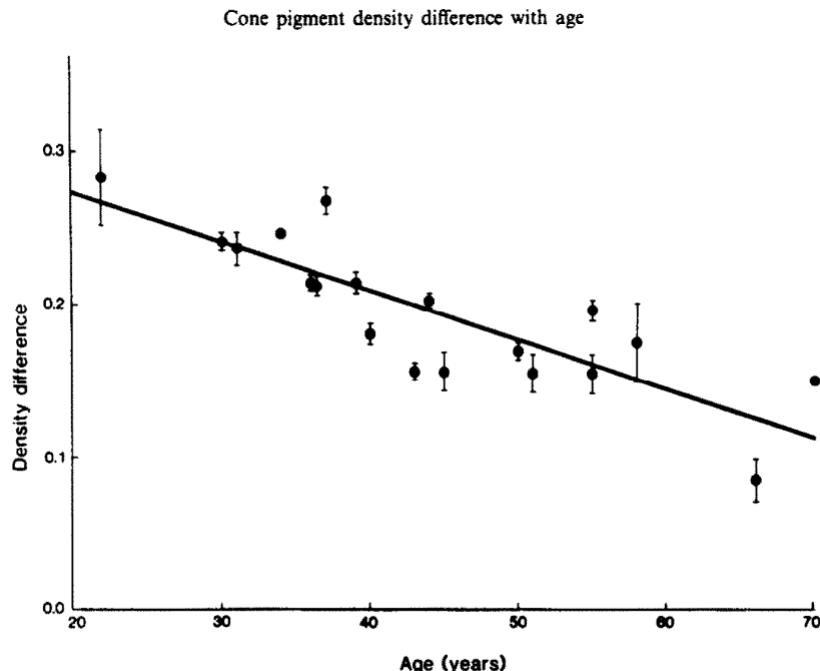


Fig. 1. Cone pigment density difference as a function of subject's age. The density difference is for the spectral peak at 560 nm and is from the central 2 degrees of visual field. Points are the mean  $\pm$  SEM for multiple bleaches per subject.

FIGURE 7.3 – Densité des cônes en fonction de l'âge (figure tirée de [Kilbride *et al.*, 1986])

## 7.2 Âge

L'ergonomie du système d'affichage et la physionomie du visage ne sont pas les seuls facteurs influençant la perception. Un certain nombre d'autres paramètres entrent en jeu, comme notamment le genre (homme ou femme) pour la vision des couleurs [Fairchild, 2005] mais surtout, pour la quasi totalité des grandeurs de la vision : l'âge de l'observateur.

On peut dès lors évoquer la modification, avec l'accumulation des années, de facteurs comme le diamètre pupillaire, l'acuité visuelle ou la sensibilité au contraste [Owsley *et al.*, 1983] (Cf. Fig. 3.10), la réduction de la capacité à extraire des informations dans le champ de vision [Ball *et al.*, 1988; Gross *et al.*, 2008; Rogé & Pébayle, 2009; Rogé *et al.*, 2004; Sekuler & Bennett, 2000], la qualité de l'image rétinienne [Artal *et al.*, 1993], la densité de cônes dans la fovéa [Kilbride *et al.*, 1986], la structure de l'œil et du cristallin [Cook *et al.*, 1994].

Par ailleurs, on note les excellentes vidéos de Craig Blackwell sur le thème du vieillissement de l'œil et de la vision<sup>1</sup>.

La finalité (ici) de toutes ces études sur le vieillissement de l'œil est de pouvoir intégrer les résultats dans un modèle de vision. Nous verrons dans la prochaine section ce qu'est et de quoi peut être composé un modèle de vision.

1. Craig Blackwell. 3 décembre 2014. Aging Eye 1 : Vision [Vidéo en ligne]. Repéré à <https://www.youtube.com/watch?v=FtcZ4an-VGo>

## **Deuxième partie**

# **Première approche : modèles de vision**



# INTRODUCTION

Le terme « modèle de vision » est très générique et englobe en fait un certain nombre de briques que l'on va vouloir mettre, ou non, suivant les applications et la finesse du modèle recherché. Définitions possibles :

- [Moreau & Tisseau, 2006] : « Algorithme qui prend en entrée une image et qui renvoie l'interprétation qu'en fait le système visuel humain. »
- [Pattanaik *et al.*, 1998] : « Les modèles de vision ont pour objectif de relier la perception visuelle d'un observateur dans une scène virtuelle à la perception visuelle d'un sujet regardant un dispositif d'affichage. Le but final étant de spécifier une transformation dont l'objectif serait de corréler visuellement la scène virtuelle à l'image affichée. »

Un modèle de vision est donc une construction théorique qui vise à reproduire et prédire le fonctionnement de la vision humaine. Le modèle va ainsi intégrer tout un système de modélisation du passage de la lumière dans l'œil, de la captation par la rétine puis du traitement de la lumière par le cerveau. Le modèle de vision est à différencier du prédateur de vision en cela que ce dernier est inclus, avec d'autres sous-modèles, dans le modèle de vision. Un modèle de vision peut notamment aussi intégrer une prédiction des effets de l'âge (comme évoqué précédemment, voir par la suite) mais aussi des comportements plus spécifiques comme le mécanisme d'attention visuelle, ou la reproduction de tons.

Nous allons maintenant présenter dans les sections suivantes un état de l'Art partiel des modèles de vision, prédateurs de vision et modèles de reproduction de tons, puis détailler deux modèles particulièrement intéressants :

- Modèle de vision avec vieillissement du système optique [Mantiuk & Ramponi, 2015].
- Prédateur de vision chromatique complet [Pattanaik *et al.*, 1998].

---

## INTRODUCTION

---

## CHAPITRE

# 1

## ÉTAT DE L'ART

L'état de l'Art ici présenté n'a pas volonté d'être exhaustif sur tous les modèles de vision tant il en existe de types et de variations. On présente néanmoins les grands classiques ainsi que certaines de leurs ramifications.

Si on a défini dans la section précédent le modèle de vision, il reste encore à déterminer les actions et objectifs de deux types de sous-modèles, le Visual Differences Predictor (VDP) et la reproduction de tons :

- La reproduction de tons, ou *tonemapping*, est une technique qui vise à contrebalancer les limites techniques des capteurs d'image par rapport à l'œil humain. En effet, celui-ci est -comme on a pu voir précédemment- capable d'intégrer en temps réel un très grand nombre de couleurs et surtout de grands écarts de luminosité, tandis qu'un capteur technologique est bien plus grandement limité. Le tonemapping est alors une technique pour ramener les informations hors-cadre du capteur dans son champ d'action.
- Le VDP est un type d'algorithme qui est amené à donner, objectivement, les différences entre deux traitements d'une même image, par exemple entre l'image initiale et une image post-compression. Si le VDP est modélisé sur le système visuel humain, on pourra savoir si les différences entre traitements d'image sont visibles à l'œil nu. Un VDP HDR aura pour objectif de s'assurer que l'interpolation des luminances non-affichables en luminances affichables est efficace.

Tout d'abord, il convient de citer les deux modèles fondateurs et sur lesquels la quasi majorité des autres modèles de vision sont adaptés : le *Visual Differences Predictor* de [Daly, 1992] et le *Visual Discrimination Model* de [Lubin, 1995]. Si ces deux modèles sont des références absolues, ils restent néanmoins des modèles achromatiques (le traitement et le résultat se font uniquement sur le canal de luminance de l'image). Leur équivalent chromatique (traitement sur tous les canaux de l'image, luminance et couleur) est le modèle

développé par [Pattanaik *et al.*, 1998]. Ce dernier sera développé en détail dans la section suivante. Enfin, [?] ont formalisé les premiers la méthode de reproduction de tons, méthode reprise très largement dans de nombreux modèles par la suite. On en profite également pour pointer vers deux références qui contiennent elle même de très nombreuses références sur le sujet : [Moreau & Tisseau, 2006] et [Bradley, 1999].

## 1.1 Modèles principaux

Le modèle de [Daly, 1992] est donc un modèle achromatique qui se base sur une simulation de l'adaptation des cellules de la rétine à la luminance. La fonction prend aussi en compte la vision spatiale (c'est à dire la sensibilité aux fréquences spatiales) avec l'intégration de fonctions de sensibilité au contraste (CSF, évoquées précédemment). Enfin, le modèle simule aussi les traitement corticaux (c'est à dire la gestion de l'information nerveuse par tout un réseau de neurones) via une fonction dite *cortex*.

Parallèlement, le modèle de [Lubin, 1995] a lui une structure plus logique que le modèle précédent, du point de vue du déroulement de la chaîne de vision (de la perception au traitement), néanmoins le traitement se fait toujours de manière achromatique. Il simule l'optique de l'œil et le comportement de la rétine, puis effectue un calcul des contrastes locaux (c'est à dire par zone de l'image perçue) puis il prend en compte l'orientation de l'image et enfin termine par une phase de transduction<sup>1</sup> (voir section suivante pour une définition plus complète).

## 1.2 Pérennités des modèles principaux

[Lukin, 2009] et [Bradley, 1999] ont eu, par rapport au modèle de Daly, des approches radicalement différentes. Le premier a cherché à améliorer la modélisation en complexifiant la fonction de cortex tandis que le second l'a remplacée par une autre approche : une fonction basée sur un modèle en ondelettes. Les améliorations de Lukin se concentrent sur une fixation plus précise des seuils utilisés dans la modélisation, l'ajout de la modélisation de l'effet de facilitation pendant l'effet de masque (traitements cognitifs) ainsi que la séparation de l'amplitude et de la phase de la réponse au modèle. L'amplitude servant à calculer le masque de contraste tandis que la phase prend part dans la modélisation de l'effet de facilitation).

## 1.3 Modèles de Mantiuk *et al.* [2004, 2005, 2015]

Une famille de modèles de vision très intéressante, et basée sur le modèle Daly, est conçue par Rafal Mantiuk et ses collaborateurs : [Mantiuk *et al.*, 2004, 2005; Mantiuk & Ramponi,

---

1. Transducteur. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/transducteur/>

2015]. Le dernier de ces modèles, qui vient en amélioration des précédents, est d'ailleurs développé en détail par la suite.

[Mantiuk *et al.*, 2004] proposent un modèle de vision pour les images HDR (High Dynamic Range). Leur hypothèse de base étant que le modèle de Daly est fonctionnel mais reste éloigné de la vision humaine en cela qu'il n'accepte pas une amplitude de luminance suffisante et n'intègre donc pas de mécanisme d'adaptation du système visuel à la luminosité ambiante. Deux modifications principales du modèle initial sont alors opérées :

- Le modèle de Daly part du principe que les photo-récepteurs donnent des réponses égales et subissent des pertes égales quel que soit le niveau de luminance (haut ou bas) : une nouvelle transformation de l'information de luminance, cette fois non-linéaire, est proposée.
- Une simple fonction de contraste étant insuffisante pour traiter des images avec de grands écarts de luminance, une nouvelle approche est proposée : l'image est filtrée par niveau de luminance puis chaque niveau se voit appliquer une fonction de sensibilité au contraste (CSF).

Ensuite, [Mantiuk *et al.*, 2005] apportent une deuxième modification au modèle. La modélisation du passage de la lumière dans l'optique de l'œil est modifiée d'un fonction de diffusion classique (Point Spread Function) au profit d'une fonction de [Deeley *et al.*, 1991] (Eq. 1.1) :

$$OTF(\rho, d) = \exp \left[ - \left( \frac{\rho}{20.9 - 2.1 \cdot d} \right)^{1.3 - 0.07 \cdot d} \right] \quad (1.1)$$

Avec  $\rho$  la fréquence spatiale du stimulus qui traverse l'œil et  $d$  le diamètre de la pupille. Ce dernier est quand à lui donné par le modèle de [Moon & Spencer, 1944] (Eq. 1.2) :

$$d = 4.9 - 3 \cdot \tanh [0.4 (\log(Y_{adapt}) + 1.0)] \quad (1.2)$$

Enfin, [Mantiuk & Ramponi, 2015] apportent une modification fondamentale dans leur modèle, en rajoutant les effets de l'âge de l'observateur simulé sur la perception des erreurs. Ce modèle est largement discuté dans une section suivante (Section 3).

## 1.4 Reproduction de tons

Après la formalisation de la technique par [Rushmeier *et al.*, 1995], les algorithmes de reproduction de tons se divisent en trois grandes familles : le traitement des images statiques par des méthodes globales, le traitement des images statiques cette fois ci par des méthodes locales et enfin le traitement des images dynamiques ; [Moreau & Tisseau, 2006] relèvent un grand nombre de modèles du genre. Cette dernière famille est entièrement revue par [Drago *et al.*, 2003].

Enfin, [Ferwerda *et al.*, 1996] proposent un modèle très complet, basé sur le modèle de reproduction de tons à base de contrastes de Ward. Ce modèle fonctionne avec des techniques de calcul d'éclairage plus avancées telle que la *Global Illumination*, et ce, à des niveaux de luminosité allant de photopique à scotopique. Sont modélisés : la capture des changements de seuil de visibilité, les changements d'apparence des couleurs, les changements d'acuité visuelle ainsi que la sensibilité à tous ces facteurs au cours du temps. On peut par exemple obtenir toutes les images intermédiaires d'un processus d'adaptation de l'œil à un changement brutal de luminosité, que ce soit du clair vers l'obscur ou inversement (Cf. Fig. 1.1).

Nous allons maintenant revenir en détails sur deux modèles particulièrement complets et rigoureux, chacun dans leur domaine : d'un côté le grand mimétisme de la vision humaine et de l'autre la prise en compte avancée des spécificités de l'observateur, et notamment son âge.

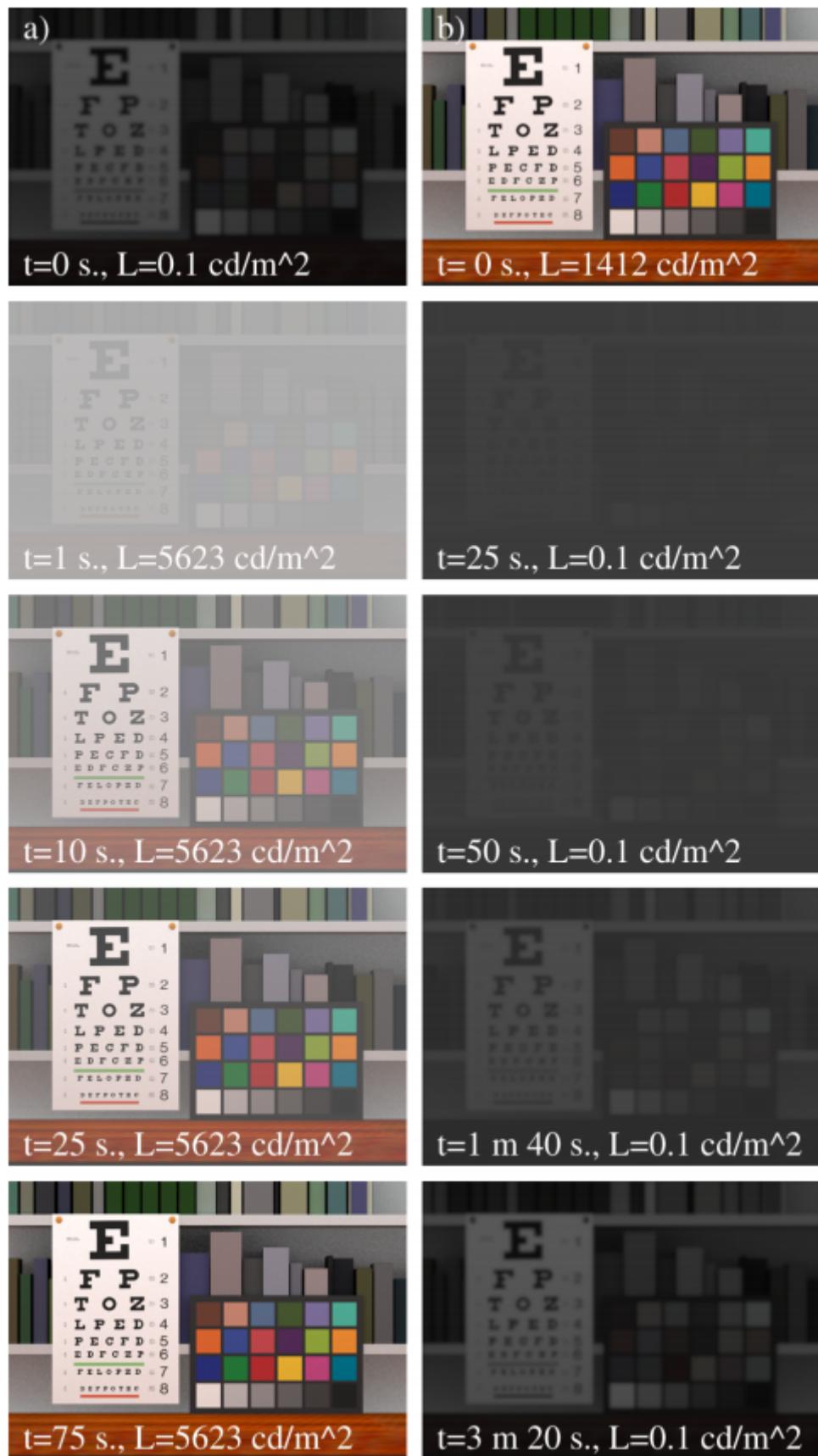


FIGURE 1.1 – Différentes étapes du modèle d'adaptation lumineuse. Images tirées de [Ferwerda *et al.*, 1996]

---

## CHAPITRE 1. ÉTAT DE L'ART

---

## CHAPITRE

# 2

## MODÈLE DE PATTANAIK *ET AL.*

Le modèle de vision proposé par [Pattanaik *et al.*, 1998] se positionne face à deux problématiques majeures dans le domaine de l'imagerie informatique : afficher des images ayant un intervalle de luminosité très grand et surtout arriver à afficher des images qui correspondent à la perception que la vision humaine aurait.

Il part du principe que beaucoup de modèles ne sont capables que de gérer des scènes uniquement achromatiques en entrée et non pas des scènes avec des caractéristiques plus proches d'un scène naturelle : lumière du jour, amplitude de luminosité étendue, acuité non-uniforme de l'observateur, ...

Les auteurs du modèle mettent donc en avant trois points clefs qu'il convient d'utiliser ou d'atteindre dans un modèle de vision :

- Fonctions de sensibilité au contraste (CSFs) : Cf. section sur le contraste.
- Mécanisme d'adaptation locale de la vision : permet d'améliorer grandement la capacité de voir distinctement dans un intervalle de luminosité très étendu.
- Modèle plus vaste qu'un modèle de seuil : en général, les modèles permettent de définir le seuil en-dessous au au-dessus duquel le visible devient invisible ou inversement. Un modèle plus vaste doit décrire l'intégralité du comportement et non pas uniquement au voisinage du seuil.

Le modèle proposé est divisé en deux parties principales : le modèle visuel à proprement parler et un modèle d'affichage (Cf. Fig. 2.1). L'algorithme visuel travaille une image d'entrée afin d'en extraire les contrastes des canaux chromatiques et achromatique. Enfin, l'algorithme d'affichage reconstruit une image afin de l'afficher.

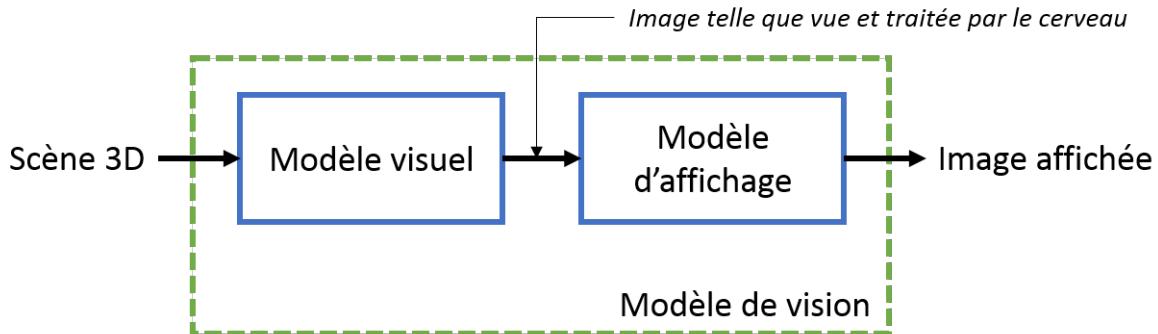


FIGURE 2.1 – Schéma synthétique du déroulé du modèle de [Pattanaik et al., 1998]

Le modèle emprunte ses composantes à de nombreux auteurs différents. Il se déroule comme suit (Cf. Fig. 2.2) :

- Décomposition de l'image via des filtres Gaussiens
- Compensation de la dispersion optique dans l'œil, modèle de [Westheimer, 1986]
- Compensation de l'éblouissement, modèle de [Spencer et al., 1995]
- Échantillonnage spectral de tous les étages de la pyramide sauf le plus bas (dernier) [Fairchild, 1998; Wyszecki & Stiles, 2000]
- Calibration des signaux d'entrée
- Traitement spatial des images de chaque canal [Burt & Adelson, 1983; Peli, 1990]
- Conversion en canaux chromatiques et achromatique (théorie des processus antagoniques) [Fairchild, 1998; Hunt, 1995]
- Passage par des transducteurs de contraste [Watson & Solomon, 1997]
- Combinaison des signaux des cônes et des bâtonnets
- Traitement de l'étage de la pyramide restant (dernier) [Fairchild, 1998]
- Résultats du modèle visuel
- Reconstruction de l'image par traitement inversé pour l'affichage

L'image doit être décomposée à l'aide de filtres gaussiens, de telle manière à ce que les bandes-passantes (ce qui est gardé dans le filtre) correspondent à des fréquences spatiales bien définies : 16, 8, 4, 2, 1 et 0.5 cpd<sup>1</sup>.

On réalise ensuite un échantillonnage spectral dans le but de reproduire la réponse des différents cônes et bâtonnets présents dans la rétine. Cet échantillonnage se fait par l'intégration de la distribution de radiance pour chaque pixel après multiplication par l'efficacité lumineuse des photo-récepteurs mis en jeu. L'efficacité pour les cônes est donnée par les

1. cpd : cycles par degrés.

courbes de Hunt, Pointer et Estevez [Fairchild, 1998] tandis que l'efficacité des bâtonnets est donnée par l'observateur scotopique de la CIE [Wyszecki & Stiles, 2000].

Cependant, dans la majeure partie des cas, la valeur de radiance du pixel n'est pas disponible et doit être calculée pour les cônes par une transformation linéaire depuis les valeurs du tri-stimulus CIE 1931 XYZ (Eq. 2.1). Pour les bâtonnets, une approximation est nécessaire, en utilisant l'Eq. 2.2 :

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.1)$$

$$R = -0.702 \cdot X + 1.039 \cdot Y + 0.433 \cdot Z \quad (2.2)$$

Les différents canaux sont ensuite soumis à un traitement spatial, une décomposition en pyramide laplacienne de 7 étages [Burt & Adelson, 1983]. Les 6 premiers étages sont ensuite convertis en images contrastées [Peli, 1990] au moyen d'un gain de luminance (Eq. 2.3) :

$$\begin{aligned} G_{cone}(I) &= \frac{1}{0.555 \cdot (I+1.0)^{0.85}} \\ G_{rod}(I) &= \left[ \frac{10}{I^2+10} \right] \cdot \left[ \frac{1}{0.908 \cdot (I+0.001)^{0.85}} \right] \end{aligned} \quad (2.3)$$

On obtient au final une image au contraste adapté pour un certain niveau ( $ACI_n$  - Image avec contraste adapté au niveau  $n$ ), avec  $LP_n$  la réponse au filtre passe-bas de l'image de niveau  $n$  ( $n$  allant de 1 à 6, en fonction de l'étage de la pyramide concerné) (Eq. 2.4) :

$$ACI_n = G(LP_{n+1}) \cdot [LP_n - LP_{n+1}] \quad (2.4)$$

Les images contrastée obtenue juste avant, sont ensuite transformées encore, jusqu'au système  $AC_1C_2$  qui reprend la Théorie des processus antagoniques, c'est à dire la découpage du traitement de l'image par le cerveau en un canal achromatique ( $A$ ), et deux canaux chromatiques ( $C_1$  et  $C_2$ ), voir sections précédentes sur la couleur. La transformation se fait au moyen d'une banale matrice de passage [Fairchild, 1998; Hunt, 1995] :

$$\begin{bmatrix} A \\ C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 2.0 & 1.0 & 0.05 \\ 1.0 & -1.09 & 0.09 \\ 0.11 & 0.11 & -0.22 \end{bmatrix} \cdot \begin{bmatrix} L \\ M \\ S \end{bmatrix} \quad (2.5)$$

Le signal passe en suite à travers des transducteurs basés sur le contraste pour reproduire la sensibilité au contraste de l'œil. Pour rappel, un transducteur<sup>1</sup> est une fonction ou un mécanisme qui permet de transformer une grandeur en une autre, comme par exemple

1. Transducteur. (s.d.). Dans *Dictionnaire Larousse en ligne*. Repéré à <http://www.larousse.fr/dictionnaires/francais/transducteur/>

Fréquence spatiale (cpd)	0.5	1.0	2.0	4.0	8.0	16.0
p (achromatique)	1.93	1.35	1.15	1.04	1.15	1.40
p (chromatique)	1.93	1.93	2.35	2.85	-	-
p (bâtonnets)	3.39	3.39	4.50	7.64	-	-

TABLE 2.1 – Valeurs de la puissance dans les fonctions de transducteurs de [Watson & Solomon, 1997]

un mouvement en électricité ou une lumière en électricité. Les transducteurs sont ici utilisés pour supprimer de l'image le contraste qui serait imperceptible dans les conditions (d'éclairage) données. Leurs valeurs sont limitées à 50 pour reproduire les limitations en luminosité du système visuel. Les auteurs du modèle utilisent les équations de [Watson & Solomon, 1997] modifiées pour être inversibles (Eq. 2.6) :

$$\begin{aligned} T_{cone,achromatic}(c) &= \begin{cases} 22.4 \cdot \sqrt{\frac{c}{0.536}} & \text{si } c \geq 0.536 \\ 22.4 \cdot \left(\frac{c}{0.536}\right)^p & \text{sinon} \end{cases} \\ T_{cone,chromatic}(c) &= \begin{cases} 22.4 \cdot \sqrt{\frac{c}{0.176}} & \text{si } c \geq 0.176 \\ 22.4 \cdot \left(\frac{c}{0.176}\right)^p & \text{sinon} \end{cases} \\ T_{rod}(c) &= \begin{cases} 22.4 \cdot \sqrt{\frac{c}{0.0335}} & \text{si } c \geq 0.0335 \\ 22.4 \cdot \left(\frac{c}{0.0335}\right)^p & \text{sinon} \end{cases} \end{aligned} \quad (2.6)$$

Les valeurs de p sont données dans la Table 2.

Le canal chromatique final est en fait une combinaison des valeurs de la composante A des tri-stimulus des cônes et des bâtonnets. Le calcul se fait au moyen de l'Eq. 2.7. La valeur 7 est obtenue par calibration :

$$A_{total} = A_{cones} + \frac{A_{batonnets}}{7} \quad (2.7)$$

Cette étape conclue le traitement des 6 premiers étages de la pyramide de décomposition.

Après quoi, le traitement du dernier et septième étage de la pyramide se fait via des transducteurs passe bas (Eq. 2.8) [Fairchild, 1998] :

$$\begin{cases} T_{LP,cones,achromatique}(LP) = 30.5 \cdot \sqrt{LP} \\ T_{LP,cones,chromatique}(LP) = 53.3 \cdot \sqrt{LP} \\ T_{LP,batonnets}(LP) = 122 \cdot \sqrt{LP} \end{cases} \quad (2.8)$$

Au point intermédiaire, à l'issue du modèle visuel, on obtient donc les résultats suivants :

- 1 canal achromatique
- 2 canaux chromatiques
- 6 images issues d'un filtre spatial bande-passante

- 1 image issue d'un filtre passe-bas

Enfin, on effectue le processus de reconstruction de l'image en faisant un traitement étape-par-étape inverse par rapport au modèle visuel de manière à recréer le signal perçu par les cônes pour l'afficher. On prend les fonctions inverses des transducteurs évoqués précédemment (Cf. Eq. ....). On passe d'une décomposition de la lumière en mode chromatique/achromatique ( $AC_1C_2$ ) au découpage reproduisant le fonctionnement des cônes et bâtonnets ( $LMS$ ), en utilisant la matrice inverse et enfin on applique un mécanisme de gain inverse en fonction des conditions d'affichage. Pour finir, on effectue la dernière transformation en allant réalisant une dernière transformation : de  $LMS$  à l'espace colorimétrique adapté (CIECAM, CIELAB, CIELUV, RGB, ...)

En conclusion, [Pattanaik *et al.*, 1998] ont mis sur pied un modèle extrêmement précis et pertinent, composé de sous-modèles disponibles dans la littérature. Cependant, ce modèle est statique, c'est à dire qu'il ne prend pas en compte la composante temporelle et ne fonctionne que pour traiter une image et non pas intégré dans une boucle de calcul d'image en temps réel. Nous allons maintenant aborder le modèle de [Mantiuk & Ramponi, 2015].

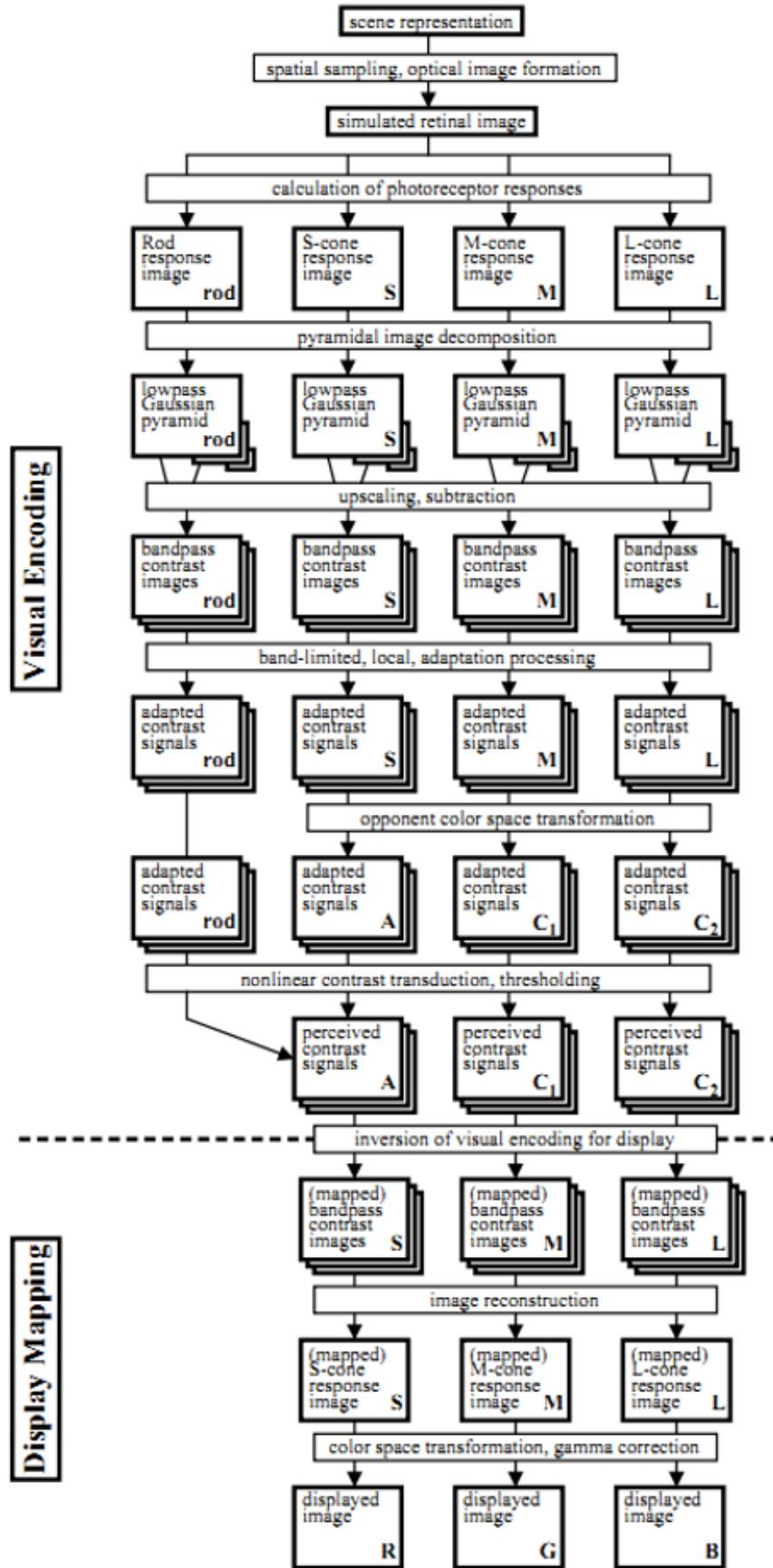


FIGURE 2.2 – Schéma exhaustif du traitement de l'image par le modèle de [Pattanaik et al., 1998]

## CHAPITRE

### 3

# MODÈLE DE MANTIUK & RAMPONI

Comme on a pu voir précédemment, l'âge de l'observateur à une grande influence sur un grand nombre de paramètres dans la vision ; tant de manière physiologique que cognitive. Cependant, le modèle que nous présentons ici, celui de [Mantiuk & Ramponi, 2015], est basé sur le prédicteur HDR-VDP-2 (voir précédemment) altéré selon trois grands axes : le vieillissement du cristallin, la réduction du diamètre de la pupille avec l'âge, les éblouissement invalidants et une composante de dégradation cognitive.

La contribution la plus importante à la dégradation de la vision au cours du temps vient du vieillissement du cristallin. La lentille continue de grossir et de générer des nouvelles fibres, compressant et accumulant ainsi les fibres plus anciennes dans la région centrale du cristallin. Celui-ci devient donc de plus en plus rigide et de moins en moins transparent.

Le cristallin est modélisé comme un filtre optique dépendant de la longueur d'onde et qui réduit la quantité de lumière que la rétine reçoit. Cependant ce filtre a peu d'effet en vision photopique (vision de jour) et est plus efficace en dessous de  $3Cd/m^2$ , c'est à dire dans les domaines mésopique et scotopique. [Mantiuk & Ramponi, 2015] reprennent ici le modèle de [Pokorny *et al.*, 1987]. pour le vieillissement du cristallin (Cf. Eq. 3.1) :

$$O_D(a) = \begin{cases} L_1 \cdot (1 + 0.02 \cdot (a - 32)) + L_2 & a \leq 60 \\ L_1 \cdot (1.56 + 0.0667 \cdot (a - 60)) + L_2 & a \geq 60 \end{cases} \quad (3.1)$$

Avec  $O_D$  la densité optique du cristallin,  $a$  l'âge du sujet et  $L_1$ ,  $L_2$  des constantes issues de tables dans le modèle de Porkony *et al.*.

[Mantiuk & Ramponi, 2015] incluent aussi dans leur modèle un mécanisme d'efficacité lumineuse (valable seulement dans le domaine photopique) qui a fut validé et entériné par

la CIE en 1924. Les résultats liés à ce modèle sont disponibles chez [Sagawa & Takahashi, 2001].

L'éblouissement (ou *Sensibility Glare* en anglais) est une pollution lumineuse qui intervient en présence de lumière très vive directement dans le champ de vision. Cette pollution atténue les hautes fréquences spatiales et réduit le contraste perçu sur la rétine. La prise en compte de la sensibilité à l'éblouissement en fonction de l'âge est elle aussi étayée par un modèle de la CIE : [Vos & van den Berg, 1999].

Le diamètre pupillaire moyen tend à diminuer avec l'âge, on parle alors de *myosis* (l'effet contraire étant la *mydriase*). Ce phénomène a été modélisé par un grand nombre de chercheurs différents ; [Watson & Yellott, 2012] proposent à la fois une review des différents modèles mais surtout un modèle unifié pour l'évolution du diamètre de la pupille. C'est ce modèle qui est utilisé dans le modèle de vision avec vieillissement de l'œil de [Mantiuk & Ramponi, 2015].

Le modèle est le suivant (Cf. Eq. 3.2) :

$$\begin{cases} S(l, a, f) = (a - 28.58) \cdot (0.02132 - 0.009562 \cdot D_{SD}) \\ D_{SD} = 7.75 - \frac{5.75 \cdot k}{k+2} \\ k = \left(\frac{L \cdot f}{846}\right)^{0.41} \end{cases} \quad (3.2)$$

Avec  $S$  le diamètre en question,  $a$  l'âge de l'individu,  $L$  le niveau de luminosité auquel l'œil est adapté (en  $Cd/m^2$ ) et  $f$  la zone de champ.

Ensuite, [Burton *et al.*, 1993] ont mesuré la perte de sensibilité causée par l'affectation des récepteurs et des processus cognitifs entre des jeunes adultes et des personnes âgées. Ces mesures ont été réalisées avec une laser permettant d'ignorer le système optique et d'envoyer directement une image sur la rétine. Les résultats donnent de légères variations entre les populations d'âge différent. [Mantiuk & Ramponi, 2015] proposent alors leur propre modèle empirique (Eq. 3.3) :

$$\log_{10}(\Delta S) = -(\beta \cdot \log_2(\rho + a)) \cdot \max(a - 24; 0) \quad (3.3)$$

Avec  $a$  l'âge du sujet,  $\rho$  la fréquence spatiale du stimulus (en  $cpd$ ), et  $\alpha = 0.75$  et  $\beta = 0.00195$  sont des valeurs empiriques proposées par les auteurs du modèle.

Enfin, cette modélisation est toutefois à prendre avec recul car un grand nombre de valeurs utilisées ici ont été mesurées pendant des expérimentations réalisées sur des écrans CRT (Cathodiques) ayant un spectre de luminosité assez réduit (de 0.1 à  $80Cd/m^2$ ).

# CONCLUSION

L'étude des modèles de visions et leurs traductions vers le domaine de l'informatique semble être un domaine déjà fortement décrit et bien maîtrisé. Une contribution directe dans ce domaine n'est alors ni très évidente ni très pertinente. On pourrait à la rigueur tenter d'appliquer le modèle de Pattanaik *et al.* en temps réel mais la portée scientifique semble limitée. On décide alors de laisser cette approche de côté et de changer d'angle d'attaque.

Plutôt que de s'intéresser à une capture d'image respectant le même processus que la vision humaine, on se penche sur un autre aspect jouant un rôle prépondérant dans le réalisme : l'affichage correct de ce qui a été capturé dans la scène 3D. En effet, si une fois l'image capturée proprement (via une mise en place du modèle de Pattanaik *et al.* par exemple), si le système n'est pas capable de restituer l'image dans la même gamme de perception que celle de l'œil humain, le réalisme vis à vis de l'utilisateur sera amoindri.

## CONCLUSION

---

## **Troisième partie**

# **Deuxième approche : score de réalisme**



# INTRODUCTION

Pour rappel, on avait posé dans la définition du réalisme (*cf.* Chapitre I.1) le cadre d'étude suivant :

« Pour le cas de cette thèse, la définition du réalisme qui aura été retenue - et qui sera sous-entendue quand on utilisera le mot *réalisme* seul- est celle de la proximité physiologique avec le système visuel humain. [...] Le réalisme physiologique doit, par construction et par définition, se baser sur le système visuel humain. »

La première approche, qui était de développer un "modèle de vision" traduit pour la Réalité Virtuelle et les caméras qui extraient les images des scènes 3D, s'est vite révélée être une impasse ou en tout cas une voie offrant peu de débouchés tant ces modèles étaient déjà largement décrits dans la littérature.

On choisit alors une approche différente : plutôt que de "faire" du réalisme, il pourrait être intéressant de "quantifier" le réalisme qu'un-e système/application fait. En effet, si on entend souvent parler de réalisme, son appréciation est souvent binaire : soit un système ou une application est réaliste, soit il/elle ne l'est pas. De plus, il est difficile de pouvoir dire concrètement ce qui rend un système moins réaliste qu'un autre. Pour ce faire, il faudrait pouvoir être capable de mettre des valeurs et des critères objectifs à remplir. On aurait alors un moyen efficace et juste de juger du réalisme d'un système et/ou de comparaison entre plusieurs éléments.

A notre connaissance, il existe peu voire très peu de modèles ayant pour but de quantifier objectivement le réalisme physiologique dans la littérature. Ce type de modèle ne doit pas être confondu avec les méthodes de qualifications de la qualité de l'image qui elles sont basées sur une comparaison entre une image source et cette même image après qu'elle soit passée à travers tout un processus de transformation (comme par exemple, un encodage vidéo) [Čadík, 2004; Winkler, 2000] (*cf.* Figure 1). Actuellement, il existe déjà un grand

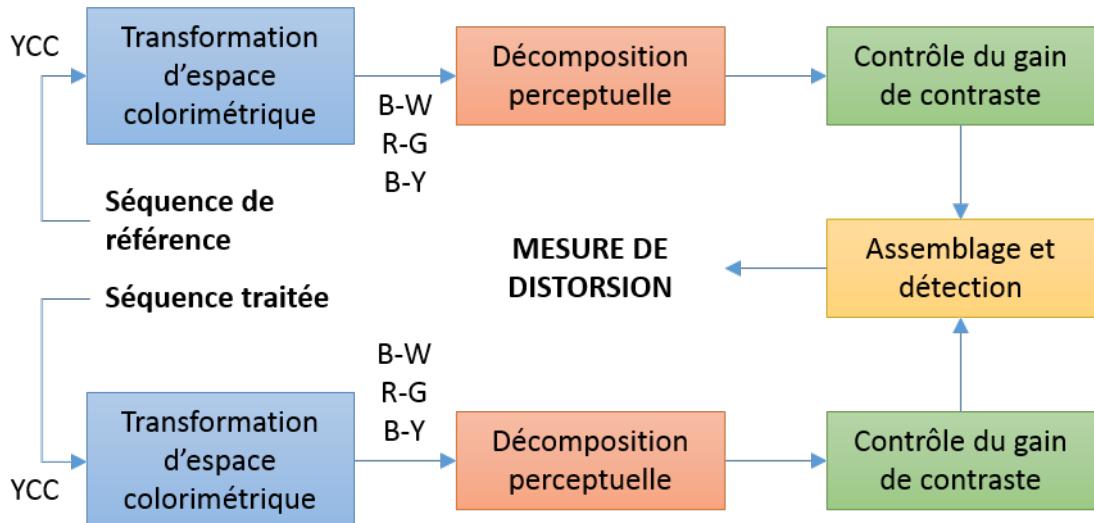


FIGURE 1 – Fonctionnement du processus de mesure de qualité d'image, pour des séquences vidéo 2D. Traduction d'une figure reprise de Winkler [2000]

nombre de méthodes pour la qualité d'image, notamment pour des images 2D (images statiques ou vidéos). Ces dernières années, l'accent a plutôt été mis sur le développement de techniques de qualification de la qualité pour des images 3D. On peut trouver un certain nombre de review de ces techniques dans la littérature [Moorthy & Bovik, 2013; Moorthy *et al.*, 2013] voire même avec une dimension subjective en plus [Beghdadi *et al.*, 2013].

Le réalisme en tant que tel est quant à lui en général plutôt mesuré par une question directe sur le réalisme de l'expérience à laquelle le sujet vient de participer (ce qui est à relier avec la quatrième acceptation du réalisme, le réalisme *psychologique* et non pas le réalisme *physiologique*). Des approches plus détaillées peuvent se servir de questionnaires [Fiard *et al.*, 2014; Fuccentese *et al.*, 2015].

Néanmoins, on trouve quelques approches objectives pour la définition du réalisme bien que ce ne soit pas forcément leur objectif principal :

- Le modèle de Rose [Burgess, 1999; Rose, 1948].
- Le modèle d'observateur idéal [Geisler, 2003].

## Modèle de Rose

Le modèle de Rose est une première tentative (à l'époque pour la télévision en noir et blanc !) de caractérisation de "l'enregistreur d'image" idéal, vis à vis de la vision humaine. Le modèle est conçu pour la capture d'un objet unique, c'est à dire que le modèle donne les caractéristiques qu'un objet unique doit avoir pour être capturé par l'enregistreur idéal. On peut le décrire par l'équation suivante (Eq. 1), avec  $B$  la luminance de l'objet en foot-Lamberts<sup>1</sup>,  $C$  le contraste de l'objet comparé au background (en pourcentage) et  $\alpha$  l'angle

1. NB :  $1 fL \approx 3.426 \text{ cd/m}^2$

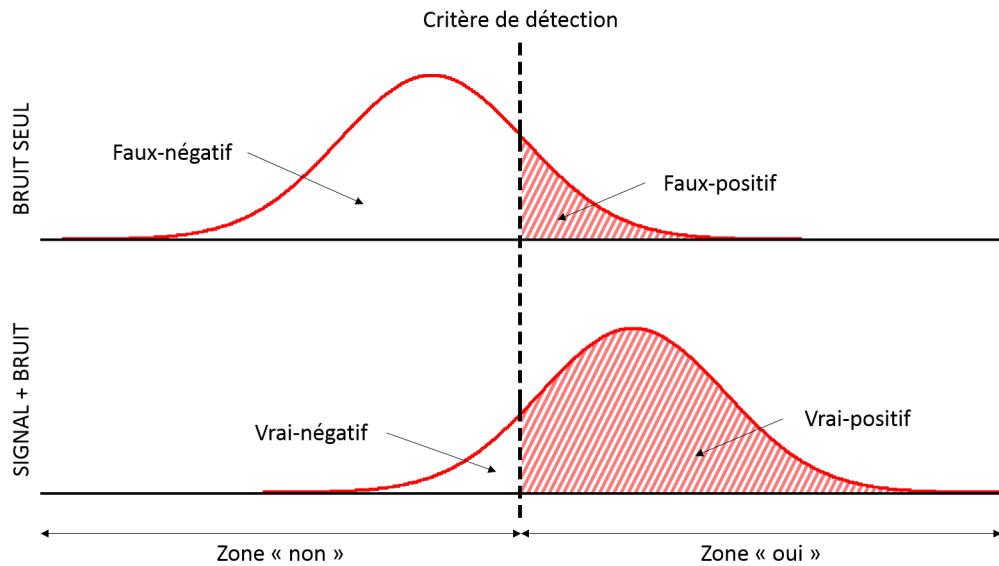


FIGURE 2 – Illustration de la théorie de détection du signal. Répartition des probabilités de détection autour d'un seuil de détection déterminé (critère de détection)

de visuel sous lequel l'objet est vu par l'enregistreur.

$$BC^2\alpha^2 = \text{constant} \quad (1)$$

## Observateur idéal & théorie de détection du signal

La définition d'un "observateur idéal" est basée sur une théorie appelée "détection du signal" (Signal Dedection Theory). Cette théorie décrit la capacité d'un système à discerner un pattern spécifique au milieu d'autres patterns ou bien au milieu d'un bruit (cf. Fig. 2). La théorie de détection du signal est également basée sur la théorie bayésienne qui est l'interprétation du concept de probabilités comme état de savoir -le pattern est visible ou non- plutôt qu'une fréquence ou une propension d'apparition d'un phénomène. La théorie de détection du signal inclue également des probabilités de détection et des seuils de vision. Cependant, les observateurs idéaux sont utilisés pour une unique tâche bien spécifique (au contraire de la vision que se veut généraliste) telle que la détection de photon, la détection de pattern, l'identification ...

---

## INTRODUCTION

---

## CHAPITRE

# 1

# SCORE DE RÉALISME

## 1.1 Ambition

On a vu dans les modèles présenté précédemment dans l'introduction, et notamment dans celui de Rose qui mêle dans une équation mathématique simple des grandeurs essentielles pour le réalisme (contraste, luminance, taille), une approche pragmatique de la définition du réalisme physiologique. Néanmoins, il manque toujours une dimension à ces modèles : les modèles de Rose et d'observateur idéal prédisent pour l'un les qualités optimales d'affichage et pour l'autre la moment à partir duquel l'œil humain va percevoir l'objet puis une description de la probabilité de plus en plus élevée de le percevoir, mais jamais les deux en même temps.

Au delà de ces modèles, on souhaite être capable de pouvoir quantifier le biais qu'il existe entre ce qui est affiché dans un environnement virtuel et la qualité de perception de l'utilisateur. C'est à dire non seulement être capable de dire quelles sont les conditions optimales pour que le système visuel humain interagisse avec l'affichage immersif mais aussi pouvoir apprécier les conditions non-optimales, tel que le système de probabilité dans le modèle d'observateur idéal. On souhaite néanmoins réfléchir à une autre approche que celle probabiliste pour se rapprocher d'un démarche plus pragmatique, plus «ingénieur».

Cette méthode pourrait alors indiquer à quel point l'interaction entre le système visuel humain et l'affichage est différente de l'interaction qui aurait lieu dans le monde naturel et ce de manière tangible, mesurée. Pour que l'évaluation soit fiable et répétable, la quantification doit être complètement objective et doit dépendre de critères physiques du système d'affichage (hardware).

On propose alors une évaluation de la performance d'un système via un score qui dépeindrait, pour un système immersif donné et pour une situation donnée, à quel point le

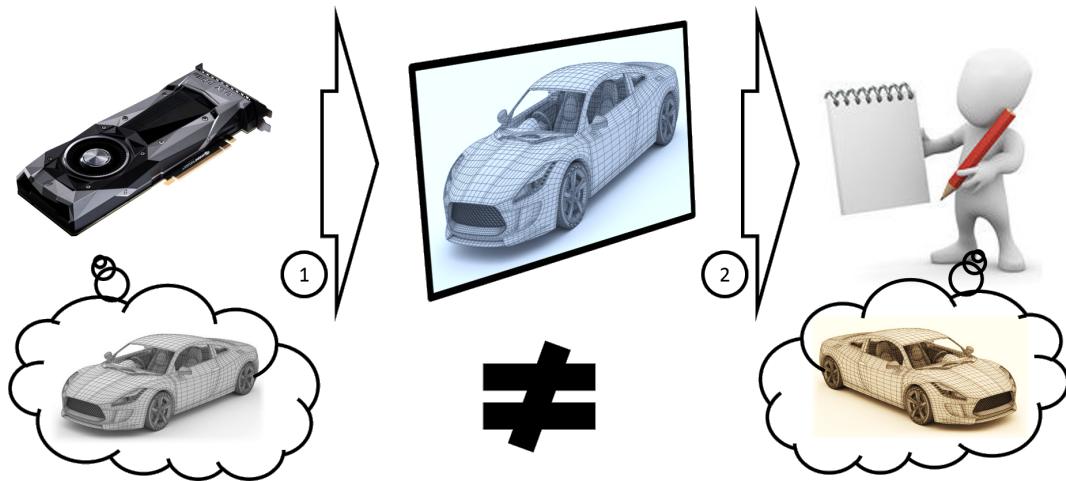


FIGURE 1.1 – Illustration de la différence qu'il existe entre les consignes données par la carte graphique (1) et les informations visuelles reçues par l'observateur (2).

simulateur est efficace pour transmettre le bon niveau d'informations visuelles et, étant dans un système de Réalité Virtuelle, le bon niveau d'indices d'immersion.

## 1.2 Méthodologie

On propose donc un système de score qui serait basé sur le système visuel humain. Ce score est en fait la réunion d'un certain nombre (à définir, voir par la suite) de sous-scores chacun lié à une caractéristique de la vision et/ou de l'immersion liée à la vision (c'est à dire les critères immersifs dérivés directement du système visuel ou du système en lui-même). Chaque critère est jugé indépendamment puis contribue à la note globale via une pondération.

La première étape était donc logiquement d'avoir une connaissance et une compréhension approfondie du système visuel humain. Comme on a pu voir dans la première partie, qui présente cette étude de l'état de l'Art, on s'est intéressé d'abord au fonctionnement de l'œil en lui-même puis de ce qu'il se passe après, que ce soit le traitement nerveux ou le traitement par le cerveau. L'accent a ensuite été mis sur le contraste et la couleur qui sont des deux grandes caractéristiques de l'image ainsi que la perception visuelle de la profondeur. Enfin, certaines influences, classées ici abusivement comme « physiologiques », modifient le processus normal de la vision et doivent être, si possible, prises en compte.

Deuxièmement, une fois cette connaissance acquise il a fallut découper le processus de vision et les facteurs d'immersions en différents critères. Ce processus a été long et compliqué, et fait l'objet d'un chapitre pour le détailler (voir par la suite). Ces critères se veulent indépendants les uns des autres mais des relations ont quand même parfois été relevées entre certains critères. Par exemple, la précision de la vision dépend dans certains cas de la luminosité : une situation bien éclairée permet de voir bien plus précisément qu'une situation mal éclairée. Pour plus de simplicité et de clarté, on a regroupé les critères par domaine.

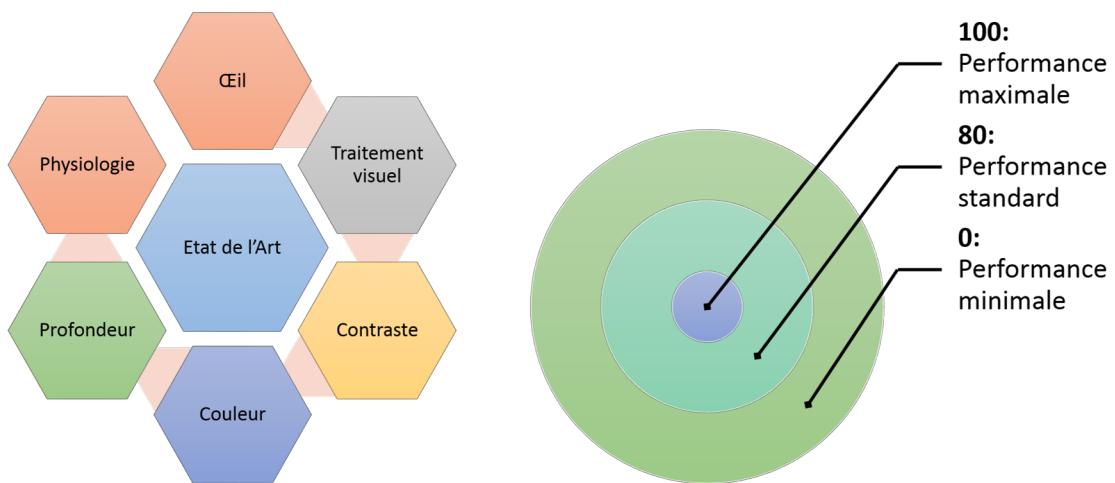


FIGURE 1.2 – Résumé des étapes d'état de l'Art pour la construction du modèle et répartition des objectifs de représentation des notes intermédiaires dans le score de réalisme.

Troisièmement, une fois les critères établis, il convient de les évaluer. Pour ce faire, on propose d'attribuer une note comprise en 0 et 100 à chacun des critères du modèle. Cette note modélise le fonctionnement du critère en question : la note de 0 représente les conditions dans lesquelles le système de vision n'est pas capable d'utiliser le mécanisme (ou peut l'utiliser mais n'en tire aucune information utile) tandis que la note de 100 représente le fonctionnement ou la performance maximal-e du mécanisme de vision. On rajoute également une note intermédiaire à 80 qui représente le fonctionnement nominal du mécanisme de vision. La valeur de 80 est plus probablement la valeur de score à viser car la valeur de 100 représente la maximum possible de l'œil humain et pourrait être considéré comme de la sur-qualité par rapport à la performance nominale. Il est évident que dans certaines applications, atteindre la valeur de 100 peut être une nécessité. Cette démarche est confortée par le fait que la Commission Internationale pour l'Eclairage (CIE) procède de la même manière (noter sur une valeur de 100 en adjugeant 80 à la valeur nominale) dans son modèle de performance visuelle en fonction de la luminosité et du contraste (voir plus loin la partie spécifique sur la luminance et le contraste).

La notation en elle-même se fait, principalement, basée sur la littérature : on cherche les valeurs correspondantes aux fonctionnements minimal, nominal et maximal. On cherche ensuite une loi mathématique simple permettant de décrire l'évolution de la note en fonction des valeurs clefs du mécanisme visuel. Si on a pu appliquer cette méthode pour la majorité des cas, certains critères ne se prêtent pas à un tel traitement et ont bénéficié d'une notation particulière. Cette notation sera explicitée plus bas, au cas par cas, dans la présentation des critères retenus dans la dernière version du modèle.



# CHAPITRE

## 2

# PROPOSITION DE MODÈLE

Pour faciliter la lecture et la compréhension du chapitre, on propose d'expliquer sommairement ici certaines distinctions entre des termes proches qui seront utilisés. Tous ces termes seront explicités en détail par la suite dans les deux prochains chapitres.

L'« acuité monoscopique », aussi connue sous le nom d'acuité visuelle est définie par le *Vulgaris Médical*<sup>2</sup> de la sorte :

L'acuité visuelle est la capacité de discernement des informations apportées au cerveau par la vue. C'est la performance visuelle, elle mesure le pouvoir de l'œil à distinguer nettement les détails, avec et sans lunettes.

Elle ne doit pas être confondue avec « l'acuité stéréoscopique » qui est la capacité de l'œil à distinguer une très petite différence de profondeur à une distance donnée.

De même, le « champ de vision » est ce qu'un individu perçoit à un instant  $t$  sans possibilité de bouger la tête ou l'œil dans son orbite. Le « champ de regard » englobe tout ce qu'un individu peut voir en ayant le droit de bouger la tête et les yeux. Par définition, le champ de regard est donc supérieur au champ de vision.

Enfin, la gamme dynamique (Dynamic Range en anglais) est, dans le domaine de l'image, le rapport entre le minimum et le maximum de luminance atteignable. Il est défini par rapport à la courbe de gamma obtenue avec un écran cathodique. Le HDR (ou High Dynamic Range) représente un ratio entre le minimum et le maximum de luminance possible supérieur au ratio standard. HDR sous entend en général une technique d'imagerie pour augmenter artificiellement les extrema lumineux pendant la capture d'image.

<sup>2</sup>. Acuité visuelle. Dans *Vulgaris Médical*. Consulté le 20/09/2017, <http://www.vulgaris-medical.com/encyclopedie-medicale/acuite-visuelle> (Lien court : <https://lc.cx/GfSj>)

## 2.1 Propositions préliminaires

Avant d'arriver à la modélisation finale du score (voir ci-après) nous sommes passé par un certain nombre de pistes et de modélisations. Nous allons présenter ici les différentes étapes qui ont vu le jour.

### 2.1.1 Première modélisation (Fig. 2.1)

Initialement, la proposition était de faire un score simple qui réunirait les critères de base, accompagné d'un score avancé qui servirait à raffiner l'évaluation du système. La composition du modèle est basée sur les capacités de la vision qui sont l'acuité, l'attention périphérique, la coordination des yeux, la perception de la profondeur, le focus et la perception des couleurs. Si le focus reste un problème insoluble à cause du conflit accommodation-vergence toutes les autres capacités sont néanmoins traitable. Un critère répondant à la coordination des yeux n'est cependant pas encore proposé dans cette première modélisation. Le score simple était alors décliné selon les critères suivants :

- le contraste
- l'acuité monoscopique
- l'acuité stéréoscopique
- la latence
- le nombre d'images par seconde
- le respect de l'échelle 1 :1
- la couverture du nombre de couleurs visibles
- la taille de l'HDR

Tandis que le test avancé rajoutait en plus du score simple qui restait à effectuer :

- le champ de vision
- le champ de regard
- les maximum et minimum de luminance

Le tout étant rassemblé dans une équation de somme pondérée (Eq. 2.1), avec  $\Sigma$  la valeur du score général du système,  $\lambda_i$  la pondération du critère  $i$ ,  $\sigma_i$  la valeur observée du critère et  $F_{100}$  la fonction donnant la note entre 0 et 100 du critère.

$$\Sigma = \frac{1}{\sum_{i=1}^n \lambda_i} \sum_{i=1}^n \lambda_i F_{100,i}(\sigma_i) \quad (2.1)$$



FIGURE 2.1 – Première modélisation du score de réalisme.

### 2.1.2 Seconde modélisation (Fig. 2.2)

La seconde étape de la modélisation est assez proche de la première puisqu'on garde une liste quasiment similaire de critères de base : contraste, acuités monoscopique et stéréoscopique, images par seconde, couleur, luminosité et champ de vision. On retire le concept de score simple et de score avancé pour se concentrer sur un mélange des critères des deux scores précédemment envisagés en enlevant cependant les critères de champ de regard, de latence et d'échelle 1 :1. Ces trois derniers facteurs sont isolés dans une catégorie appelée « facteurs limitant » et doivent interdire à la note globale de passer au delà d'un certain seuil en fonction de leur valeur. Ces facteurs sont de type booléens : si le système propose une valeur moins bonne que celle demandé par le facteur limitant, la note globale est automatiquement tronquée.

Il était également envisagé de permettre un système de logique combinatoire entre les facteurs limitant. Par exemple, si le score de la latence est inférieur à 50 la note globale est tronquée à 90 alors que si le score de la latence est inférieur à 60 et que le score de champ de regard est inférieur à 40 le score global est tronqué cette fois à 75.

C'est à partir de cette modélisation qu'apparaissent les jalons spécifiques de performance : la note de 0 pour le minimum de performance possible ou acceptable, la note de 80 pour la performance standard du système visuel humain et la note de 100 pour sa performance maximale. Un jalon optionnel (en fonction de la littérature) à 60 pour une performance dite « acceptable » a même été un temps envisagé avant d'être rejeté faute de pertinence et de répétabilité sur tous les critères.

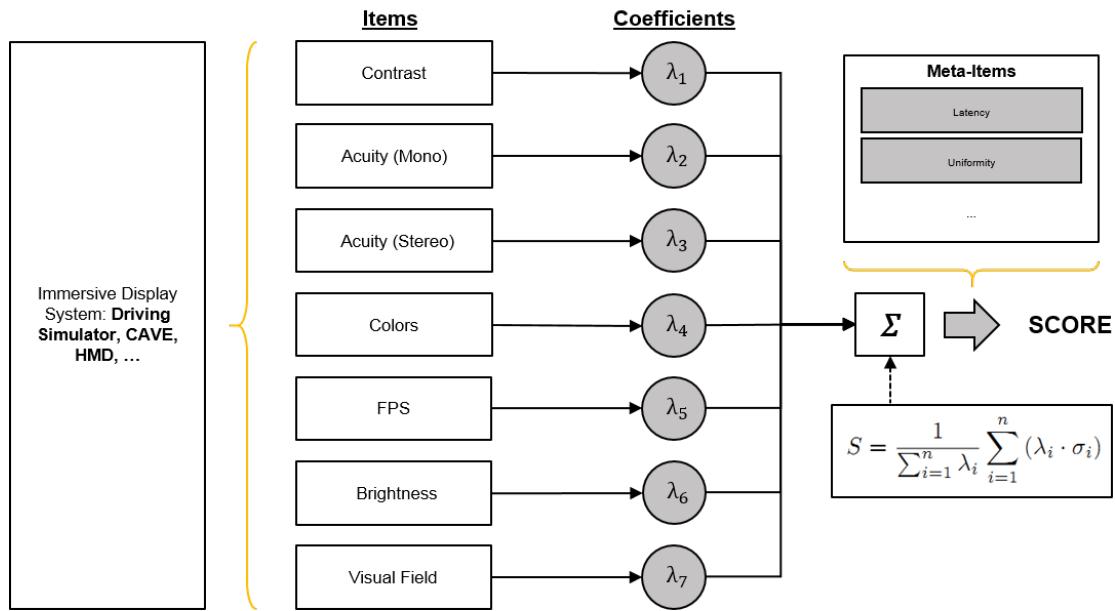


FIGURE 2.2 – Deuxième modélisation du score de réalisme.

### 2.1.3 Troisième modélisation (Fig. 2.3)

La troisième modélisation est sûrement celle qui introduit le plus de changements tant dans la structure que dans la réflexion et le nombre de critères retenus. Le nombre de critère est doublé, passant de 7 à 14. Le concept de facteurs limitant est supprimé. A l'inverse, les critères ne sont plus regroupés en un seul et même ensemble mais sont répartis dans 3 pôles : le pôle « Affichage », le pôle « Immersion » et le pôle « Observateur ». Pour le pôle observateur on choisit d'extraire des autres pôles et de réunir tous les critères nécessitant des informations sur la personne, typiquement sa position. Par exemple, dans un dispositif de type CAVE, la luminosité reste la même indépendamment où que soit la personne tandis que la taille perçue du pixel dépendra de la distance à l'écran.

Chaque pôle peut ainsi recevoir un score qui lui est propre et le système évalué peut ainsi avoir un score d'affichage honnête mais une mauvaise note en immersion sans que cela dégrade l'évaluation de l'affichage. On met ainsi en évidence trois scores de premier niveau : le score d'affichage, le score d'immersion et le score perceptuel (en rapport avec la catégorie observateur). Les scores d'affichage et d'immersion sont ensuite réunis dans un score technologique, par opposition au score perceptuel pour respecter la dichotomie réalisée entre les critères systémiques pur des critères nécessitant des informations sur l'observateur. Enfin, les scores technologique et perceptuel sont assemblés afin de donner le score global du système immersif évalué.

Si la majorité des critères ont une note basée sur des points clef de la littérature, certains points sont plus compliqués à aborder de la sorte étant donné qu'ils ont un fonctionnement plutôt binaire. On les appelle « critères booléens ». Les critères concernés sont les suivants : tracking, stéréoscopie, vergence et accommodation. Par exemple, le critère de stéréoscopie peut n'être jugé que sur la présence ou l'absence de stéréoscopie dans le système immersif. Pour ce critère en particulier, une piste de recherche sera évoquée plus loin.

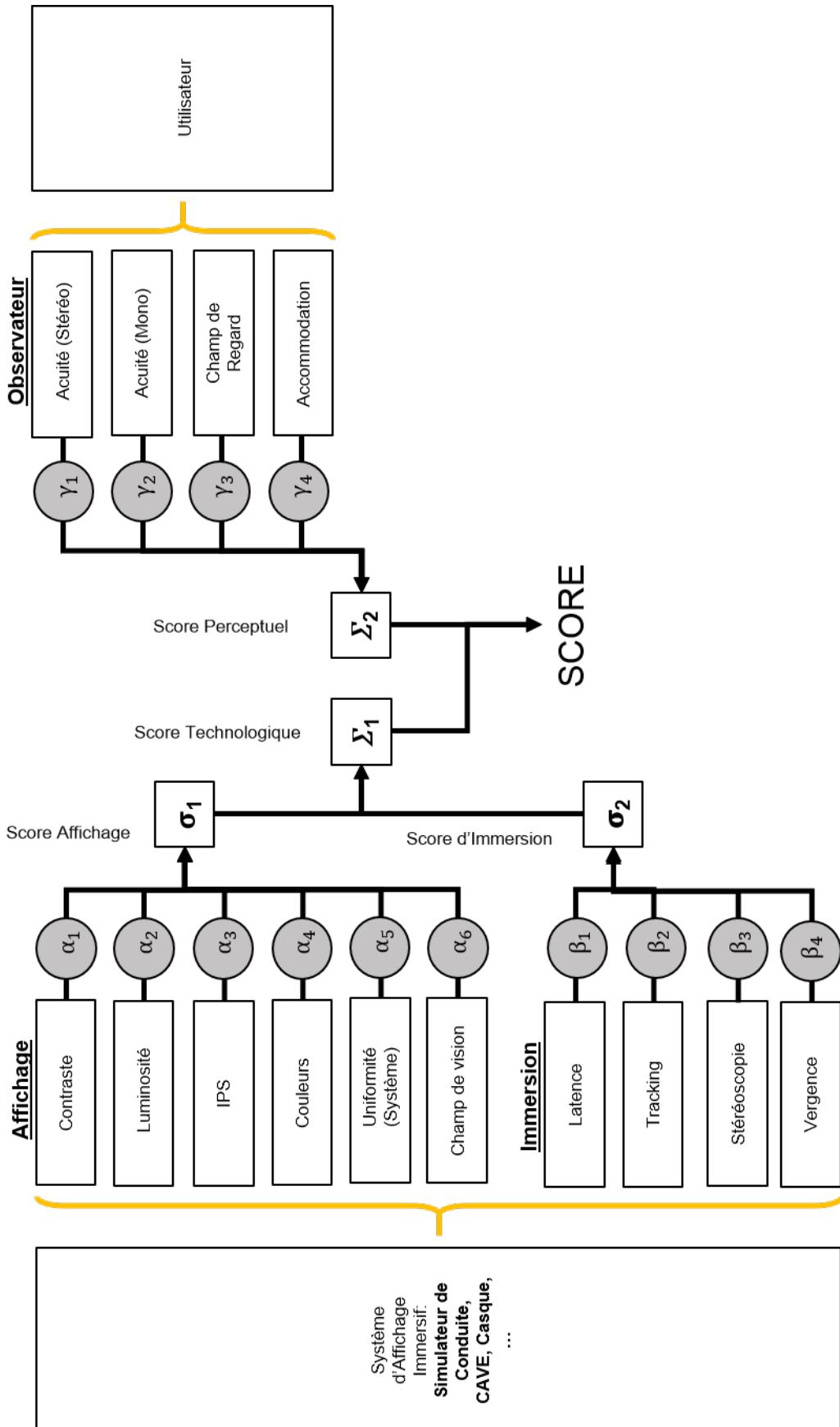


FIGURE 2.3 – Troisième modélisation du score de réalisme.

## 2.2 Modèle définitif

En définitive, après un certain nombre d’itérations, de reconstructions, de nouvelles réflexions et de correction -notamment grâce à l’avis de reviewers des publications qui présentent ce modèle de score- on parvient à une version que l’on estime satisfaisante.

On propose un modèle de score divisé en douze critères distribués en deux sections : une partie indices de vision et une partie indices d’immersion. La division des critères est faite telle que la première partie représente les informations qu’un œil recevrait à l’intérieur d’un simulateur tandis que la deuxième catégorie concerne ce que le système immersif offre à l’utilisateur pour participer à l’immersion. L’immersion est évidemment également réalisée via d’autres processus and nous avons seulement couvert les indices liés au hardware. Les critères sont répertoriés comme suit. Une vue d’ensemble du modèle est présentée Fig. 2.4.

- **Indices de vision :** contraste et luminosité (luminance), images par seconde, nombre de couleurs affichables, champ de vision, acuités monoscopique et stéréoscopique.
- **Indices d’immersion :** latence, champ de regard, stéréoscopie, tracking, uniformité et convergence des caméras (laquelle est directement liée au tracking des yeux).

Les critères de luminance et de contraste sont fusionnés. De part la nature du contraste (un ratio de luminances) ces deux critères semblent indissociables. Le critère de vergence est simplement renommé en « convergence des caméras » car il portait à confusion. Le critère « Accommodation » qui jugeait de la présence ou non du conflit accommodation-vergence n’était pas pertinent car il n’existe aujourd’hui aucun système permettant de visualiser des images en 3D sans générer ce type de conflit ; il a donc été supprimé.

La plus grosse modification reste le passage en deux catégories contre trois dans la précédente modélisation. La catégorie observateur portait à confusion, regroupait des critères qui avaient toute légitimité pour être dans les deux autres catégories. Les critères (moins le critère Accommodation, supprimé) ont donc été redistribués dans leur catégorie légitime. Celles ci sont en plus renommées de manière plus pertinente.

Maintenant que nous avons présenté comment le modèle se présentait, il convient de rentrer plus avant dans le détail et de présenter les critères retenus, un par un. A l’image des différentes modélisations du score, on présentera rapidement les étapes de notation quand elles existent.

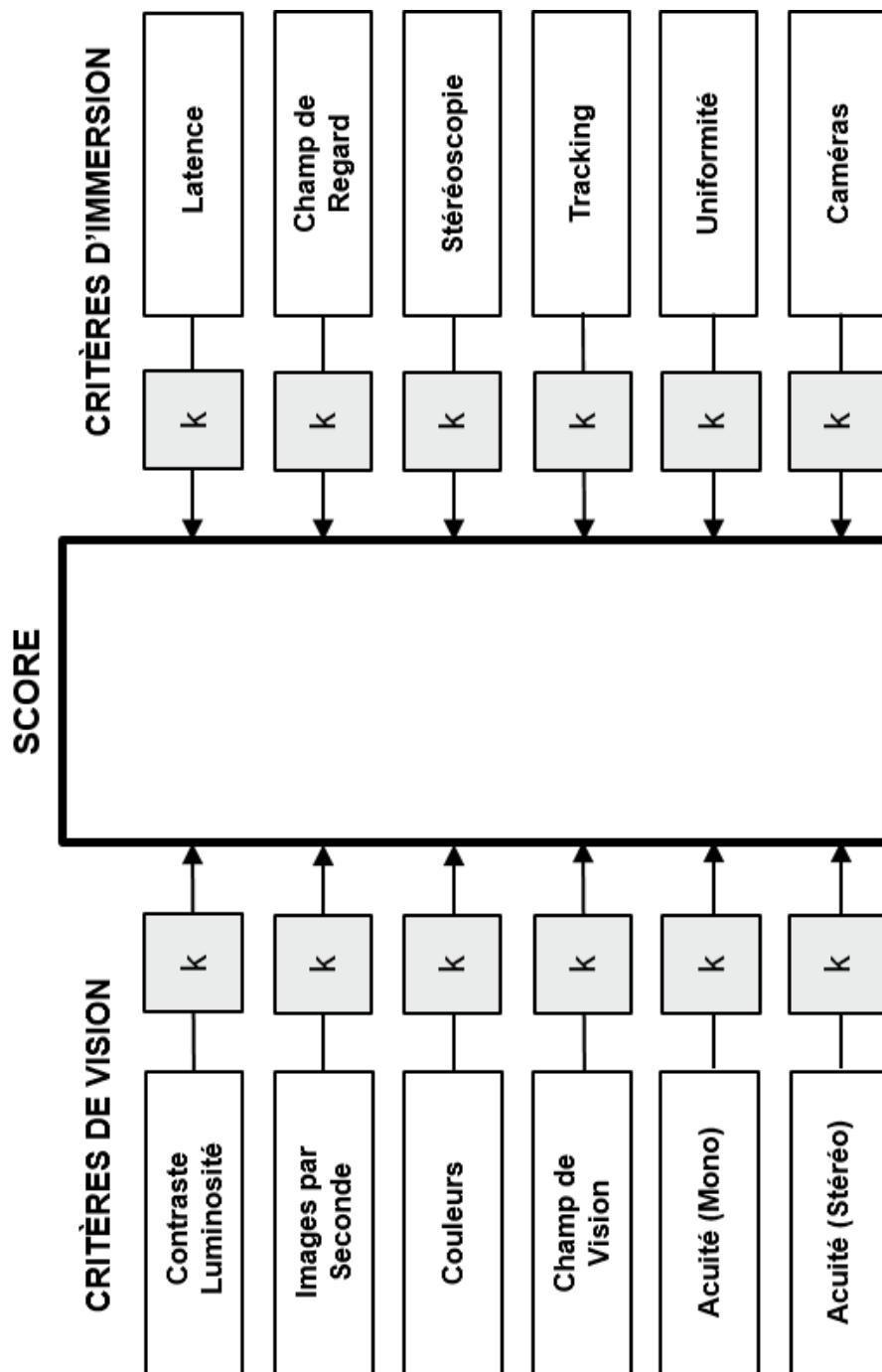


FIGURE 2.4 – Dernière modélisation du score de réalisme.



## CHAPITRE

# 3

## 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.

### 3.1 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.

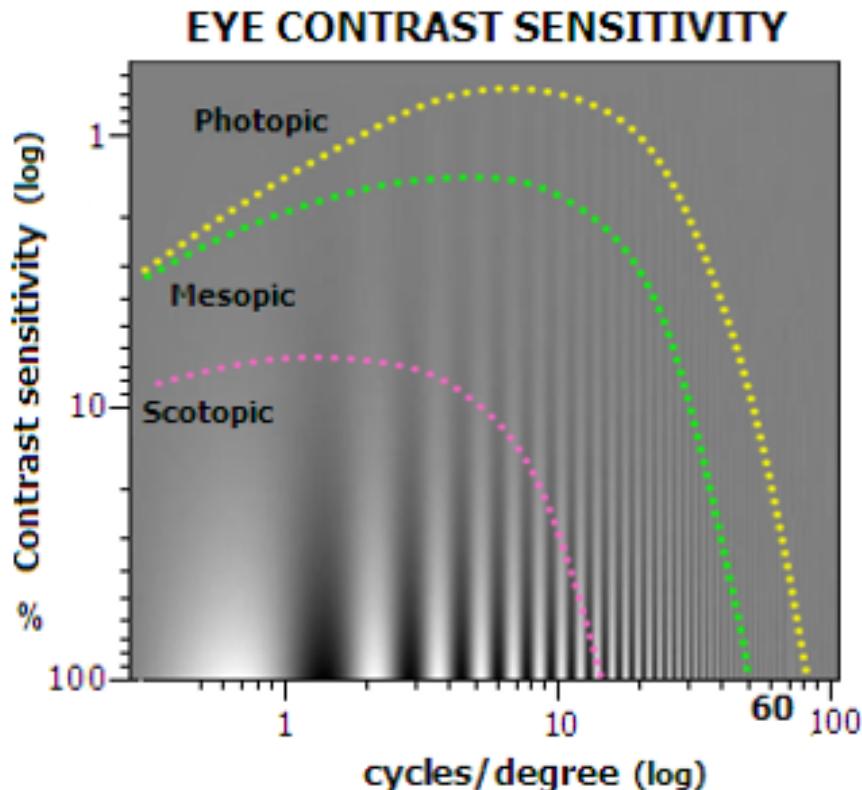


FIGURE 3.1 – Fonctions de sensibilité au contraste pour les domaines photopiques, scotopiques et mésopiques.

### 3.1.1 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. 3.1 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.

### 3.1.2 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. 3.2. 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.

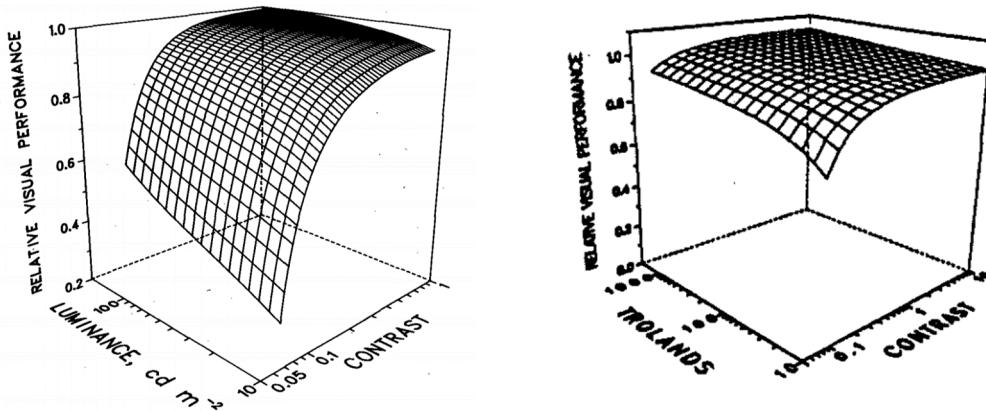


FIGURE 3.2 – Exemple de performance visuelle pour deux tailles de cible différentes

## 3.2 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).

### 3.2.1 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. 3.3. 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.

### 3.2.2 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

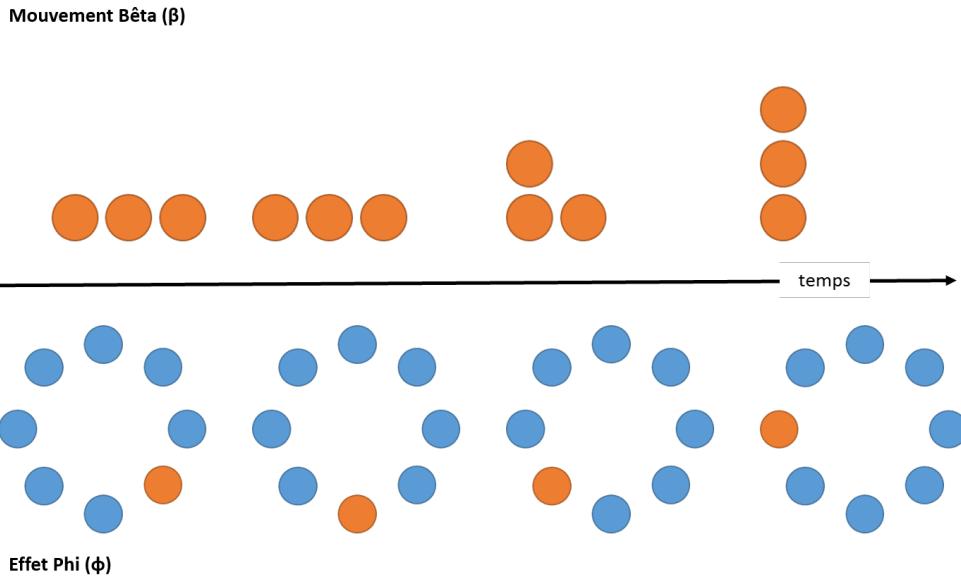


FIGURE 3.3 – Illustration du « mouvement beta » et de l’« effet phi »

suivante (Eq. 3.1) :

$$r = \frac{\text{Amplitude de l'harmonique fondamentale}}{\text{Luminance moyenne}} \quad (3.1)$$

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.

### 3.2.3 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.

### 3.2.4 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. 3.2), dont le tracé est donné en Fig. 3.4 :

$$F_{FPS}(f) = \begin{cases} 0 & f < 16 \\ 126.5 - \frac{367.1}{\sqrt{f-7.6}} & \text{else} \\ 100 & f > 200 \end{cases} \quad (3.2)$$

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.

## 3.3 Couleurs

### 3.3.1 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>

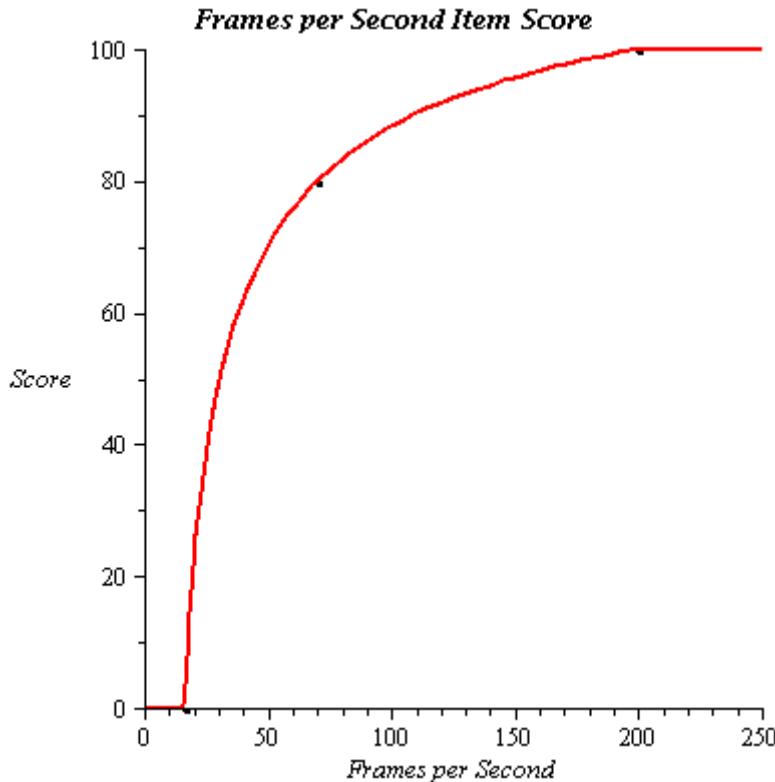


FIGURE 3.4 – Tracé de la fonction de notation du critère « images par seconde »

[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).

### 3.3.2 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. 3.1), on présente un certain nombre de normes avec la proportion d'espace colorimétrique

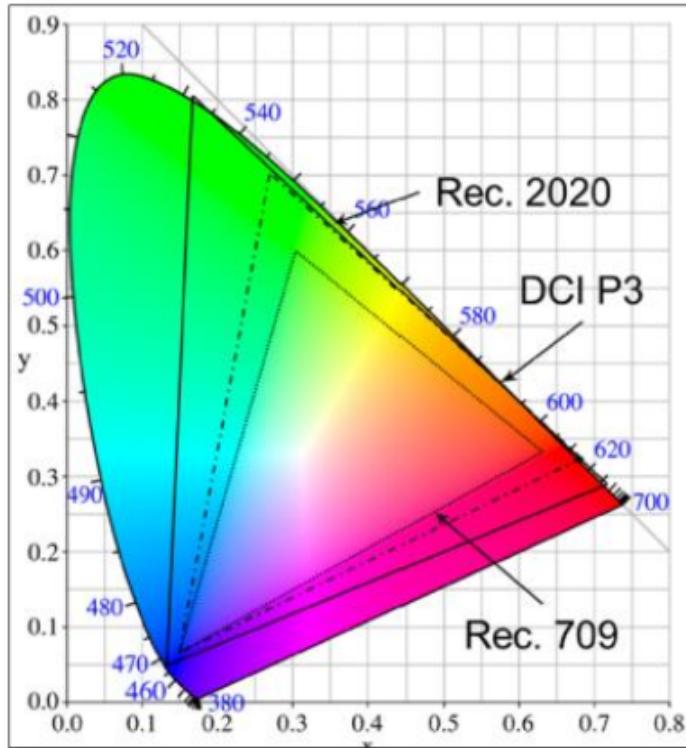


FIGURE 3.5 – Tracé de différents gamuts sur l'espace colorimétrique CIE RGB 1931.

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. 3.5

TABLE 3.1 – 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%

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

Pas mal mais caractérise la lumières émises 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 ?)

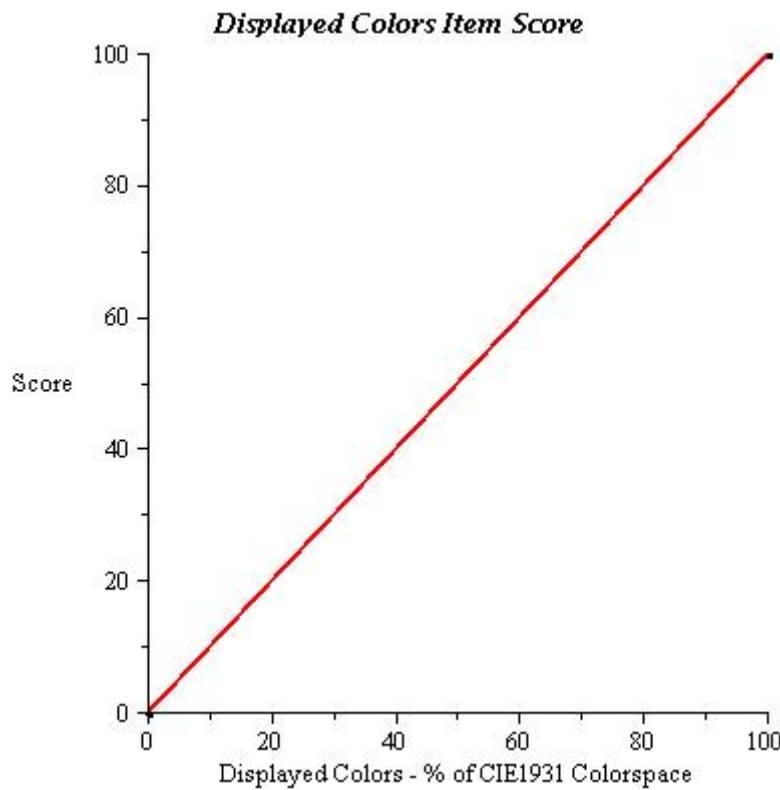


FIGURE 3.6 – Tracé de la fonction de notation du critère « couleurs »

### 3.3.4 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. 3.6). 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 :

$$F_{color}(c) = \begin{cases} c & c \leq 100 \\ 100 & c > 100 \end{cases} \quad (3.3)$$

Il pourrait être intéressant de réaliser une expérimentation sur les besoins en couleurs (à différentier 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.

## 3.4 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

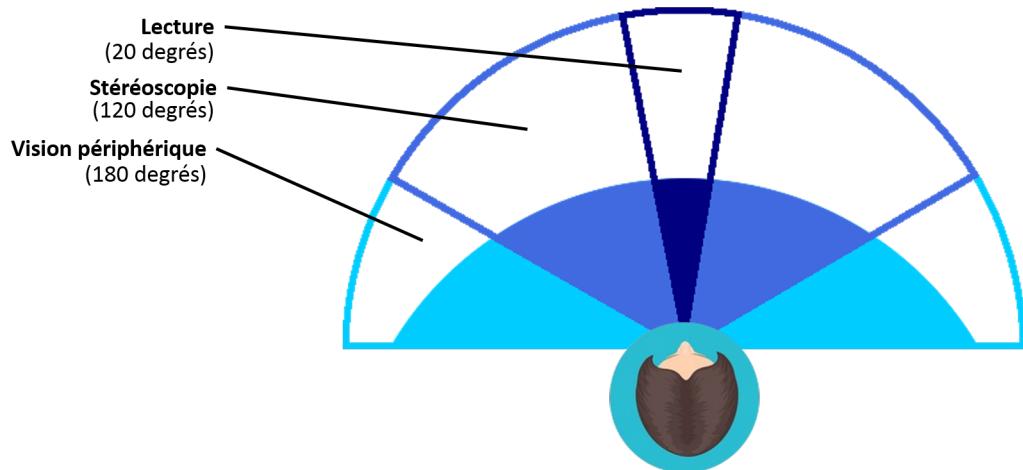


FIGURE 3.7 – Répartition des zones visuelles sur l’axe horizontal du champ de vision.

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).

### 3.4.1 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 fovéale 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. 3.7. 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. 3.4) :

$$F_h(h) = \begin{cases} 0 & h < 20 \\ 19.6\sqrt{h} - 0.5 \cdot h - 78.3 & \text{otherwise} \\ 100 & h > 180 \end{cases} \quad (3.4)$$

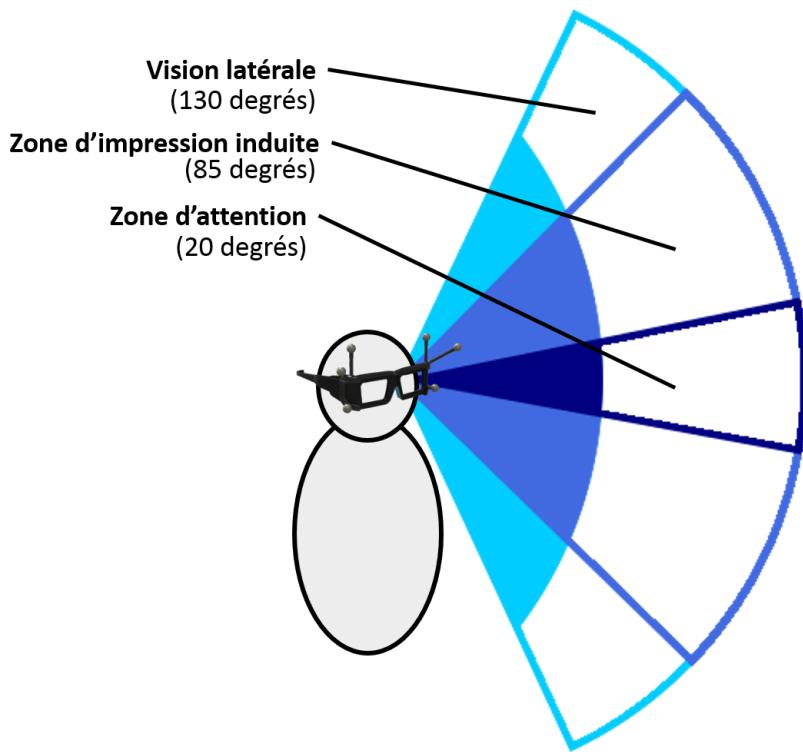


FIGURE 3.8 – Répartition des zones visuelles sur l'axe vertical du champ de vision.

### 3.4.2 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. 3.8). L'équation propre à l'axe vertical est, avec  $v$  la valeur en degrés du champ de vision vertical (V-FOV) (Eq. 3.5) :

$$F_v(v) = \begin{cases} 0 & v < 20 \\ 32.0\sqrt{v} - 1.1 \cdot v - 121.1 & \text{otherwise} \\ 100 & v > 130 \end{cases} \quad (3.5)$$

### 3.4.3 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. 3.6) :

$$\begin{cases} k_h = \frac{180}{180+130} = 0.58 \\ k_v = 1 - k_h = 0.42 \end{cases} \quad (3.6)$$

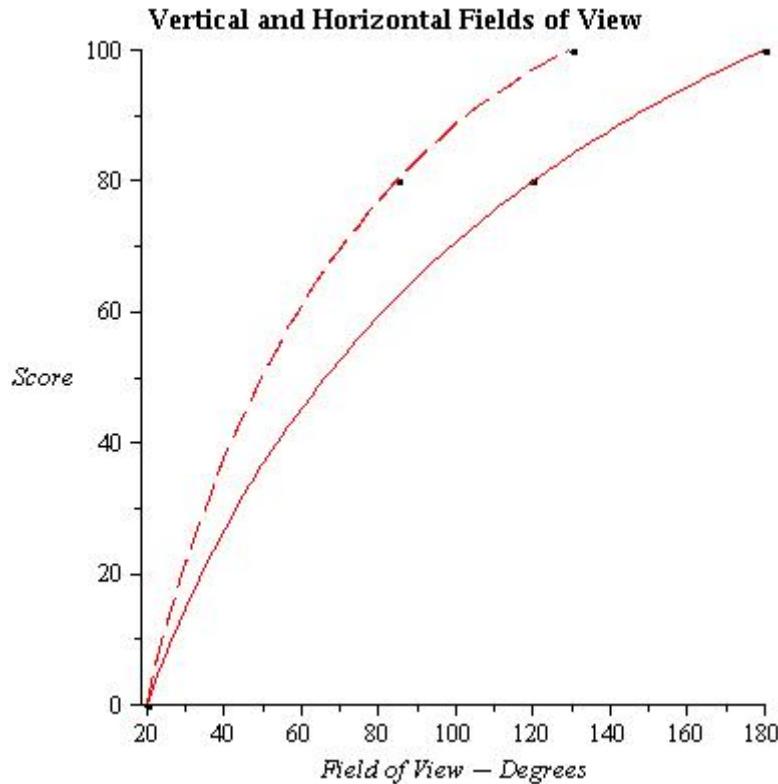


FIGURE 3.9 – Tracé de la sous-fonction horizontale de notation du critère « champ de vision » (ligne continue) et de la sous-fonction verticale (pointillés).

### 3.4.4 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. 3.7) :

$$F_{FOV}(h, v) = k_h \cdot F_h(h) + k_v \cdot F_v(v) \quad (3.7)$$

## 3.5 Acuité monoscopique

### 3.5.1 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	
		1.25	20/16	6.3/5	6/4.8	5/4	4/3	1/0.8	0.8	-0.1	105	
		<b>1.0</b>	<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>	<b>1.0</b>	<b>0</b>	<b>100</b>	
		0.8	20/25	6.3/8	6/7.5	5/6.3	4/5	1/1.25	1.25	+0.1	95	
Low Vision	Mild 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	
		0.5	20/40	6.3/12.5	6/12	5/10	4/8	1/2	2.0	0.3	85	
		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	
Low Vision	Moderate Vision Loss	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	
Low Vision	Severe Vision Loss	<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	
Profound Vision Loss	Profound Vision Loss	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	
(Near-) Blindness	Near-Blindness	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 3.10 – Classification ICD-9-CM

La figure 3.10 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.

### 3.5.2 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 3.2).

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 3.2 – 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.

### 3.5.3 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. 3.8, tracé en Fig. 3.11) est la suivante :

$$F_{mono\_acuity}(\alpha) = \begin{cases} 0 & \alpha > 3.5 \\ 128.9 - 68.8\sqrt{\alpha} - \frac{0.1}{\alpha} & \text{else} \\ 100 & \alpha < \frac{1}{6} \end{cases} \quad (3.8)$$

## 3.6 Acuité stéréoscopique

### 3.6.1 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. 3.9) :

$$\Delta r = 0.001r^2 \quad (3.9)$$

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

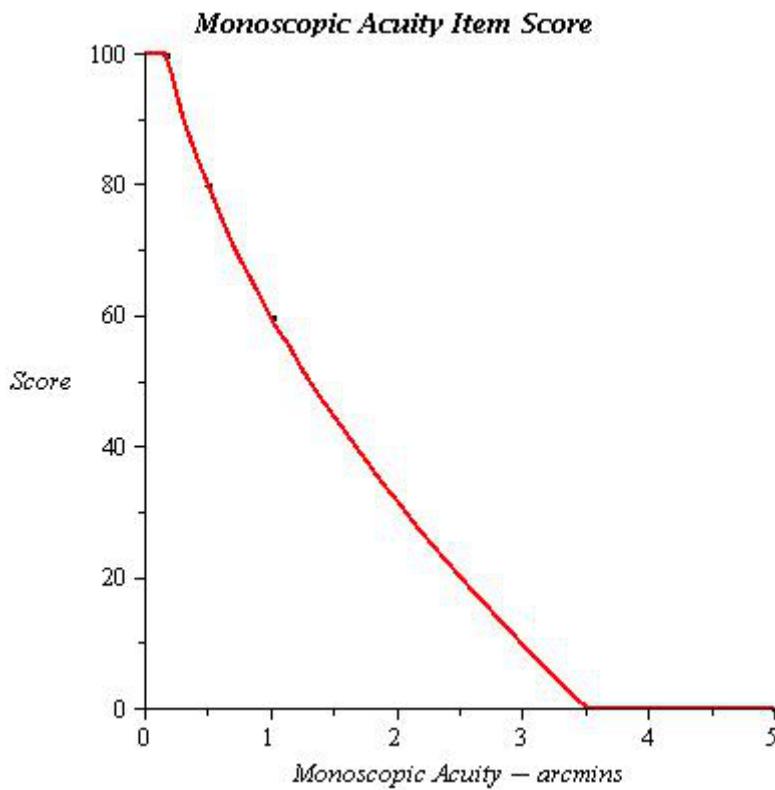


FIGURE 3.11 – Tracé de la fonction de notation du critère « acuité monoscopique »

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. 3.13) 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. 3.12).

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. 3.10) [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.

$$\Delta\nu_{min} = \frac{d_{IPD} * \Delta r}{r^2} \quad (3.10)$$

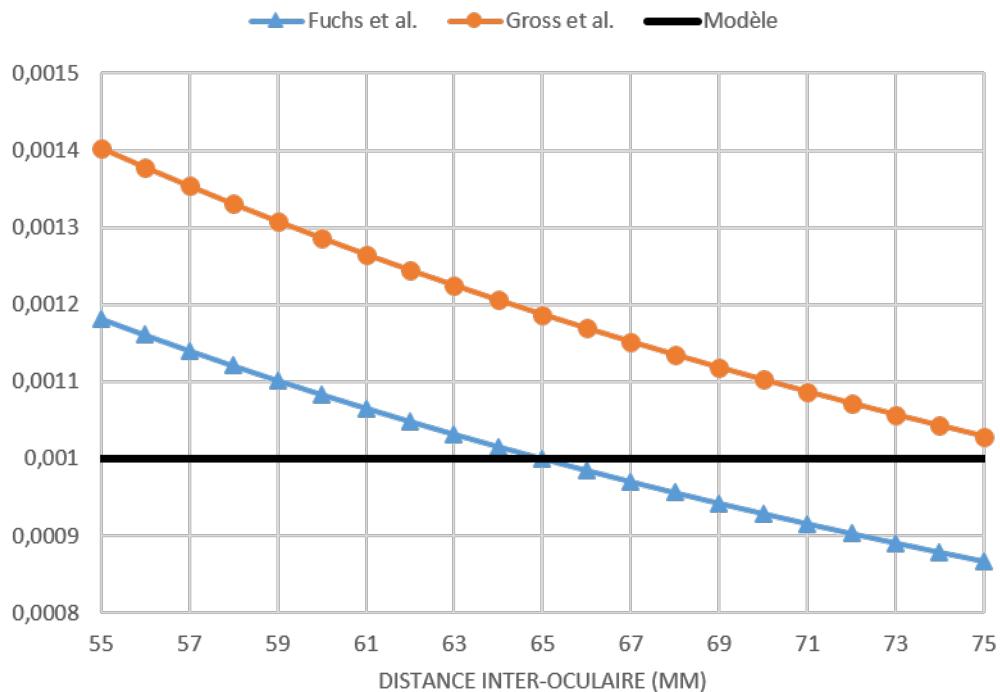


FIGURE 3.12 – Variation de la constante du modèle d’acuité stéréoscopique en fonction de la distance interoculaire.

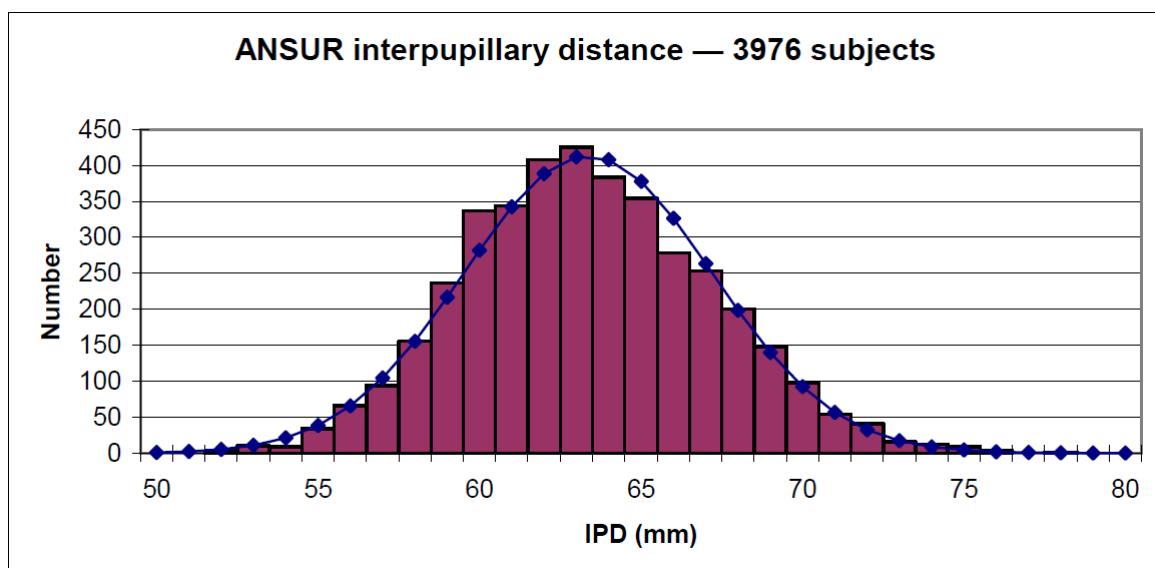


FIGURE 3.13 – Variation de la distance interoculaire sur la population. Image tirée de [Dodgson, 2004]

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.

### 3.6.2 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. 3.11) :

$$F_{stereo\_acuity}(x, r) = \begin{cases} 100 & \alpha < \Delta\nu_{min} \\ 100 * \frac{\Delta\nu_{min}}{\alpha} & \text{else} \end{cases} \quad (3.11)$$

#### 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. 3.14.

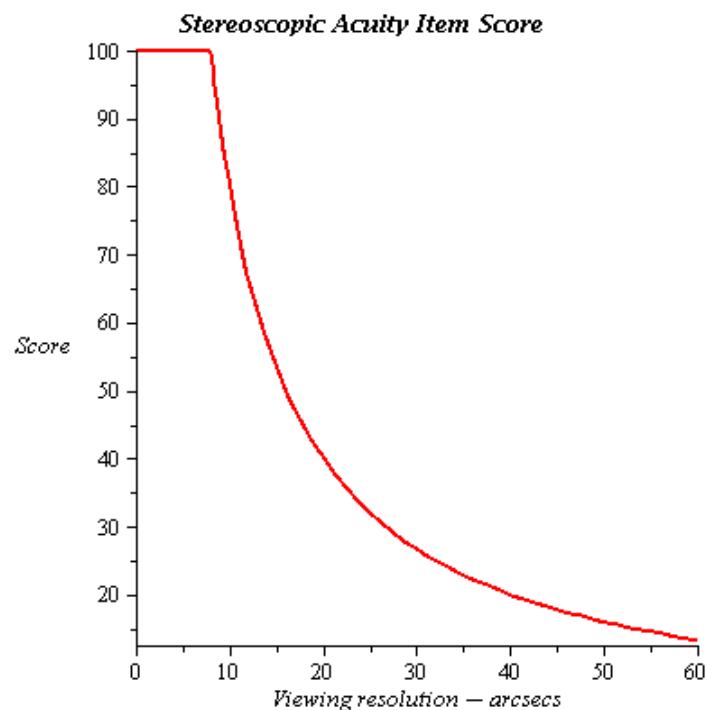


FIGURE 3.14 – Tracé de la fonction de notation du critère « acuité stéréoscopique » ( $\Delta\nu_{min} = 8 \text{ arcsecs}$ )

## CHAPITRE

# 4

## INDICES D'IMMERSION

### 4.1 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-calculation » : le temps écoulé entre un mouvement tracké et de l'ordre de recalcul de l'image à une nouvelle position. Plus simplement, c'est le temps d'acquisition des capteurs de mouvements.
- « Pré-calculation 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.

## 4.2 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.

### 4.2.1 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. 4.1), 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 :

$$\begin{cases} F_{H-FOR}(h) = \frac{100}{h_{max}-h_0} \cdot (h - h_0) \\ F_{V-FOR}(v) = \frac{100}{v_{max}-v_0} \cdot (v - v_0) \end{cases} \quad (4.1)$$

### 4.2.2 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 interpréter. 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. 4.1). 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

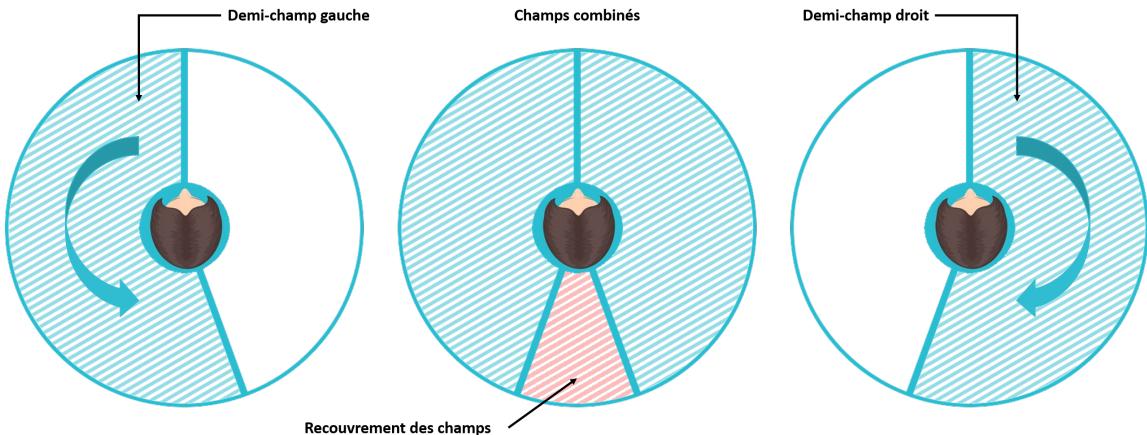


FIGURE 4.1 – Recouvrement des demi-champs de regard (temporal et nasal).

obtient (Eq. 4.2) :

$$\begin{cases} k_h = \frac{400}{400+310} = 0.56 \\ k_v = 1 - k_h = 0.44 \end{cases} \quad (4.2)$$

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. 4.3) :

$$F_{FOR}(h, v) = k_h \cdot F_{H-FOR}(h) + k_v \cdot F_{V-FOR}(v) \quad (4.3)$$

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.

## 4.3 Stéréoscopie

### 4.3.1 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, ont des images d'un point de vue toujours 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 4.2 – 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. 4.2). 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. 4.2). 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.

### 4.3.2 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. 4.3 un exemple à titre d'illustration uniquement.

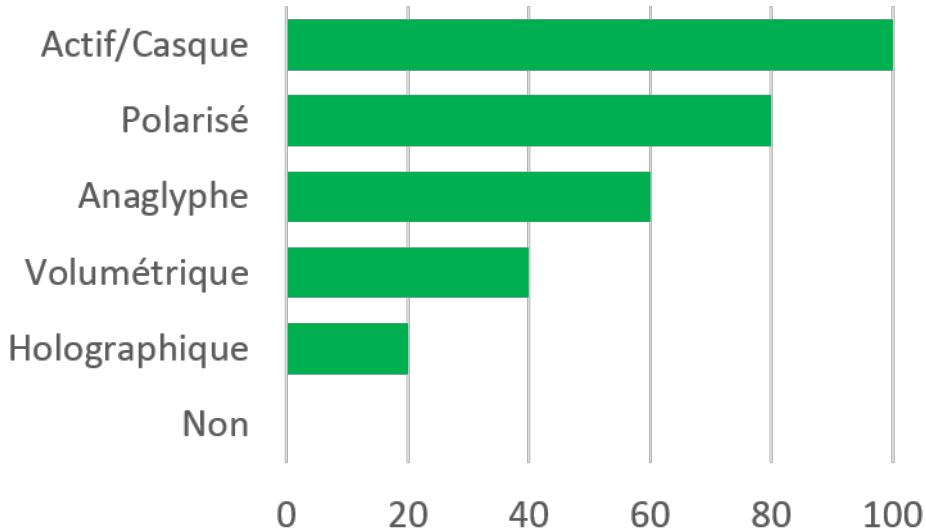


FIGURE 4.3 – Illustration d'une notation du critère « Stéréoscopie » en fonction de la technologie

## 4.4 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. 4.4). 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. 4.4 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 4.4 – « Body » de tracking monté sur des lunettes de stéréoscopie & ensemble caméra-émetteur infrarouge pour le tracking.

## 4.5 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 différenciations 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.

## 4.6 Orientation des caméras

### 4.6.1 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. 4.5), 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. 4.6). 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 4.5 – Exemple de système de tracking de regard portatif.

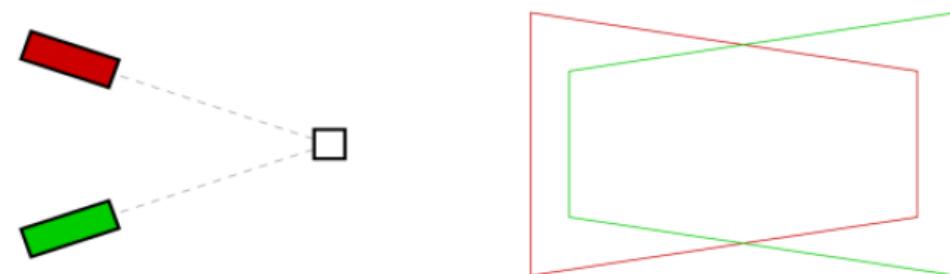
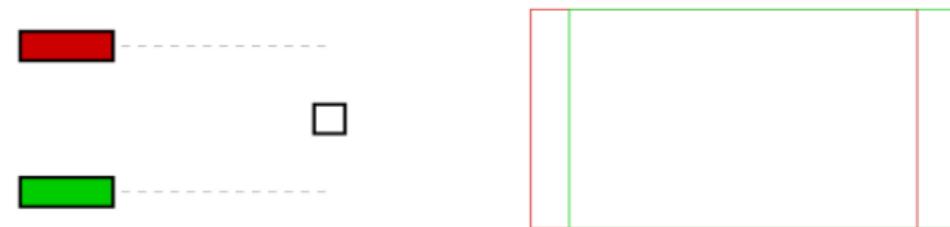


FIGURE 4.6 – Le carré est vu comme deux rectangles dans une configuration « caméras à convergence nulle » et comme deux trapèzes lorsque les caméras sont convergentes. Il faut donc retordre artificiellement les trapèzes pour qu'ils correspondent à la projection qui se fait sur une surface rectangulaire (image tirée de [Aurat, 2016]).

#### 4.6.2 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.



# CHAPITRE

# 5

## PONDÉRATION DU MODÈLE

Il existe deux niveaux de pondération dans le modèle : un premier, local, à l'intérieur de certains critères (champ de vision et champ de regard), et un deuxième, global, qui donne un poids à chacun des critères les uns par rapport aux autres. On a déjà abordé la pondération locale et ses hypothèses, on s'intéresse ici au second niveau de pondération : au global.

La pondération complète du modèle est un travail conséquent qui nécessite un engagement entier. On ne prétend donc pas donner la solution mais simplement des propositions pour une première approche théorique.

### 5.1 Contraste, luminance, taille et champs visuels

Comme on a pu voir dans les chapitres précédents, la construction du modèle se base sur la vision et sa modélisation. Une première piste importante pour la pondération peut être relevée à l'étude du modèle de Rose [Rose, 1948], qui, on le rappelle, est décrit par une simple équation (Eq. 5.1) :

$$BC^2\alpha^2 = \text{constant} \quad (5.1)$$

On voit clairement qu'une importance supérieure est attribuée aux critères de contraste et de taille par rapport au critère de luminance avec des puissances d'ordre 2 par rapport à une puissance d'ordre 1. On peut donc dire que le contraste et la taille de l'objet nécessite une pondération (plus) forte. Si on ne peut pas relier directement la taille de l'objet perçu à l'acuité visuelle, il est important de noter que plus l'acuité est forte, plus l'objet sera

perçu finement. Enfin, par expérience, il nous semble même que le contraste est un critère capital et qu'il devrait recevoir la plus forte des pondérations.

La pondération (locale et globale) des champs de vision et de regard est fortement liée à l'application qui est faite dans le système immersif. On a pu le constater via une expérimentation qui sera décrite dans le chapitre suivant.

TABLE 5.1 – Force relative des critères du modèle non liés à la perception de la profondeur

Paramètre	Force relative
Contraste	Fort
Taille de l'objet	Fort
Champ de vision	Variable
Champ de regard	Variable
Luminance	Faible

## 5.2 Tracking, stéréoscopie, convergence des caméras et couleurs

Une autre piste très intéressante est issue de l'importance relative des indices de profondeur entre eux. On a déjà présenté les différents indices qui permettent au cerveau de reconstruire une carte de profondeur (*cf.* première partie sur l'état de l'Art). On s'intéresse ici à un classement par importance de ces indices.

Il existe quelques études qui ont comparé des indices de profondeurs entre eux afin de déterminer le(s)quel(s) est/sont prépondérant(s) sur les autres. Par exemple, [Mazur & Reising, 1990] compare la stéréoscopie avec la perspective aérienne et la comparaison de taille entre des objets familiers. [Reinhart *et al.*, 1990] ont comparé la taille relative des objets entre eux, la stéréoscopie (disparités binoculaires), l'interposition et la luminosité de l'objet, tandis que [Surdick *et al.*, 1994, 1997] comparent quant à eux 7 indices différents :

- la luminance relative,
- la taille relative,
- la hauteur relative,
- la perspective linéaire,
- l'effet de rapprochement,
- le gradient de texture,
- la stéréoscopie.

Les résultats de Mazur & Reising [1990] et de Surdick *et al.* [1994] montrent une domination des indices de profondeur que sont la stéréoscopie et les perspectives aérienne et

linéaire. Les résultats de Reinhart *et al.* [1990] mettent en avant l'apport de la stéréoscopie par rapport aux indices monoculaires pour la perception de la profondeur de manière plus tranchée. On note néanmoins que la grande majorité des indices mis en concurrence sont des indices graphiques qui sont donc en dehors de notre cadre d'étude.

Parallèlement, [Mikkola *et al.*, 2010] relèvent que la stéréoscopie et la parallaxe de mouvement sont considérées comme les indices les plus importants (résultats à rapprocher de ceux de [Rogers & Graham, 1979] qui montrent que la parallaxe de mouvement est un indice de profondeur pertinent même lorsque tous les autres indices sont inexistant). Ils ont de leur côté conduit des expérimentations dans lesquelles ils demandaient à leurs sujets de noter de manière subjective l'impression de profondeur ressentie : la stéréoscopie arrive encore une fois avec de bon résultats tandis que des indices monoculaires tels que les ombres ou les textures des objets semblent avoir un impact plus réduit. Ces

Enfin, [Mehrabi *et al.*, 2013] ont réalisé d'excellentes classifications sur les indices de profondeur, leurs atouts et leurs limitations (on trouvera une adaptation en Table 5.2) ainsi qu'une table sur les différentes techniques d'affichage en spécifiant notamment les indices de profondeurs disponibles dans chacun des cas. La classification est divisée en deux parties avec des indices ayant une contribution soit fort soit faible mais permet déjà une bonne approximation pour un idée générale de pondération.

TABLE 5.2 – Force relative des indices de perception de la profondeur. Table adaptée depuis [Mehrabi *et al.*, 2013]

<b>Indice de profondeur</b>	<b>Indice binoculaire</b>	<b>Force relative</b>	<b>Portée</b>
Parallaxe binoculaire	Oui	Fort	2,5 m – 20 m
Parallaxe monoculaire	Non	Fort	0 – ∞
Taille de l'image rétinienne	Non	Fort	0 – ∞
Perspective linéaire	Non	Fort	0 – ∞
Gradient de texture	Non	Fort	0 – ∞
Recouvrement	Non	Fort	0 – ∞
Accommodation	Non	Faible	0 – 2 m
Convergence	Oui	Faible	0 – 10 m
Perspective aérienne	Non	Faible	<i>Long. dist. seulement</i>
Ombrage	Non	Faible	0 – ∞
Couleurs	Non	Faible	0 – ∞

### 5.3 Conclusion

Il reste néanmoins des critères pour lesquels la littérature ne donne pas immédiatement un a priori de pondération tels que le nombre d'images par seconde, la latence ou l'uniformité entre les écrans. par défaut on leur attribue une importance relative plus faible que les facteurs « forts » qui semblent très prépondérants sur les autres.

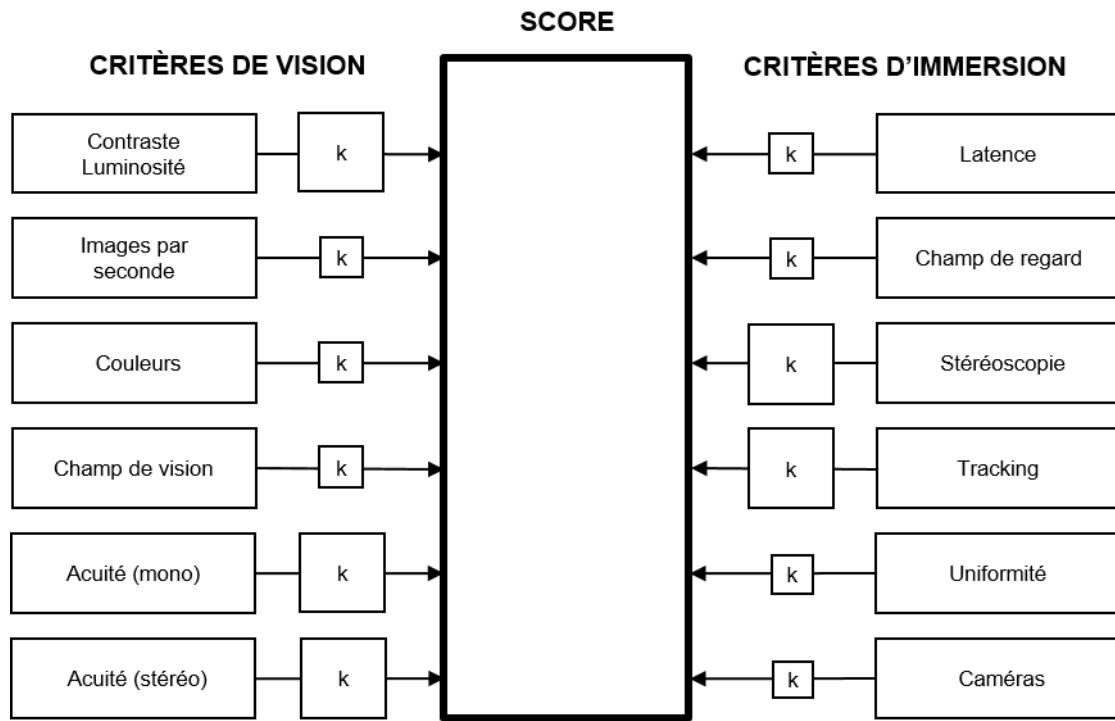


FIGURE 5.1 – Modélisation avec ordre de grandeur de la pondération.

On a maintenant une meilleure idée générale des tendances qu'il peut exister pour une pondération du modèle dans le cas général. Dans des cas spécifiques (*cf.* chapitre suivant) comme la conduite, la pondération peut fortement varier en fonction des besoins de l'utilisateur. On présente ces ordres de grandeur en Fig. 5.1 en modélisant les pondérations « fortes » avec une boîte plus grande que les pondérations « faibles ».

Il convient de rappeler que cette proposition est une ébauche et que la définition précise d'une (ou de plusieurs) pondération(s) reste un sujet à part entière que nous n'avons pas à proprement parler traité.

# CHAPITRE

# 6

# ETUDE EXPÉRIMENTALE

## 6.1 Objectif

On a réalisé une expérimentation pour comparer les notes données par notre modèle de score et les notes données par des utilisateurs d'un système d'affichage immersif aux différentes caractéristiques que l'on retrouve dans le modèle : fluidité, couleurs, contraste, ... Cette expérimentation n'a pas pour but -et ne peut pas- valider notre modèle : elle compare deux types de données très différentes, les unes objectives et les autres subjectives. C'est seulement une comparaison entre un modèle théorique qui donne un score qui reflète à quel point la partie hardware du système donne un signal proche de celui défini dans les modèles de vision, et un score donné subjectivement par les sujets et qui reflète à quel point les utilisateurs ont apprécié les différentes capacités du système.

Cependant, cette expérimentation nous permet de montrer la différence qu'il peut exister entre l'appréciation d'un critère noté subjectivement et sa note de performance théorique. Cela nous permet aussi de récupérer des indices sur la manière dont les utilisateurs se comportent dans un système immersif et dans le cas spécifique de cette simulation ; ces indices nous aideront ensuite dans certaines de nos hypothèses.

## 6.2 Apparatus

32 sujets ont été réunis pour cette expérimentation : 23 hommes pour 9 femmes, âgés de 20 à 27 ans, avec une moyenne d'âge de 25 ans (écart-type : 1,8 ans). Elle a été faite dans une CAVE 4 faces dont les dimensions sont présentées en Table 6.1. La taille des pixels était de 2,25 mm et les sujets étaient assis de telle manière à avoir les yeux à 2 m de l'écran de face. L'expérimentation était écologique : les sujets devaient simplement conduire dans

INSUFFICIENT		GOOD		MAX
1	2	3	4	5
			X	

FIGURE 6.1 – Echelle « aidée » pour noter subjectivement de 1 à 5 les différents critères du système

le simulateur. La conduite était faite au moyen d'un ensemble Logitech G25 avec volant et pédales, le tout connecté au logiciel de simulation SCANeR Studio<sup>1</sup>. Les sujets étaient assis dans un vrai siège de voiture posé au sol. L'environnement virtuel de simulation était un simple paysage en boucle fermée mêlant à la fois mer, bord de lac, ville et montagne.

TABLE 6.1 – Dimensions du CAVE « P3I » dans lequel a eu lieu l'étude expérimentale sur la comparaison entre les notes objectives et des notes subjectives.

Face	Taille	Résolution
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

Chaque sujet devait faire la même chose : conduire pendant 8 minutes sans autre tâche particulière que de regarder tout autour de soi. La vitesse de la voiture était limitée à 30km/h, et ce, pour plusieurs raisons : rester dans le même cas d'usage en basse vitesse pour tout le monde et ne pas rendre malade les sujets. Après les 8 minutes, les sujets devaient noter via une échelle « aidée » de 1 à 5 (*cf. Fig. 6.1*) les différents critères de notre modèle, par comparaison à ce qu'ils pourraient voir dans « la vraie vie ».

Une comparaison a ensuite été faite entre les moyennes des notes subjectives des sujets et les valeurs théoriques objectives du modèle de score. Aucune pondération n'était appliquée sur les critères (mis à part les pondérations internes aux critères de champ de vision et de champ de regard).

Pour représenter la variation de taille, et donc de hauteur des yeux par rapport au sol en position assise, entre les différents sujets les valeurs théoriques de champ de vision et de champ de regard sont calculées deux fois avec 1,10 m et 1,15 m de hauteur d'yeux (par rapport au sol, en position assise sur le siège de conduite). On moyenne ensuite ces deux valeurs.

On a sélectionné seulement des sujets jeunes (de 20 à 27 ans) et dont la vision était parfaite (ou corrigée pour être parfaite). Cela a été fait afin de rester sur des sujets dont les capacités de vision (vision du contraste, de la luminance, acuité, ...) n'ont pas commencé à décroître en fonction de l'âge, processus qui commence autour de 30 ans.

1. <http://www.oktal.fr/en/automotive/range-of-simulators/software> (lien raccourci : <https://lc.cx/NxyR>)

## 6.3 Résultats

Afin d'analyser les résultats, la population de sujets a été divisée en 8 sous-catégories. Les caractéristiques de ces différentes sous-catégories peuvent être trouvées dans la Table 6.2. Aucun résultat statistique ne peut être tiré de la catégorie « femme + gamer » puisque qu'elle ne comporte qu'un seul membre.

TABLE 6.2 – Population et âge des sous-catégories de l'étude expérimentale sur la comparaison entre les notes objectives et des notes subjectives.

Sous-catégorie	Population	Âge moyen	Écart-Type
Total	32	24.8 ans	1.8
Hommes (M)	23	24.8 ans	2
Femmes (W)	9	24.8 ans	1.7
Gamers (G)	13	24.5 ans	2.3
Non gamers (NG)	19	25 ans	1.5
M+G	12	24.7 ans	2.3
M+NG	11	24.9 ans	1.6
W+G	1	22 ans	0
W+NG	8	25.1 ans	1.5

Cinq valeurs théoriques du modèle étaient assez mûres pour être calculées et être comparées aux notes données par les sujets :

- le nombre d'images par seconde
- les couleurs
- l'acuité monoscopique
- le champ de vision
- le champ de regard

Les autres critères n'étaient soit pas encore notable via le modèle de score (luminance, contraste, latence et uniformité) soit difficile à noter par un simple questionnaire sans procéder pendant l'expérimentation à une tâche particulière mettant en avant la caractéristique (acuité stéréoscopique, tracking et stéréoscopie).

Les valeurs des moyennes et des écarts types des cinq critères retenus pour cette étude expérimentale sont données en Fig. 6.2.

## CHAPITRE 6. ETUDE EXPÉRIMENTALE

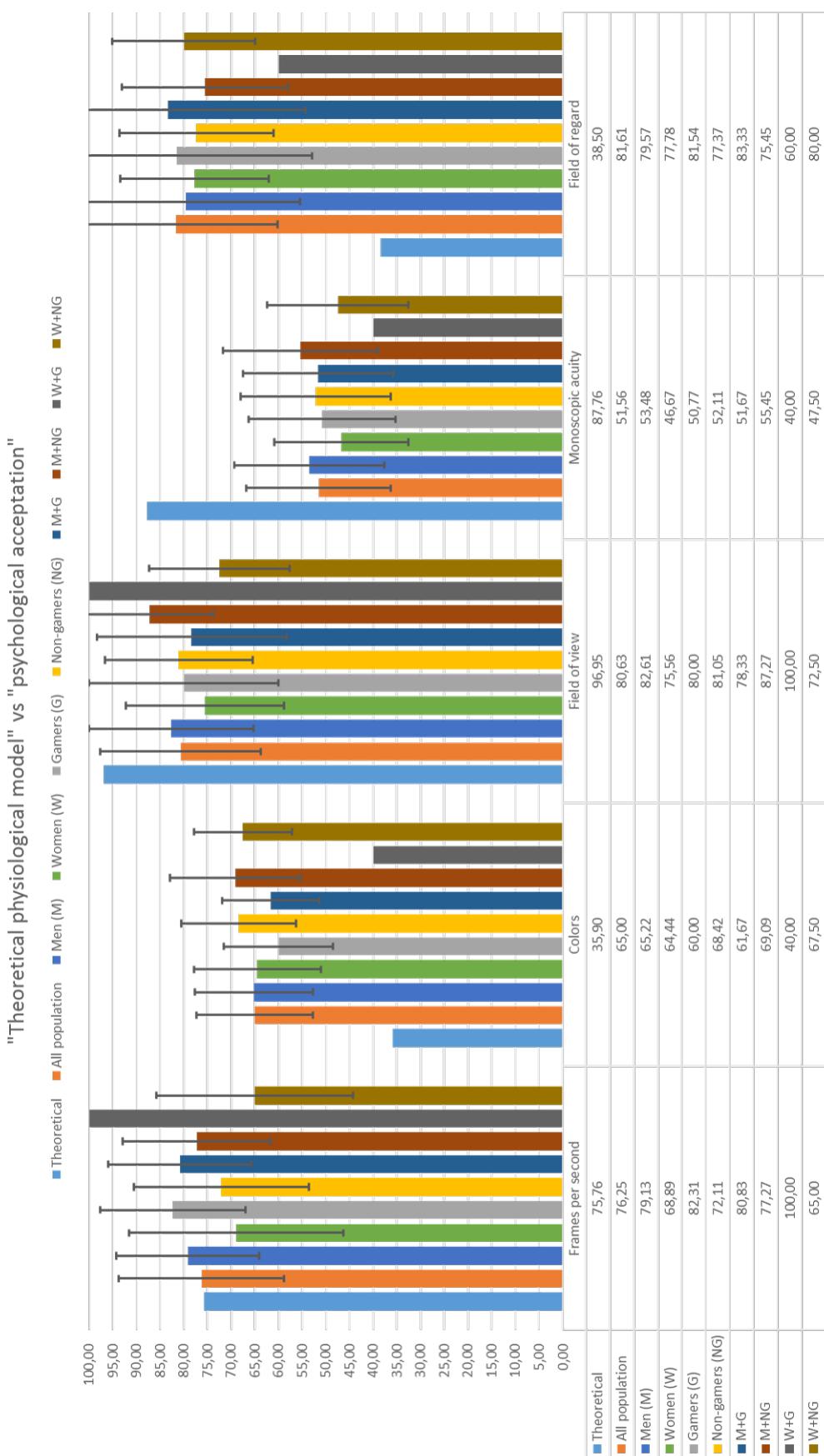


FIGURE 6.2 – Résultats des valeurs théoriques calculées et des valeurs subjectives relevées.

## 6.4 Discussion

Les résultats de la comparaison entre le modèle théorique et l'acceptation psychologique par les sujets peuvent être divisés en trois catégories. Premièrement les valeurs relevées sur les questionnaires correspondent plutôt bien aux valeurs données par le score théorique. C'est le cas par exemple pour le critère « images par seconde ».

Deuxièmement, et en opposition à la première catégorie de résultats, deux critères diffèrent fortement par rapport aux valeurs d'acceptation par les sujets : dans un sens (pour les « couleurs » les valeurs subjectives sont supérieures aux valeurs objectives) comme dans l'autre (pour l'« acuité monoscopique » les valeurs théoriques sont supérieures aux valeurs d'acceptation).

Enfin, la troisième catégorie contient des critères qui ont des valeurs objectives très différentes (96,95 pour le « Champ de Vision » et un pauvre 38,50 pour le « Champ de Regard ») mais qui ont pourtant des valeurs subjectives très proches l'une de l'autre.

### 6.4.1 Couleurs et acuité monoscopique

Ces deux critères montrent une grande différence de notation entre leur score théorique et leur score subjectif même si, à l'intérieur des sous catégories, les moyennes et les écarts-type sont assez proches.

D'un côté, les sujets étaient globalement satisfaits de la quantité de couleurs bien que ce soit en fait assez peu par rapport à toutes les couleurs visibles réellement. Une raison possible à cette divergence est l'absence de référence au monde réel pour juger par comparaison et l'utilisation massive d'appareils utilisant le même espace limité de couleurs (smartphones, télévisions, cinéma, ...).

De l'autre côté, l'acuité monoscopique n'a pas été bien notée alors qu'elle aurait du paraître suffisante. On peut l'expliquer aussi par l'habitude d'avoir une très forte demande sur la résolution des écrans, que ce soit pour les téléphones, les ordinateurs ou les écrans divers.

### 6.4.2 Champ de vision & champ de regard

Bien que légèrement plus élevée, la valeur théorique du champ de vision est assez proche de sa valeur d'acceptation associée et est, en plus, contenue dans la majorité des écarts-type des sous catégories. Cependant, la valeur théorique du champ de regard est extrêmement basse comparée à sa valeur d'acceptation et éloignée de tous les écarts-type des sous-populations.

La valeur théorique basse vient du fait que, même si l'angle de vision horizontal est très bon, il n'y a pas de face supérieure ce qui joue très négativement dans la notation de l'axe vertical du critère. A cause de la pondération quasi équivalente pour les deux axes, la note globale du critère est tirée vers le bas par la partie verticale qui est très basse.

## CHAPITRE 6. ETUDE EXPÉRIMENTALE

---

La tâche de conduite nécessite la quasi totalité de l'attention du conducteur à l'avant de la voiture et implique majoritairement l'axe horizontal de la vision. C'est pourquoi les sujets n'ont pas eu besoin d'utiliser particulièrement d'autres écrans que celui face à eux ni besoin d'images qui seraient hors-écrans. Cela peut expliquer pourquoi les participants ont noté le champ de regard si proche du champ de vision, car ils ont rarement dépassé les bornes du champ de vision et, quand c'était le cas, seulement un petit peu. Cela implique, dans ce cas précis, que notre hypothèse de pondération des axes en fonction de leur taille respective n'est pas correcte. Un jeu de pondération spécifique à la conduite devrait être calculé et appliqué, limitant considérablement l'influence de l'axe vertical.

Enfin, on remarque une tendance intéressante à mettre en avant (bien que non soutenue par une preuve statistique). Les femmes (joueuses ou non) semblent être plus sévère dans leur notation que les hommes, et ce, quelle que soit la note théorique associée.

# CONCLUSION

La première approche de la thèse, développer un "modèle de vision" traduit pour la Réalité Virtuelle, s'est vite révélée être une impasse. On choisit alors une approche différente : plutôt que de "faire" du réalisme, il pourrait être intéressant de "quantifier" le réalisme. A notre connaissance, il existe peu voire très peu de modèles ayant pour but de quantifier objectivement le réalisme physiologique dans la littérature.

On propose alors une évaluation de la performance d'un système via un score basé sur le système visuel humain qui dépeindrait, pour un système immersif donné et pour une situation donnée, à quel point le simulateur est efficace pour transmettre le bon niveau d'informations visuelles et, étant dans un système de Réalité Virtuelle, le bon niveau d'indices d'immersion. Ce score est en fait la réunion de douze critères chacun lié à une caractéristique de la vision et/ou de l'immersion visuelle (c'est à dire les critères immersifs dérivés directement du système visuel ou du système en lui-même). Chaque critère est jugé indépendamment puis contribue à la note globale via une pondération. La notation en elle même se fait, principalement, basée sur la littérature. Si on a pu appliquer cette méthode pour la majorité des cas, certains critères ne se prêtent pas à un tel traitement et ont bénéficié d'une notation particulière.

Les critères sont répertoriés en deux sections :

- **Indices de vision** : contraste et luminosité (luminance), images par seconde, nombre de couleurs affichables, champ de vision, acuités monoscopique et stéréoscopique.
- **Indices d'immersion** : latence, champ de regard, stéréoscopie, tracking, uniformité et convergence des caméras.

Après une étude théorique fouillée pour déterminer la plus grande majorité des critères, on a réalisé une expérimentation pour comparer les notes données par notre modèle de

## CONCLUSION

---

score et les notes données par des utilisateurs. Cette expérimentation n'avait pas pour but -et ne peut pas- valider notre modèle mais elle nous a permis de montrer la différence qu'il peut exister entre l'appréciation d'un critère noté subjectivement et sa note de performance théorique.

Néanmoins, il reste des points à développer. Certains critères ne sont pas aboutis et doivent être complétés par la pratique. C'est le cas par exemple du contraste et de la luminance, ainsi que de la latence. Le traitement expérimental de ces critères sera abordé dans les deux prochaines parties. Enfin, même si on a pu rapidement aborder la pondération intra-critère (champ de vision et champ de regard) ainsi que la légitimité des hypothèses sur lesquelles elle est construite (étude expérimentale), il restera encore tout un travail autour de la pondération inter-critères. Le sujet est extrêmement vaste et complexe car il dépend fortement de ce que l'on fait dans le système immersif.

Nous allons maintenant présenter le travail d'expérimentation qui a été réalisé autour de la validation du modèle de Rea en Réalité Virtuelle, pour le critère « contraste - luminance ».

## **Quatrième partie**

# **Partie expérimentale : contraste et luminance**



# INTRODUCTION

Dans le chapitre précédent, on a décrit deux approches possibles pour le critère de contraste et de luminance : les fonctions de sensibilité au contraste et la performance visuelle relative. Si les premières sont assez limitées, notamment car elles ne déterminent qu'un seuil à partir duquel un élément devient visible et non pas sa visibilité en fonction des conditions d'illumination, le concept de performance visuelle est intéressant. On en a présenté les deux principales modélisations : le modèle de Blackwell et de la CIE [Blackwell & Blackwell, 1971], puissant mais hermétique, et le modèle de Rea et ses évolutions [Rea, 1986]. On a également pointé le fait que ces modèles, bien qu'intéressants, n'étaient pas à l'origine conçus pour la réalité virtuelle et nécessitaient donc une vérification expérimentale.

Il a été nécessaire de choisir entre les deux modèles, celui de Rea et celui de la CIE/Blackwell. Le modèle de Blackwell semble plus complet avec une portée d'action comprise en 1 et  $10000\text{ cd/m}^2$  et un calcul basé sur trois processus de vision, décrits comme principaux, impliqués dans la reconnaissance des détails de la tache à effectuer, dans le maintien des yeux en position fixe pendant les périodes de fixation, et enfin, dans la réalisation de mouvements rapides des yeux (les saccades). Néanmoins, le modèle de la CIE est une boîte noire générée à partir de la mise en commun des travaux d'un certain nombre de chercheurs. Il n'existe pas d'expérimentation détaillée qui puisse être refaite, et à fortiori, encore moins en réalité virtuelle.

De l'autre côté, la modélisation de Rea est plus limitée en portée (entre 12 et  $169\text{ cd/m}^2$ ) mais présente en détail tout le protocole qui a été mis en œuvre pour développer les équations de performance visuelle. De plus, l'intervalle de fonctionnement du modèle correspond relativement bien aux luminances atteignables dans un simulateur et n'est donc pas très contraignant.

L'objectif de cette partie est donc de vérifier par l'expérimentation, dans un simulateur, les prédictions de performance du modèle de Rea. Pour ce faire, on a transposé en réalité virtuelle l'une des expérimentations mise en place par Rea et Ouellette pour déterminer leur modèle.

---

## INTRODUCTION

---

## CHAPITRE

# 1

## MODÈLES DE REA

### 1.1 Première modélisation de la performance visuelle relative

Les sections suivantes décrivent brièvement le contexte et les premiers travaux de Rea qui l'ont mené à proposer sa propre modélisation de la performance visuelle [Rea, 1987, 1986], définie comme étant la vitesse et la précision atteintes pendant la réalisation d'une tache visuelle.

#### 1.1.1 Modèle précurseur

Le premier modèle de Rea est directement inspiré de l'effet de compression proposé par [Naka & Rushton, 1966] qui modélise qu'à partir d'un certain niveau d'intensité, lorsque l'intensité du stimulus augmente la réponse sensorielle associée augmente de moins en moins jusqu'à atteindre une forme de plateau (Fig. 1.1) Les différentes courbes représentent les longueurs d'onde des lumières de couleur utilisées. Cet effet est modélisé de la manière suivante :  $\frac{R}{R_{max}} = \frac{I^n}{I^n + k^n}$ .

Dans cette équation, le rapport de la réponse sensorielle ( $R$ ) sur la réponse sensorielle maximale ( $R_{max}$ ) est égal au rapport de l'intensité  $I$  montée à une puissance  $n$  déterminée sur la somme de cette même intensité  $I$  montée à la puissance  $n$  et d'une intensité  $k$ , également montée à la puissance  $n$ , telle que la réponse à l'intensité  $k$  est égale à la moitié de la réponse sensorielle maximale :  $R(k) = \frac{R_{max}}{2}$ .

On peut faire l'analogie suivante pour donner un bon exemple de l'effet de compression : si dans une pièce on allume une deuxième ampoule d'égale intensité, la sensation de lumi-

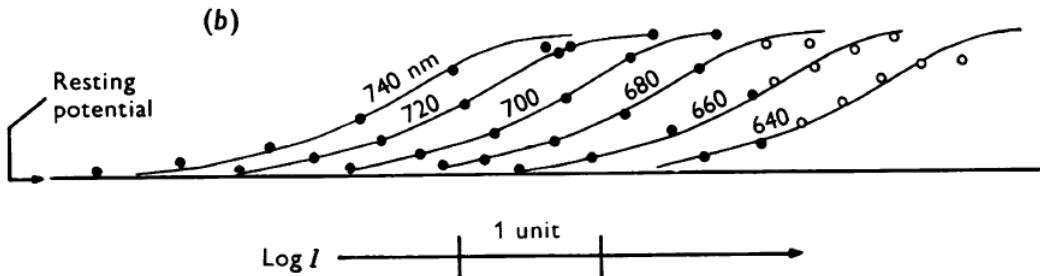


FIGURE 1.1 – Illustration de l’effet de compression. Image tirée de [Naka & Rushton, 1966]

nosité va grandement augmenter. Au contraire, si on rajoute une ampoule dans une pièce dans laquelle déjà mille ampoules sont allumées, la différence de luminosité perçue sera minime.

### 1.1.2 Application à la performance visuelle

Rea s'est imposé deux règles pour la conception de son modèle : la performance visuelle doit être issue d'une performance à la réalisation d'un tache et le modèle doit être cohérent avec la littérature, notamment avec l'effet de compression que l'on vient de décrire.

Le stimulus est décrit (Eq. 1.3) comme la différence entre le seuil de contraste  $C_t$ , calculé avec la formule de Blackwell (Eq. 1.1), qui représente le contraste minimal à partir duquel la perception devient possible en fonction de la luminosité du fond ( $L_B$ ) et le contraste visuel  $C_V$  qui correspond au contraste entre la luminosité du fond et la luminosité de la tache visuelle (Eq. 1.1).

$$C_t = 0.048 \left[ \left( \frac{0.308}{L_B} \right)^{0.4} + 1.0 \right]^{2.5} \quad (1.1)$$

$$C_V = \frac{L_B - L_T}{L_B} \quad (1.2)$$

$$\Delta C = C_V - C_t \quad (1.3)$$

Les calculs suivants sont ensuite tirés de régressions polynomiales de degré 2 faites avec leurs données expérimentales :

$$\begin{cases} n = 0.882 + 4.38\theta_1 - 6.05\theta_1^2 \\ \theta_1 = \log(\log(L_B)) \end{cases} \quad (1.4)$$

$$\begin{cases} k = -2.25 + 1.77\theta_2 - 0.217\theta_2^2 \\ \theta_2 = \log(L_B) \end{cases} \quad (1.5)$$

$$VP_{max} = 0.0628 + 0.0120\theta_2 - 0.00268\theta_2^2 \quad (1.6)$$

Au final, la performance visuelle s'écrit de la manière suivante (Eq. 1.7) :

$$VP = \frac{(\Delta C)^n}{(\Delta C)^n + (k/L_B)^n} \quad (1.7)$$

Et  $RVP = \frac{VP}{f}$  avec  $f$  la valeur de  $VP_{max}$  dans les meilleures conditions (luminance et contraste maximaux). Dans le cas de l'expérimentation de Rea,  $f = 0.076$ .

Les résultats de cette expérimentation permettent à l'auteur de dégager trois tendances :

- A contraste constant, la performance augmente avec la luminance,
- La performance augmente plus rapidement avec le contraste lorsque les conditions de luminance sont plus élevées,
- La performance varie assez peu entre des conditions moyenne et supérieure de contraste.

## 1.2 Méthode des temps de réaction

Rea et Ouellette complètent la démarche initiale de Rea en proposant une méthode pour établir la performance visuelle d'un sujet, basée sur la mesure et la prédiction des temps de réaction de ce dernier à l'apparition d'un stimulus visuel calibré [Rea & Ouellette, 1988, 1991]. C'est cette modélisation que l'on va chercher à traduire en réalité virtuelle afin de déterminer si elle est utilisable dans le cadre de notre score de réalisme, pour le critère de contraste et de luminance.

### 1.2.1 Protocole expérimental

L'objectif était de mesurer le temps de réaction des sujets à l'apparition d'une cible sur un écran. La couleur de la cible était calibrée pour obtenir un contraste choisi par rapport au fond sur lequel elle était affichée (l'écran, en l'occurrence).

L'expérimentation était découpée en deux parties : une première série de mesures pour des cibles plus foncées que le fond sur lequel elles étaient présentées (méthode décrémentale) puis une autre série de mesures avec des cibles plus claires que le fond de présentation (méthode incrémentale).

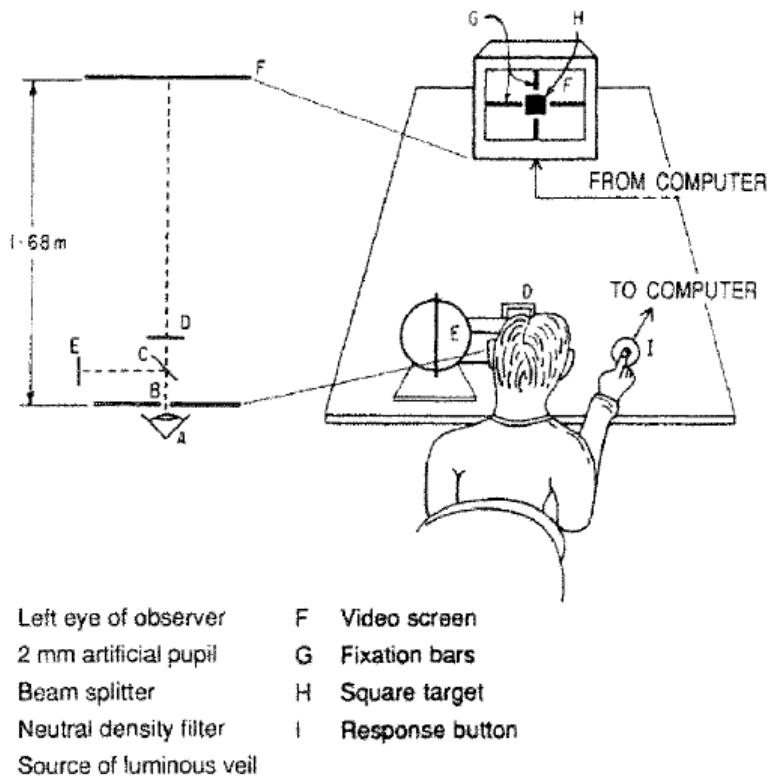


FIGURE 1.2 – Installation de l’expérimentation de Rea et Ouellette. Image tirée de [Rea & Ouellette, 1988]

Les cibles, un carré, étaient présentées à 1.68 m de l’œil du sujet sur un écran occupant 12 degrs de champ de vision horizontal et 7 degr de champ de vision vertical. Toutes les cibles étaient vues avec l’œil gauche, à travers un filtre neutre réglable et avec l’ajout d’une voile lumineux artificiel directement au niveau de l’œil (Fig. 1.2).

Le contraste du carré était calculé en utilisant l’équation suivante (Eq. 1.8) :

$$C = \frac{|(TL_b + L_v) - (TL_t + L_v)|}{TL_b + L_v} = \frac{T|L_b - L_t|}{L_a} \quad (1.8)$$

Avec  $T$  la valeur de transmittance du filtre (entre 0 et 1),  $L_b$  la luminance du fond de l’écran,  $L_t$  la luminance du carré à détecter et  $L_v$  la luminance de voile ajoutée artificiellement.  $L_a$  représente quand à elle la valeur de la luminance d’adaptation, c’est à dire la valeur pour laquelle l’œil et tout le système optique se sont réglés (avec par exemple l’adaptation du diamètre pupillaire).

Chaque apparition de cible était espacée d’une temporisation de 1.5 seconde puis d’une temporisation aléatoire variant entre 1 et 3 secondes. La taille de la tache visuelle était variable entre 0.20 et  $280 \times 10^{-5}$  steradians.

Chaque sujet a enregistré 19200 mesures de temps de réaction pour la partie décrémentale et 3625 mesures pour la partie incrémentale.

### 1.2.2 Calcul des temps de réaction théoriques

Une fois toutes les mesures effectuées, cela a permis de rétablir une équation de performance avec une protocole similaire à la première modélisation et décrit plus haut. Le détail de ces calculs est présenté en annexes et on retiendra ici seulement l'équation suivante (Eq. 1.9) :

$$R = \frac{(\Delta C_d)^{0.97}}{(\Delta C_d)^{0.97} + K^{0.97}} \quad (1.9)$$

De cette équation qui décrit la performance du sujet à détecter l'apparition de la tache visuelle en fonction des conditions d'illumination et de contraste, on dérive le temps de réaction en prenant simplement l'inverse de la performance (Eq. 1.10) :

$$RT = \frac{1}{R} \quad (1.10)$$

Cela permet d'avoir un comportement logique dans les résultats avec un temps de réaction qui diminue quand la performance augmente et inversement.

De plus, cette performance  $R$  liée au temps de réaction, ainsi calculée, peut être reliée au modèle initial de performance visuelle relative. Cela nécessite d'appliquer deux opérations sur l'ensemble des résultats mesurés dans les deux expérimentation pour les même conditions d'illumination, de contraste et de taille de cible : le calcul de  $\Delta T_{vis}$  : la variation du temps de réaction par rapport au temps obtenu dans les meilleures conditions expérimentales (Eq. 1.11) et vient ensuite une transformation linéaire (Eq. 1.12) :

$$\Delta T_{vis} = RT_{ref} - RT \quad (1.11)$$

$$RVP = RVP' \left( \frac{\Delta T_{vis} - \Delta T_{vis,r}}{\Delta T'_{vis} - \Delta T_{vis,r}} \right) \quad (1.12)$$

Avec  $RT_{ref}$  le temps de réaction « étalon » obtenu dans les meilleures conditions expérimentales,  $RVP'$  la valeur maximale de performance visuelle pour le jeu commun de conditions expérimentales,  $\Delta T'_{vis}$  la meilleure valeur pour le jeu commun et enfin  $\Delta T_{vis,r}$  l'estimation de  $\Delta T_{vis}$  au seuil de contraste de lisibilité.

On sait donc désormais calculer la performance théorique et donc les temps de réaction théoriques à l'apparition d'une tache visuelle en fonction des conditions de luminance, de contraste et de taille de la cible. Il reste donc à mesurer empiriquement nos propres temps de réactions pour les comparer.



## CHAPITRE

# 2

# MESURES PRÉLIMINAIRES

Afin de contrôler le plus finement possible les valeurs de luminance et de contraste proposées aux sujets pendant les expérimentations, il est nécessaire de mesurer avec le plus grand soin, directement sur le simulateur, toutes ces valeurs. Les mesures ont été faites dans les conditions de l'expérimentation avec un « chromameter CS-100 » (Fig. 2.1).

## 2.1 Luminance globale (luminance de fond)

La première étape est de mesurer la luminance globale du simulateur, toutes faces allumées. Dans notre expérimentation, le simulateur sert à la fois de support pour afficher les cibles visuelles à percevoir et à la fois de source principale de lumière, toutes autres lampes dans la pièce étant éteintes. La luminance globale de l'expérimentation est donc la luminance des écrans, que l'on appellera ici « luminance de fond » pour les calculs de contraste, les cibles étant affichées directement sur une couleur uniforme sur l'écran.

Même si il existe des manières théoriques de convertir une couleur et son code RGB associé en une luminance générée par l'écran, il est nécessaire pour nous de mesurer cette transformation. En effet, il existe un certain nombre de biais tels que l'influence des autres écrans sur l'écran mesuré, l'influence des sources mineures de lumière, la fatigue de la lampe du projecteur, la dégradation de l'écran. En mesurant la relation couleur luminance générée directement dans le simulateur, on se débarrasse de tous ces biais et on peut viser la fonction de transfert réelle.

On effectue l'intégralité de nos mesures sur la face avant du simulateur car l'expérimentation ne se déroulera que sur cette face. Néanmoins, toutes les autres faces du simulateur sont éclairées, comme dans les conditions expérimentales, afin de prendre en compte l'influence de la luminosité des faces latérales et de la face au sol sur la luminosité de la face



FIGURE 2.1 – Chromameter CS100

avant. On utilise cinq points de mesure sur la face avant (répartis à la manière d'un dé à 6 faces), chaque mesure par point étant triplée pour éviter tout effet indésirable.

Les mesures ne sont faites que sur des nuances de gris. Pour ce faire, on affiche des couleurs dont les composantes R, G et B sont égales. Dans le simulateur, la couleur étant codée sur 8 bits, on peut afficher 256 nuances différentes : du noir le plus « pur » (Code RGB : (0, 0, 0)) au blanc le plus « pur » (Code RGB : (255, 255, 255)). Néanmoins, par soucis de praticité, on se limite à 17 nuances de gris en faisant des incrémentations de 16 en 16 dans le codage RGB des couleurs affichées. Par la suite, comme les trois composantes R, G et B sont toujours égales, on parlera de « niveau de gris » on se référant directement à la valeur des composantes. Par exemple, le noir sera appelé « gris 0 » étant donné que c'est une nuance de gris et que ses trois composantes R, G et B sont égales à 0.

Un extrait des résultats de mesure est disponible en Table 2.1, l'ensemble des mesures, détaillé par point de mesure, est disponible en annexes.

TABLE 2.1 – Extrait des mesures de luminance globale du simulateur en fonction de la nuance de gris affichée.

Couleur affichée	Niveau de gris	Moyenne
Noir	0	0.07 cd/m <sup>2</sup>
Gris sombre	80	3.73 cd/m <sup>2</sup>
Gris clair	176	23.22 cd/m <sup>2</sup>
Blanc	255	42.82 cd/m <sup>2</sup>

On peut alors ensuite tracer le graphe représentant l'évolution de la luminance globale du simulateur en fonction du niveau de gris affiché sur toutes ses faces (Fig. 2.2). De cette courbe, on déduit alors une régression polynomiale d'ordre 3 ( $R^2 = 1$ ) qui nous donne la relation mathématique (Eq. 2.1) permettant d'anticiper la luminance en fonction du niveau de gris. Cela nous permettra par la suite de choisir précisément les couleurs à afficher pour obtenir le contraste désiré. Avec  $L_G$  la luminance globale du simulateur et  $g$  le niveau de

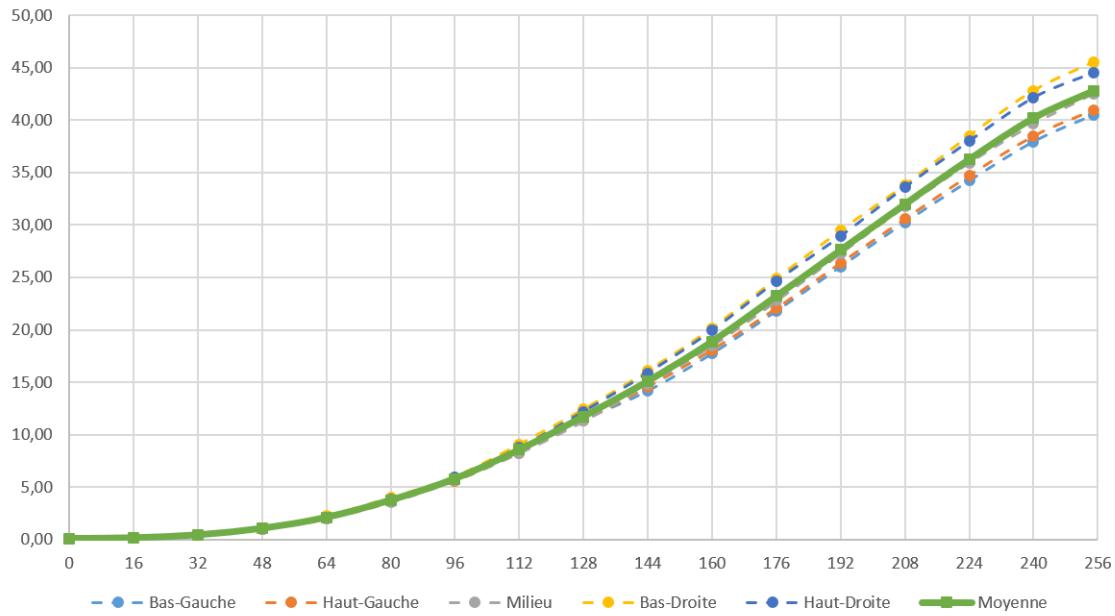


FIGURE 2.2 – Evolution de la luminance en fonction du niveau de gris sur les cinq points de mesure et leur moyenne.

gris normalisé, c'est à dire variant de 0 (= 0/255) à 1 (= 255/255).

$$L_G(g) = 0.69 - 18.37 g + 102.08 g^2 - 40.78 g^3 \quad (2.1)$$

## 2.2 Luminance de la cible (luminance de tache)

De la même manière, on réalise ensuite une série de mesures avec, en plus de l'intégralité des faces affichant un niveau de gris uni, un petit disque d'un diamètre de 3 cm, au centre de la face avant, dans une nuance de gris différente du fond. Cette cible sera la tache visuelle à détecter pour les sujets de l'expérimentation. Il est donc nécessaire de connaître également son évolution en luminance en fonction de son niveau de gris, le tout en fonction du niveau de gris du reste des écrans du simulateur qui vont très certainement influer.

On limite cette fois à 6 le nombre de conditions de niveau de gris pour le reste des écrans (les 6 niveaux qui ont été retenus pour l'expérimentation) tout en gardant l'incrément de 16 par 16 pour les niveaux de gris du disque. A cause de la taille de la cible à mesurer par rapport à la surface totale des écrans, on fait l'hypothèse que celle-ci ne perturbera pas la luminance globale tandis que le niveau de gris global influera sur la luminance du disque. De même, on ne réalise qu'une seule mesure au centre du disque, triplée encore une fois.

On présente un extrait des résultats des mesures en Table 2.2. L'ensemble des mesures est également disponible en annexes.

TABLE 2.2 – Extrait des mesures de luminance du disque en fonction de sa nuance de gris et de la nuance globale affichée.

Nuance de la cible	Fond : 0	Fond : 128	Fond : 255
0	0.06 cd/m <sup>2</sup>	5.51 cd/m <sup>2</sup>	20.30 cd/m <sup>2</sup>
64	1.13 cd/m <sup>2</sup>	6.55 cd/m <sup>2</sup>	21.40 cd/m <sup>2</sup>
144	8.20 cd/m <sup>2</sup>	13.57 cd/m <sup>2</sup>	28.40 cd/m <sup>2</sup>
255	23.37 cd/m <sup>2</sup>	28.77 cd/m <sup>2</sup>	43.40 cd/m <sup>2</sup>

On s'aperçoit que l'influence du niveau de gris du reste des écrans est très grand sur la luminance de la cible au centre de l'écran principal avec par exemple une multiplication par quasiment 300 de la luminance d'un cible noire (gris 0) sur fond noir (gris 0) par rapport à une cible noire (gris 0) sur fond blanc (gris 255).

On peut alors ensuite tracer les graphes représentant les évolutions de la luminance du disque en fonction du niveau de gris et de celui affiché sur toutes ses faces (Fig. 2.3). De ces courbes, on déduit les fonctions de transfert par régression polynomiale d'ordre 3 ( $R^2 = 1$ ) (Eq. 2.2). Avec  $L_{T,p}$  la luminance du disque sur la face avant du simulateur et  $g$  le niveau de gris normalisé de la cible et  $p$  celui du reste des écrans.

$$\begin{cases} L_{T,0}(g) = 0.44 - 10.67 g + 56.44 g^2 - 22.37 g^3 \\ L_{T,32}(g) = 0.60 - 10.67 g + 56.33 g^2 - 22.25 g^3 \\ L_{T,80}(g) = 2.11 - 10.87 g + 59.90 g^2 - 22.64 g^3 \\ L_{T,128}(g) = 5.89 - 10.74 g + 56.29 g^2 - 22.24 g^3 \\ L_{T,176}(g) = 11.45 - 10.33 g + 54.77 g^2 - 21.11 g^3 \\ L_{T,255}(g) = 20.69 - 10.85 g + 57.24 g^2 - 23.21 g^3 \end{cases} \quad (2.2)$$

## 2.3 Diamètre pupillaire

L'objectif était de vérifier une hypothèse : la luminance mesurée sur les écrans est égale à la luminance d'adaptation. Pour ce faire, toutes les mesures précédentes ont été faites par une personne équipée d'un oculomètre commercialisé par la société Ergoneers, le Dikablis Professional<sup>1</sup> (Fig. 2.4). Si on arrive à récupérer le diamètre pupillaire réel, mesuré en temps réel sur l'opérateur des mesures, on pourra, via la littérature et notamment la formule de Weale (Eq. 2.3) remonter à la luminance d'adaptation qui provoque ce diamètre pupillaire. On pourra alors la comparer à la luminance mesurée sur nos écrans et discuter notre hypothèse.

$$2r = 4.77 - 2.44 \tanh[0.3 \log_{10}(L_a)] \quad (2.3)$$

1. <https://lc.cx/dgLn>

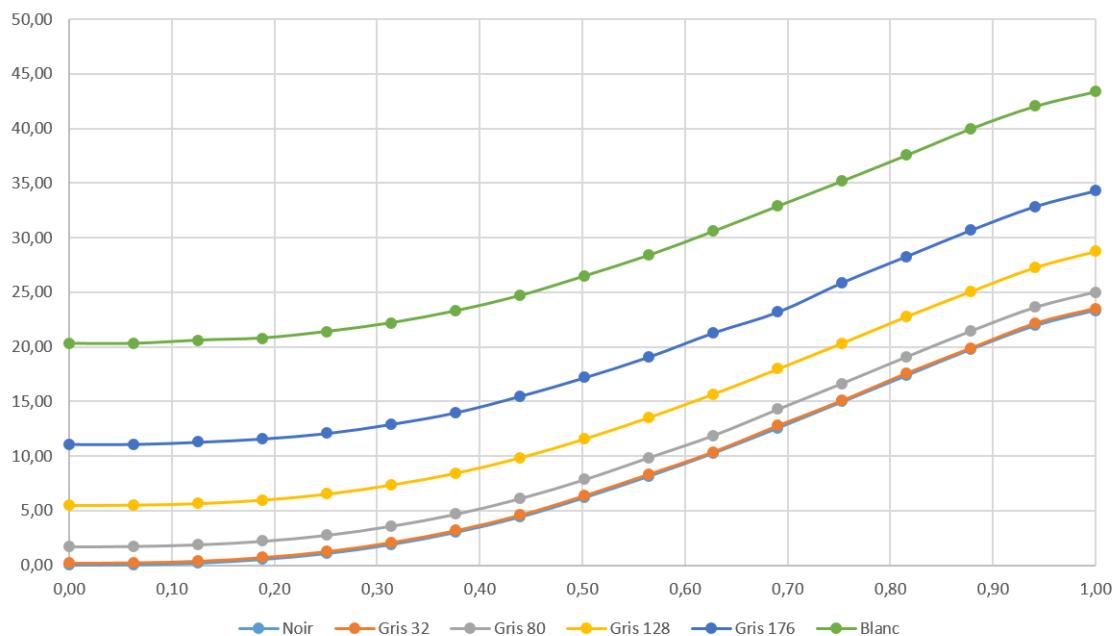


FIGURE 2.3 – Luminance de la cible en fonction de la luminance globale



FIGURE 2.4 – Oculomètre Dikablis Professional

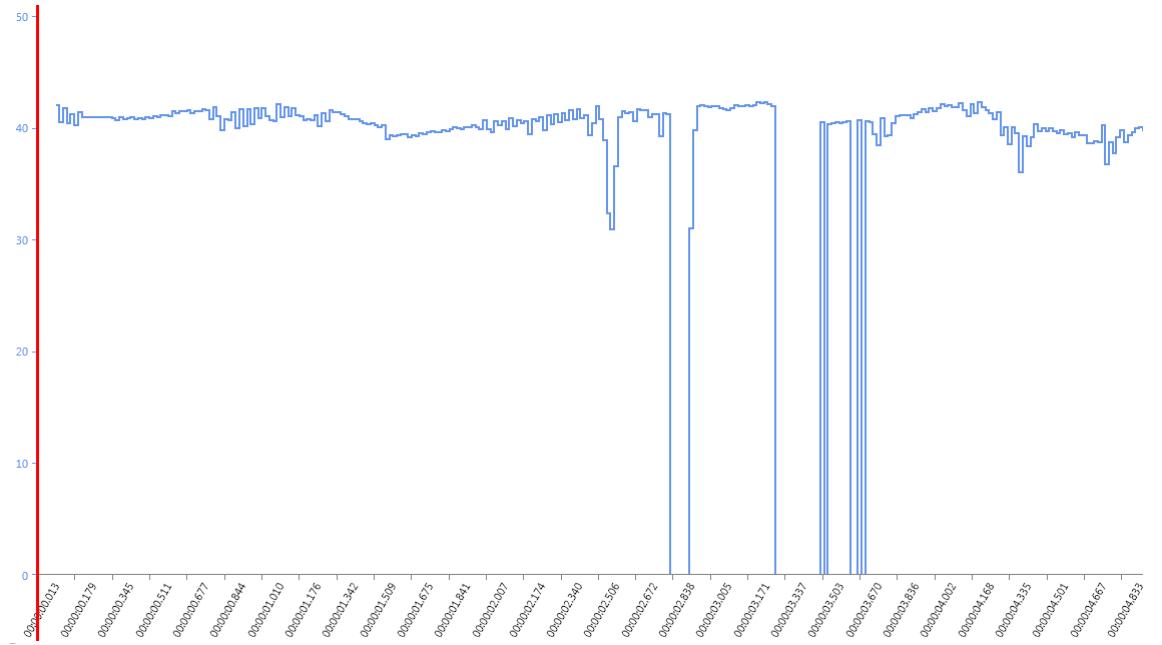


FIGURE 2.5 – Exemple d'une mesure en pixel de largeur pupillaire au cours du temps.

L'oculomètre filme les yeux dans le cadre de son fonctionnement et permet donc de relever parallèlement un certain nombre d'autres paramètres en temps réel et sur les deux yeux indépendamment. Notamment, la surface, la hauteur et la largeur pupillaire peuvent être mesurées. La Fig. 2.5 représente une mesure de largeur pupillaire au cours du temps. Les pics correspondent à un clignement des yeux. Malheureusement, ces mesures sont données en pixels et non pas directement en mm. Il faut donc faire une conversion par rapport à une référence connue (en millimètres) dans l'image.

On prend le diamètre du globe oculaire comme référence car sa taille est globalement constante (Fig. 2.6). Sur l'opérateur des mesures, cette valeur était de 25 mm. En mesurant ensuite directement sur une image, qu'on prend comme référence, la largeur de l'œil (mesure de référence) et le diamètre de la pupille on peut obtenir un ratio transformant une taille mesurée en pixel par le logiciel en millimètres réels (Eq. 2.4).

$$r = \frac{\text{diam oculaire mesure réelle (mm)} \times \text{diam pupillaire image ref (mm)}}{\text{diam oculaire image ref (mm)} \times \text{diam pupillaire image ref (px)}} \quad (2.4)$$

De même que précédemment, on prend des mesures dans cinq conditions de luminosité : condition de luminosité maximale, condition minimale et trois conditions intermédiaires. Les mesures sont disponibles dans en Table 2.3 et sur la Fig. 2.7.

La mesure est très imprécise et souffre de plusieurs facteurs d'erreurs : d'une part la mesure logicielle en pixel sur laquelle on ne peut avoir aucun contrôle, d'autre part, les mesures « à la règle » de la mesure de référence sur l'œil de l'opérateur et des valeurs sur l'image de référence pour le ratio sont assez imprécises et peuvent varier à l'ordre de grandeur du millimètre, ce qui est très important vis à vis des valeurs finales.

TABLE 2.3 – Mesure pupillaires en fonction de la luminosité

Gris	Luminance	Diam. théorique	Diam. mesuré œil gauche	Diam. mesuré œil droit
0	0.53 cd/m <sup>2</sup>	5.59 mm	5.00 mm	4.97 mm
64	2.09 cd/m <sup>2</sup>	4.54 mm	3.75 mm	3.21 mm
128	11.72 cd/m <sup>2</sup>	4.01 mm	3.42 mm	3.07 mm
192	27.62 cd/m <sup>2</sup>	3.78 mm	2.70 mm	2.53 mm
255	42.82 cd/m <sup>2</sup>	3.66 mm	2.45 mm	2.23 mm

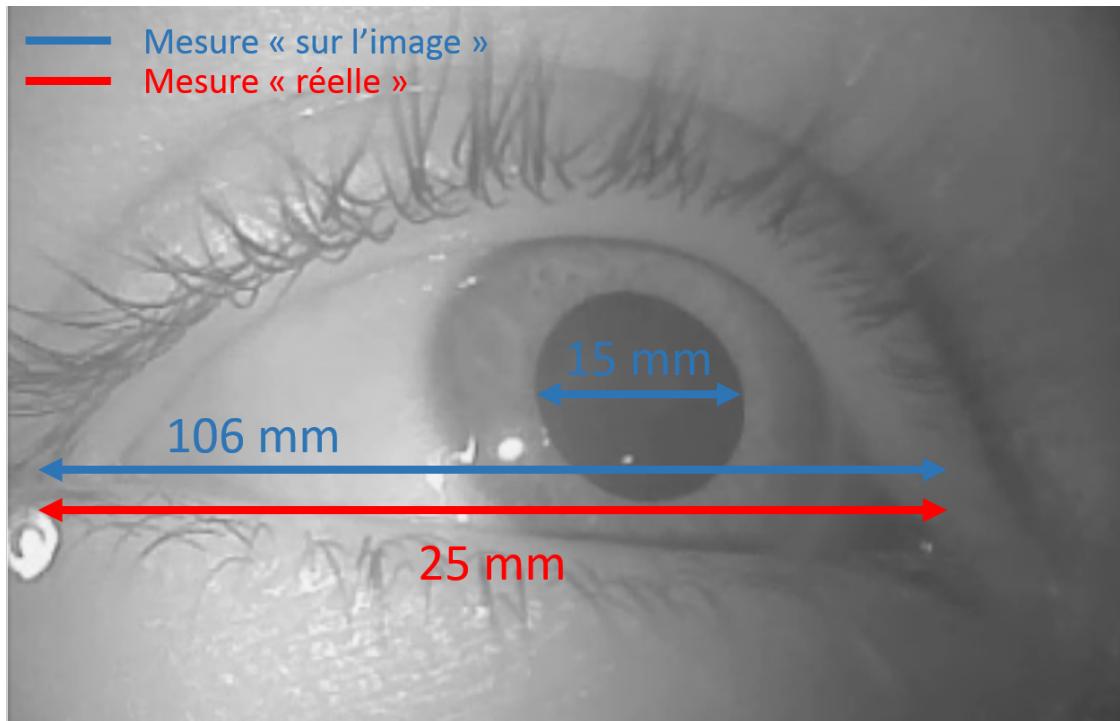


FIGURE 2.6 – Mesures pour la conversion pixel/mm.

Par conséquent nos mesures sont relativement éloignées des diamètres théoriques calculés en fonction de la luminance venant des écrans (Fig. 2.7). Cela ne permet donc pas de valider, ni d'invalider, notre hypothèse de départ qui était l'égalité entre la luminance générée par les écrans et la luminance d'adaptation des yeux. L'expérimentation devra donc se faire en posant cette hypothèse.

## 2.4 Absorption des verres des lunettes 3D

On profite également de ces séries de mesures préliminaires pour vérifier un chiffre donné par le constructeur des lunettes : seulement 17% de la luminosité globale arrive à l'œil après les lunettes stéréoscopiques. Cette donnée aurait pu nous servir directement en tant que valeur de transmittance  $T$  si on avait voulu réaliser notre expérimentation en stéréoscopie. Même si on choisira par la suite de faire passer nos sujets en conditions monoscopiques, il reste intéressant de vérifier ce chiffre de 17% avant sans justifications.

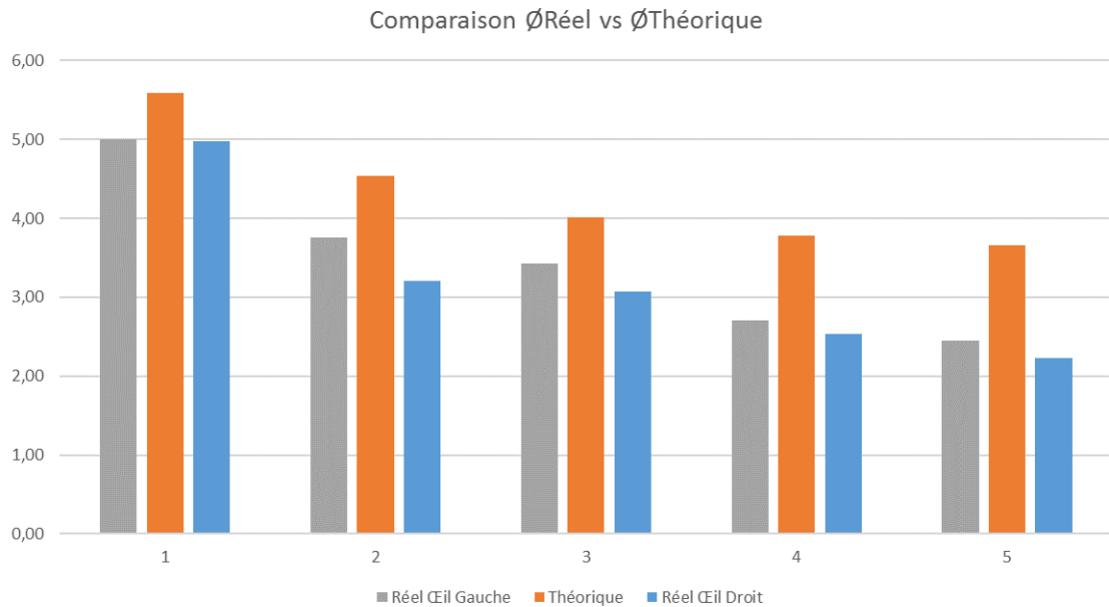


FIGURE 2.7 – Comparaison entre les mesures de diamètre pupillaire par rapport aux valeurs théoriques fonction de la luminance affichée.

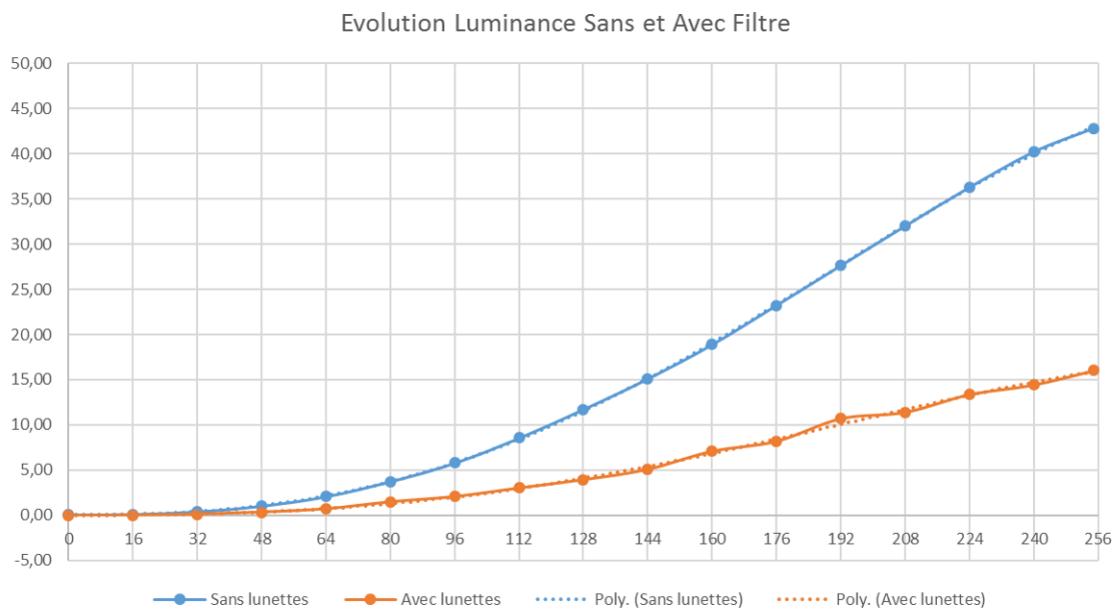


FIGURE 2.8 – Comparaison des mesures sans et avec filtre des lunettes stéréoscopiques.

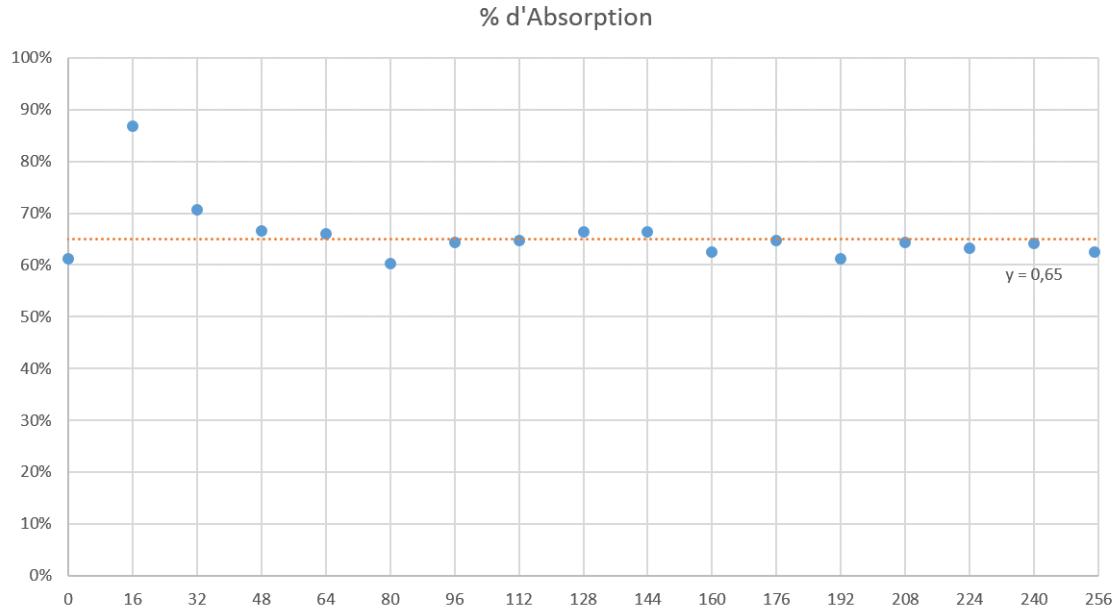


FIGURE 2.9 – Pourcentage d'absorption en fonction du niveau de gris affiché.

En parallèle des mesures décrites précédemment pour la luminance globale, on réalise également des mesures à travers un des deux verres de lunettes stéréoscopiques. Les résultats sont très probants (Fig. 2.9) et donnent une valeur d'absorption moyenne à 65% sur toute la gamme de luminances possibles dans le simulateur soit un taux d'absorption de 35%. Cependant, par construction, en stéréoscopie, les verres ne laissent passer la lumière que la moitié du temps pour permettre de nourrir chaque œil avec la bonne image. On retrouve donc bien la valeur de 17% annoncée par le constructeur.

Là encore, on déduit une régression polynomiale d'ordre 3 ( $R^2 = 1$ ) qui nous donne la relation mathématique (Eq. 2.5) permettant d'anticiper la luminance reçue à travers les verres de lunettes stéréoscopiques en fonction du niveau de gris. Avec  $L_F$  la luminance de la face avant du simulateur à travers le filtre et  $g$  le niveau de gris normalisé.

$$L_C(g) = 0.26 - 6.82 g + 36.55 g^2 - 13.86 g^3 \quad (2.5)$$

On possède désormais tous les éléments pour se lancer dans la vérification expérimentale du modèle de Rea : on connaît parfaitement le comportement en luminance du simulateur, on a vérifié le taux d'absorption des lunettes et on sait que l'hypothèse reliant la luminance de fond et la luminance d'adaptation doit être faite, faute d'avoir pu la démontrer quantitativement.



## CHAPITRE

# 3

## PROTOCOLE EXPÉRIMENTAL

### 3.1 Dispositif

#### 3.1.1 Hypothèses de travail

La première et principale hypothèse nécessaire au déroulement de cette expérimentation est que la luminance d'adaptation est considérée comme étant égale à la luminance de background : toutes les sources de luminosité parasites (voyants des ordinateurs, luminosité filtrant par la porte, ...) et la luminosité de la tache à visualiser sont jugées sans influences par rapport à la taille et à la puissance de la source lumineuse qu'est le background.

Cette hypothèse est à un impact à plusieurs niveaux dans l'expérimentation : elle caractérise la valeur de la luminance sensée arriver à l'œil mais elle permet également une simplification dans l'équation de calcul des contrastes proposés aux sujets (voir plus bas, section 3.2.2).

La seconde hypothèse est plus spécifique et concerne également l'équation de calcul du contraste. Rea et Ouellette utilisaient un filtre variable devant l'œil de leurs sujets pour atténuer l'intensité lumineuse et ont donc inclus une valeur de transmittance de ce filtre dans leur calcul du contraste. Nous n'utiliserons pas de filtre (l'expérimentation se déroule entièrement en 2D et donc sans lunettes stéréoscopique) et nous faisons l'hypothèse que cela revient à prendre une transmittance « parfaite », c'est à dire de valeur maximale  $T = 1$ .

Enfin, une dernière hypothèse concerne le mode opératoire directement. Même en ayant pris le plus grand soin possible avec un certain nombre de mesures préliminaires, on fait l'hypothèse qu'il est effectivement possible de reproduire l'expérimentation effectuée par

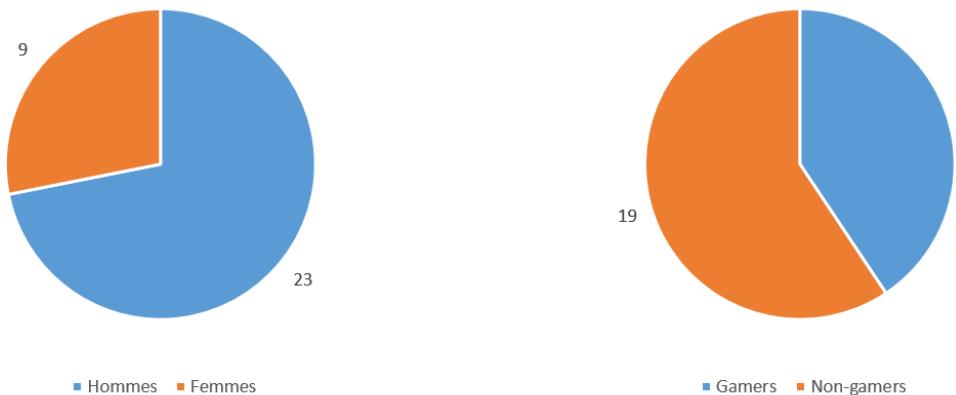


FIGURE 3.1 – Répartitions des sujets pour les expérimentation de contraste/luminance et de comparaison objective/subjective.

Rea et Ouellette et de la « traduire » en Réalité Virtuelle.

On rappelle que pour ce faire, on a calqué de la manière la fidèle possible notre mode opératoire sur le leur. Il existe néanmoins des différences, que ce soit sur l'outillage et la manière d'afficher les stimuli visuels mais également sur le nombre de sujets et surtout la quantité d'essai réalisés (dithyrambique du côté de Rea et Ouellette). L'influence du nombre d'essais réalisés est néanmoins nuancé par le fait qu'on effectue une vérification d'application du modèle contrairement à l'établissement d'une modèle à partir de zéro.

Afin d'éviter tout biais de la part des projecteurs, le passage des sujets s'est fait dans une période de temps rapprochée et avec une période de chauffe de 1h avant le premier sujet de la journée pour que les projecteurs soient stabilisés, comme recommandé par la CIE (IEC 61966-6 :2005).

### 3.1.2 Sujets & matériel

On réunit un total de 32 sujets (Fig. 3.1) pour cette expérimentation, les même que pour l'étude expérimentale décrite dans la partie précédente (Cf. 6). Ces derniers sont volontairement choisis jeunes (entre 20 et 27 ans) pour minimiser l'impact de la dégradation de la vision avec l'âge. Tous les sujets ont donc une vision parfaite, ou corrigée et assimilée parfaite. La moyenne d'âge est de 25 ans, avec un écart-type de 1,8 ans.

De même, les conditions matérielles sont identiques que pour la première expérimentation : dans un simulateur de type CAVE, les sujets sont assis dans un fauteuil de voiture, la tête bien calée dans l'appui-tête. Néanmoins, pour cette expérimentation, les sujets n'utilisent pas de lunettes stéréoscopiques. L'expérimentation se fait donc en affichage monoscopique. Les sujets sont placés de manière à avoir les yeux à 2 mètres du centre de l'écran « principal », la face avant.

### 3.1.3 Tâche à effectuer

L'expérimentation se déroulait en plusieurs étapes. La première consistait simplement en un petit questionnaire visant à relever quelques informations sur le sujet en lui-même : âge, défauts éventuels et correction de vision, sexe, pratique des jeux vidéo. S'en suivait alors l'installation du sujet dans le simulateur et une explication de la tâche à effectuer pour l'expérimentation.

De manière analogue aux expérimentations de Rea et Ouellette, les sujets devaient réagir à l'apparition de petits stimuli visuels sous la forme de disques uniformes de couleur sur un fond uniforme. La consigne était d'appuyer le plus vite possible, dès la perception du stimulus, sur un bouton spécifique (et toujours le même) d'une manette de console. Les couleurs uniformes des cibles et du fond étaient des nuances de gris choisies telles que les valeur de luminance aux niveaux des yeux et de contraste de l'un par rapport à l'autre soient parfaitement connues (voir plus bas).

La seconde étape était une session blanche d'apparition de stimuli, c'est à dire sans stockage des mesures de temps de réaction, pour assurer une bonne prise en main de la tâche à effectuer, limiter le biais de la découverte et ainsi s'assurer que les temps de réactions mesurés sont les meilleurs. Le nombre de stimuli était légèrement plus faible que pour les sessions normales. La luminance choisie pour la session d'essai était différente de toutes les autres et choisie au milieu du spectre pour éviter tout biais d'habituation.

Les étapes suivantes consistaient au déroulement des différentes sessions (voir le logigramme de déroulement global en Fig. 3.2. Le sous-logigramme « CIBLES » est présenté en Fig. 3.3), une pour chaque condition de luminance. Après chaque session, les conditions lumineuses étaient immédiatement réglées pour la session suivante puis un temps d'attente de 3 à 5 minutes était observé pour permettre aux yeux du sujet de s'adapter à la nouvelle luminance de fond. La consigne étaient de garder le regard sur les écrans du simulateur. Une fois le sujet adapté aux conditions lumineuses, il pouvait lancer de lui-même l'apparition des stimuli en appuyant sur un autre bouton spécifique de la manette de console.

L'ordre des luminances de fond proposées aux sujets était semi-aléatoire : distribution aléatoire pour chaque sujet mais de telle manière, qu'au global, autant de sujets aient rencontré la condition lumineuse  $p$  en  $k$ -ième position. Par exemple, autant de sujets ont eu la condition de luminance la plus sombre en première session, qu'en deuxième, qu'en troisième, etc.

Tous les contrastes définis et possibles (voir plus bas, section 3.2.2) sont proposés au sujet pour chaque niveau de luminance. La présentation se fait dans un ordre aléatoire grâce à l'algorithme de mélange de Fisher-Yates (aussi appelé *Algorithme P*, voir paragraphe suivant). Le mélange se fait à chaque nouvelle condition lumineuse mise en place : tous les sujets ont donc eu potentiellement une série différente à chaque fois et il n'était pas possible de prédire l'ordre avant de lancer l'apparition des cibles. L'apparition de chaque cible était espacée à chaque fois d'un temps incompressible de 1000 ms auquel s'ajoutait un temps aléatoire compris en 2000 et 4000 ms.

Le principe de fonctionnement du mélange de Fisher-Yates, pour mélanger un tableau « T »

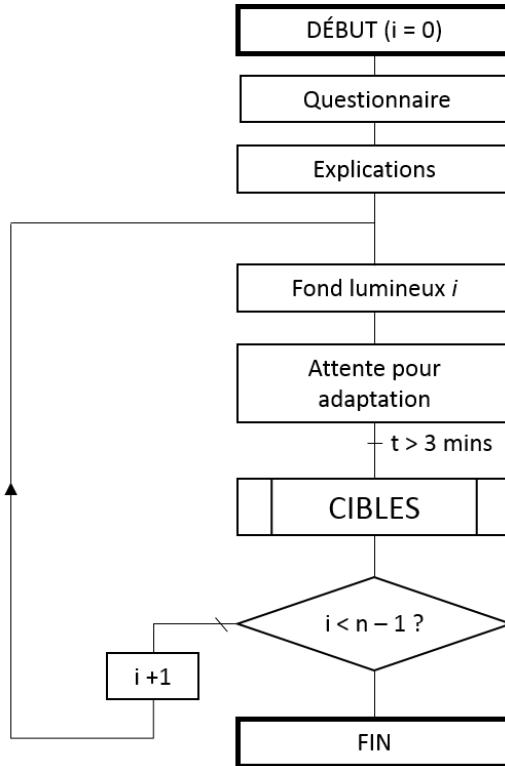


FIGURE 3.2 – Déroulement global de l’expérimentation contraste/luminance.

de  $n$  éléments (indices allant de 0 à  $n - 1$ ), est le suivant :

Pour  $i$  allant de  $n-1$  à 1 faire :  
 $j \leftarrow$  entier aléatoire entre 0 et  $i$   
échanger  $T[j]$  et  $T[i]$

L’algorithme a une complexité linéaire et donne toutes les permutations avec la même probabilité. Il a été pensé de telle sorte que chaque élément du tableau ne bouge qu’une seule fois.

Le déroulement d’une session, c’est à dire à un niveau de luminance donné, est résumé dans le logigramme en Fig. 3.3.

## 3.2 Choix des conditions expérimentales

### 3.2.1 Luminance

Les conditions lumineuses de fond sont choisies de manière arbitraire, espacées régulièrement sur l’axe 0-255 bits (valeurs minimale et maximale pour coder une couleur sur 8 bits), d’une valeur d’au moins 32 bits (soit au moins 1/8ème de la plage totale de gris). On choisit

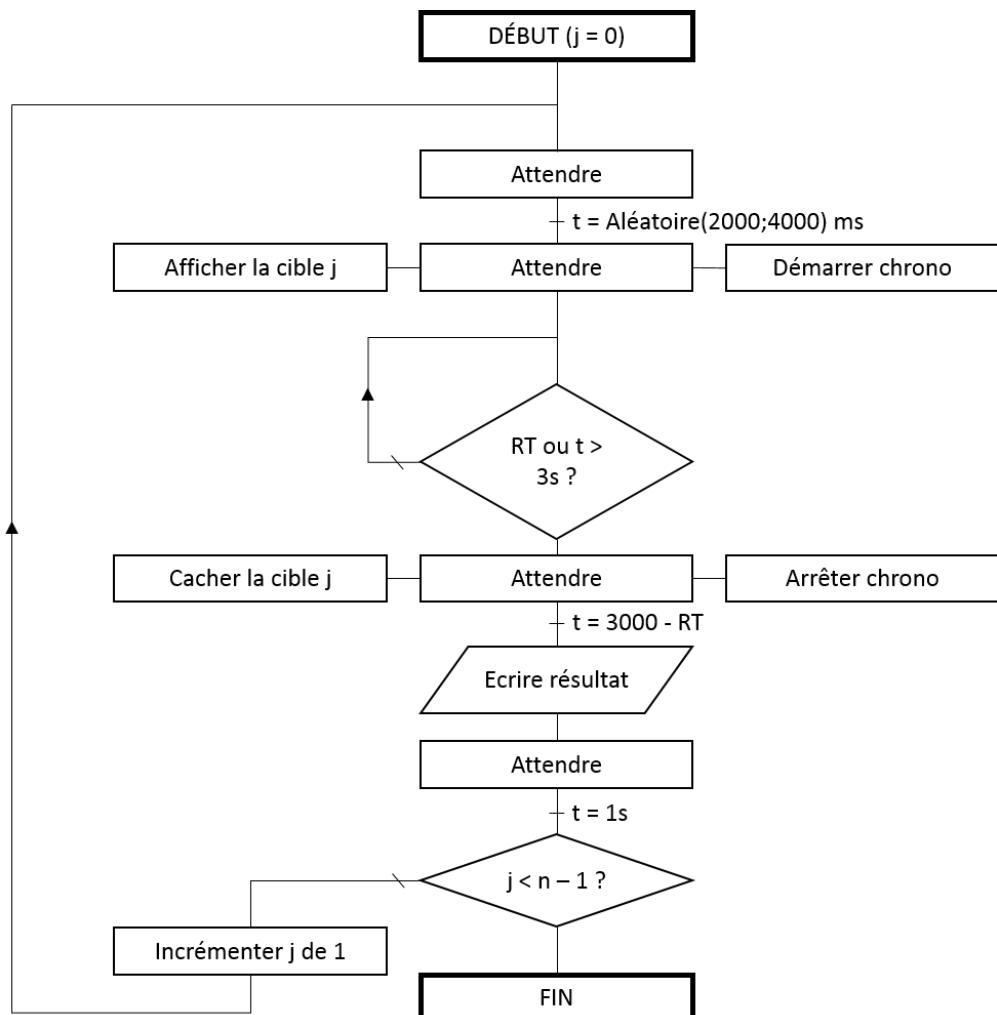


FIGURE 3.3 – Déroulement d'une session de mesure de temps de réactions pour une luminance de fond donnée.

évidemment les conditions extrêmes, à savoir la luminance liée à un background entièrement noir et la luminance liée à un background entièrement blanc. On choisit également la luminance associé au gris médian, c'est à dire 128 bits pour chacune des composantes R, G et B de la couleur affichée. Comme on a pu voir précédemment dans le chapitre sur les mesures préliminaires préalables au bon déroulement de l'expérimentation, les niveaux de luminance choisis sont associés aux niveaux de gris 0, 32, 80, 128, 176 et 255. On se limite à 6 conditions lumineuses différentes dans un soucis de temps requis pour faire passer les sujets sur tous les niveaux, pour tous les contrastes calculés.

Tous nos niveaux de luminance sont compris dans les conditions pour lesquelles Rea et Ouellette ont établi leur modèle. Les limites de ce dernier sont fixées entre 0.53 et 801  $Td$  (Trolands<sup>1</sup>) vus à travers une pupille artificielle d'un diamètre de 2 mm. Cela équivaut donc à des extrema lumineux à respectivement  $0.17 cd/m^2$  et environ  $255 cd/m^2$ ; tandis que nos extrema sont situés à  $0.53 cd/m^2$  pour la condition la plus sombre et  $42.82 cd/m^2$  pour la condition la plus claire.

### 3.2.2 Contraste

Pour calculer les différents niveaux de contraste que l'on va proposer aux sujets, on se base sur la même équation de contraste utilisée par Rea et Ouellette (Eq. 3.1). Cette dernière permet de calculer des valeurs de contraste comprises entre 0 et 1. On choisit, dans la mesure du possible (voir plus bas), de prendre des valeurs de contraste à intervalle régulier sur l'axe 0-1, espacés d'un dixième au maximum et d'un vingtième pour détailler la partie 0-0.25 de l'axe. Les valeurs de contraste cible/fond à atteindre sont donc : 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 et 1 ; soit 13 valeurs. On peut atteindre ces valeurs de contraste par valeur « inférieure » c'est à dire avec une cible plus sombre que le fond (typiquement une cible grise sur fond blanc), mais également par valeur « supérieure », c'est à dire avec une cible plus claire que le fond sur lequelle elle est affichée. Cela nous donne un total maximal théorique de 26 contrastes différents à tester par niveau de luminosité. Néanmoins, dans la plupart des cas, il n'est pas possible d'afficher tous les contrastes souhaités voire de ne pas pouvoir afficher du tout soit les contrastes « par valeur supérieure », soit les contrastes « par valeur inférieure ». Typiquement, dans la condition de fond entièrement noir, il n'est pas possible d'afficher des cibles plus sombres et on ne peut donc avoir que des contrastes « supérieurs ».

$$C = \frac{|(TL_b + L_v) - (TL_t + L_v)|}{TL_b + L_v} = \frac{T|L_b - L_t|}{L_a} \quad (3.1)$$

Il est maintenant nécessaire de déterminer quelle doit être la couleur du stimuli affiché, par rapport à la couleur du fond, pour atteindre précisément le contraste désiré. C'est ici que rentrent en jeu nos deux hypothèses préalablement établies (Cf. 3.1.1) : l'égalité entre la luminance d'adaptation et la luminance de background ( $L_a = L_B$ ) d'un côté et la transmittance prise à sa valeur maximale ( $T = 1$ ) de l'autre. On peut donc simplifier

1. Le Troland est une unité de luminance pupillaire, définie par l'équation suivante :  $Td = L \times p$ ; avec  $L$  la luminance qui arrive à l'œil en  $cd/m^2$  et  $p$  la surface de la pupille en  $mm^2$ .

l'équation 3.1 et la transformer en une équation adaptée à notre dispositif (Eq. 3.2), avec  $L_b$  la luminance du fond et  $L_t$  la luminance de la « target », le stimuli :

$$C = \frac{T|L_b - L_t|}{L_a} \Rightarrow C = \frac{|L_b - L_t|}{L_b} \quad (3.2)$$

Grâce aux régressions linéaires établies précédemment (cf. ), on peut écrire  $L_b$  et  $L_t$  comme des fonctions du niveau de gris dans notre simulateur, soit l'équation 3.3 avec  $B$  la valeur entre 0 et 255 associée à la nuance de gris du fond. De même, l'équation 3.4 avec  $T$  la valeur entre 0 et 255 associée à la nuance de gris de la cible, l'indice  $B$  étant relatif à la nuance de gris du fond :

$$L_b = f\left(\frac{B}{255}\right) \quad (3.3)$$

$$L_t = f_B\left(\frac{T}{255}\right) \quad (3.4)$$

On obtient donc, au final, l'équation suivante qui nous permet de calculer avec précision la valeur de la nuance de gris de la cible à afficher lorsque qu'on connaît la nuance de gris du fond pour afficher un contraste choisi (Eq. 3.5) :

$$C = \frac{|f\left(\frac{B}{255}\right) - f_B\left(\frac{T}{255}\right)|}{f\left(\frac{B}{255}\right)} \quad (3.5)$$

Néanmoins, comme décrit précédemment, on souhaite pouvoir calculer des contrastes « supérieurs » et « inférieurs », hors cette équation ne fonctionne que si la luminance de fond est supérieure à la luminance de la cible. Il faut donc adapter l'équation pour calculer les contrastes « supérieurs » (Eq. 3.6) :

$$C = \frac{|f\left(\frac{B}{255}\right) - f_B\left(\frac{T}{255}\right)|}{f_B\left(\frac{T}{255}\right)} \quad (3.6)$$

On présente dans les tables suivantes les valeurs de nuances de gris nécessaires pour afficher les contrastes par valeur « supérieure » (Tab. 3.1) et par valeur « inférieure » (Tab. 3.2). La première ligne donne les contrastes obtenus, la première colonne les niveaux de gris en background et le reste du tableau le niveau de gris du stimuli. On peut voir que notamment dans le cas du background le plus sombre, les valeurs ne varient que de 1 bit et donc qu'il ne serait pas possible avec notre système d'avoir un pas de variation du

contraste plus fin. On remarque que, pour les contrastes supérieurs 0.15 et 0.2 sur fond noir, la même valeur de gris est associée. Ce cas est unique et vient du fait que la régression linéaire est continue alors que l'on cherche des valeurs ponctuelles, pour des valeurs très proches ni l'arrondi ni la troncature ne permettent de les départager. On a donc ignoré une des deux valeurs pendant l'expérimentation.

TABLE 3.1 – Valeurs des nuances de gris du stimulus nécessaire pour afficher un contraste par valeur supérieur désiré.

C	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
<b>0</b>	16	17	18	18	19	20	21	22	23	24	25	26	27
<b>32</b>	33	35	36	37	38	39	41	42	44	46	47	49	50
<b>80</b>	85	88	90	93	95	98	102	106	110	114	118	121	125
<b>128</b>	133	138	142	147	151	156	164	172	180	187	195	203	211
<b>176</b>	182	190	198	206	214	222	239						
<b>255</b>													

TABLE 3.2 – Valeurs des nuances de gris du stimulus nécessaire pour afficher un contraste par valeur supérieur désiré.

C	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
<b>0</b>													
<b>32</b>	31	29	28	27	26	24	22	19	15				
<b>80</b>	79	76	74	71	69	66	61	57	52	47	52	37	31
<b>128</b>	123	118	114	109	105	101	93	85	78	70	62	52	40
<b>176</b>	167	160	153	147	140	135	123	112	101	91	79	67	50
<b>255</b>	232	218	206	196	186	177	161	146	132	117	102	86	66

A titre d'exemple, lorsque la couleur de background est un gris 128 (Code RGB = (128, 128, 128), pour afficher un stimulus contrasté par valeur supérieure à 0.3, il faut que la cible soit un gris 156, c'est à dire avec un code RGB égal à (156, 156, 156).

Maintenant que l'on a choisi les luminances de fond, que l'on sait à quelle nuance de gris elles correspondent, qu'on a choisi les contrastes que l'on voulait afficher et que l'on sait à quelles couleurs ils correspondent, on peut passer à la partie purement expérimentale et inviter des sujets à confronter leurs temps de réaction à l'apparition de nos stimuli calibrés. Les résultats théoriques et pratiques de ces expérimentations sont présentés dans le chapitre suivant.

## CHAPITRE

# 4

## RÉSULTATS

### 4.1 Prédictions du modèle théorique

On calcule les temps de réactions prévus par le modèle que l'on chercher à vérifier. Ces temps sont en millisecondes (ms). On présente ces résultats sous forme de tableau avec en colonne les différents contrastes affichés et en ligne les niveaux de gris du background (Table 4.1).

On laisse une case vide dans le tableau lorsque le modèle n'est pas capable de prédire un temps de réaction : le sujet n'est pas sensé réussir à distinguer le stimuli par rapport au fond. On remarque également que certaines valeurs sont complètement démesurées (typiquement, il faudrait théoriquement 25 secondes à un individu pour distinguer un stimuli contrasté à 0.20 sur un fond gris 80). Ces valeurs sont dues à un diviseur très proche de zéro dans les formules utilisée qui rend les valeurs de contraste proches du contraste seuil théorique très volatiles.

Pour l'ensemble des conditions, les valeurs de temps de réaction (une fois passé le seuil limite de contraste) croissent régulièrement jusqu'à tendre vers une asymptote pour les conditions les plus lumineuses.

### 4.2 Mesures réelles

Les mesures réelles sont mesurées directement par le logiciel qui affiche les stimuli. Les résultats sont ensuite écrits dans un fichier externe que l'on récupère et que l'on traite ensuite dans Excel. La précision des mesures est limitée par le nombre d'images par seconde auquel le programme tourne : à chaque image, le programme vérifie si le sujet appuie ou

## CHAPITRE 4. RÉSULTATS

---

TABLE 4.1 – Temps de réaction théoriques (en ms) prédits par le modèle de Rea et Ouellette.

C/L	0	32	80	128	176	255
<b>0.05</b>						
<b>0.10</b>						
<b>0.15</b>					10865	3594
<b>0.20</b>		25435	1924	1429	1212	
<b>0.25</b>		1686	1041	912	839	
<b>0.30</b>		1039	788	725	685	
<b>0.40</b>	2663	696	596	568	549	
<b>0.50</b>	1272	577	517	499	546	
<b>0.60</b>	929	517	473	549	449	
<b>0.70</b>	773	481	446	434	425	
<b>0.80</b>	4451	684	456	426	416	409
<b>0.90</b>	2448	626	439	412	403	396
<b>1.00</b>	1720	585	425	402	393	387

non sur le bouton de détection et stoppe le chronomètre le cas idoine. Toute action entre 2 images est donc ignorée jusqu'au calcul suivant. C'est pourquoi la précision des mesures de temps est précise à un demi-temps inter-image près. Le programme fonctionnant à 120 images par secondes (la limite se faisant au niveau du projecteur), la précision est donc de 1/240ème de seconde soit 4 ms. Au regard des valeurs attendues (les temps les plus rapides prévus dans nos conditions par le modèle sont de 400 ms) cela représente une erreur de 1% qui peut donc être négligée.

Les moyennes des temps de réaction mesurés et leurs écart-types sont listés dans la Table 4.2. Etant donné que la population de sujet était supérieure ou égale à 30, on a pu calculer l'écart type réel (Pearson) et non pas un estimateur. Les valeurs si situent entre 371 ms et 670 ms avec une moyenne des moyennes (en excluant les valeurs divergentes, voir plus loin) à 437 ms (écart-type = 53 ms) et un écart-type moyen de 96 ms (écart-type = 82 ms). On nota particulièrement la valeur de temps de réaction (2013 ms) atteinte pour un contraste de 0.05 et une luminance de fond de 0.41 cd/m<sup>2</sup>).

La performance R est définie telle que l'inverse du temps de réaction RT :  $R = 1/RT$ . On l'utilise, plutôt que le temps de réaction, pour décrire les résultats obtenus : la performance augmente quand le temps de réaction diminue et inversement, ce qui semble plus logique. Les figures 4.1, 4.2 et 4.3 montrent, pour chaque condition lumineuse, la comparaison entre la performance théorique que l'on devrait avoir et la performance réelle mesurée pendant l'expérimentation. Les points correspondant à une performance nulle décrivent les points en dessous du seuil de perception : le temps de réaction est alors « infini » et donc son inverse (la performance) tend vers 0.

## CHAPITRE 4. RÉSULTATS

---

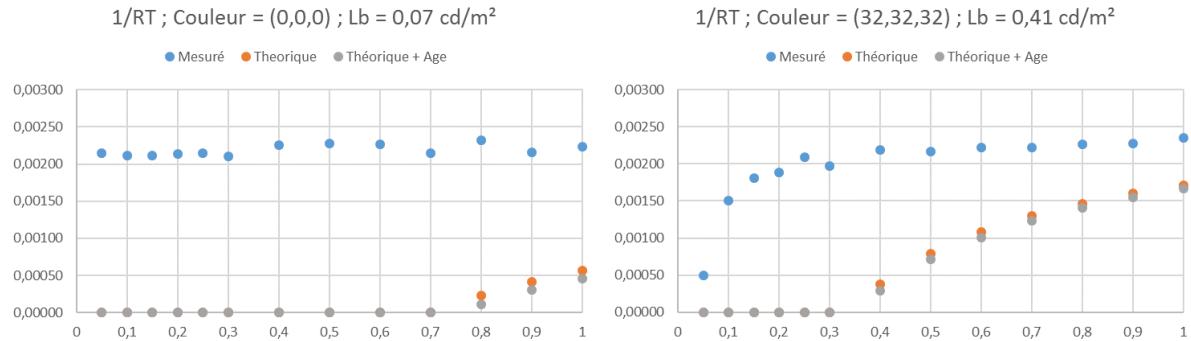


FIGURE 4.1 – Performance théorique et réelle pour l’expérimentation de détection des stimuli contrastés, conditions 0 et 32.

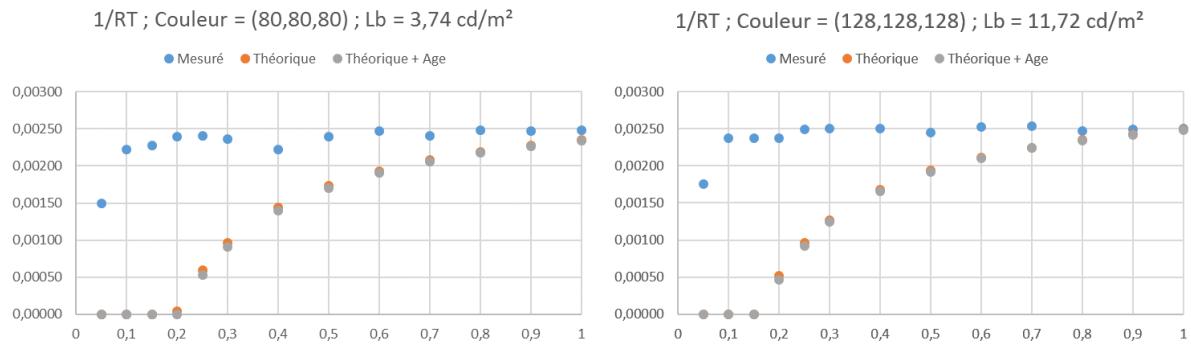


FIGURE 4.2 – Performance théorique et réelle pour l’expérimentation de détection des stimuli contrastés, conditions 80 et 128.

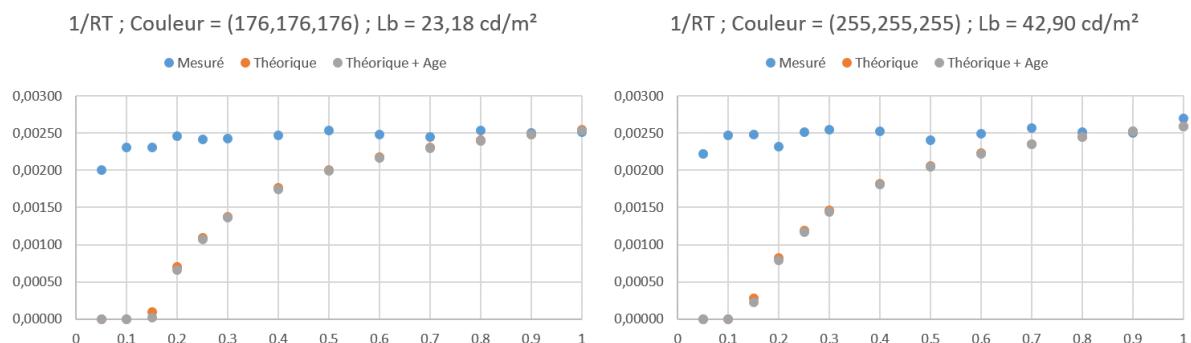


FIGURE 4.3 – Performance théorique et réelle pour l’expérimentation de détection des stimuli contrastés, conditions 176 et 255.

TABLE 4.2 – Temps de réaction moyens réels (en ms) mesurés pendant l’expérimentation et leurs écart-types moyens.

C/L	<b>0</b>	<b>32</b>	<b>80</b>	<b>128</b>	<b>176</b>	<b>255</b>
<b>0.05</b>	465 ± 58	2013 ± 429	670 ± 112	572 ± 314	500 ± 290	450 ± 155
<b>0.10</b>	474 ± 98	663 ± 447	450 ± 86	421 ± 66	434 ± 71	406 ± 50
<b>0.15</b>	474 ± 85	553 ± 286	439 ± 78	421 ± 85	434 ± 99	404 ± 48
<b>0.20</b>	468 ± 67	531 ± 323	418 ± 48	422 ± 87	407 ± 75	432 ± 88
<b>0.25</b>	465 ± 158	479 ± 63	415 ± 68	101 ± 50	141 ± 62	398 ± 92
<b>0.30</b>	475 ± 141	507 ± 215	423 ± 73	399 ± 53	412 ± 63	393 ± 72
<b>0.40</b>	443 ± 74	457 ± 59	451 ± 221	400 ± 59	406 ± 63	396 ± 52
<b>0.50</b>	440 ± 44	461 ± 78	418 ± 71	409 ± 62	394 ± 42	415 ± 67
<b>0.60</b>	440 ± 58	451 ± 66	405 ± 52	396 ± 52	403 ± 6	402 ± 81
<b>0.70</b>	466 ± 94	450 ± 80	416 ± 72	394 ± 49	408 ± 57	390 ± 48
<b>0.80</b>	432 ± 68	443 ± 67	402 ± 46	405 ± 80	394 ± 48	398 ± 67
<b>0.90</b>	464 ± 79	440 ± 61	404 ± 58	402 ± 61	400 ± 64	399 ± 77
<b>1.00</b>	448 ± 152	426 ± 41	403 ± 52	401 ± 77	398 ± 64	371 ± 48

### 4.3 Analyse et discussion

Le premier résultats qui frappe est qu’alors que le modèle de Rea prévoit que certains stimuli ne doivent pas être vus (étant en dessous du seuil), nos sujets ont été capables de voir toutes les tâches lumineuses qui leur étaient présentées, et ce, quelles que soient les conditions de luminance et de contraste.

Deuxièmement, la différence entre le comportement théorique prévu et les mesures sur nos sujets est radicale. Alors que la performance est sensée augmenter avec le contraste et la luminance, il apparaît que notre performance est toujours constante. Les sujets avaient un temps de réaction globalement identique avec des conditions lumineuses pourtant très différentes. En se fiant aux prédictions du modèle de Rea et Ouellette, les sujets ne devraient pas être aussi capables : il devraient d’abord ne rien voir puis, alors que la luminance et surtout le contraste augmentent, être de plus en plus bons. Les comportements étant aussi radicalement opposés, nous n’avons pas jugé nécessaire de déployer une méthode statistique pour vérifier si les échantillons de mesure (mesures théoriques et mesures pratiques) étaient appariés.

Au final, nos résultats ne correspondent pas du tout avec les prévisions de Rea et Ouellette mais convergent néanmoins vers les même valeurs. Bien que nous ayons fait de notre mieux pour traduire l’expérimentation initiale dans notre simulateur, il existe néanmoins toujours des différences fondamentales entre les prévisions et les mesures.

La différence principale entre les deux dispositifs expérimentaux et la quantité de lumière

## CHAPITRE 4. RÉSULTATS

---

qui arrive à l'œil, et la quantité de sources lumineuses. Du côté de Rea et Ouellette, la seule source de lumière était un petit écran situé à une distance de 1.68 *m* du sujet, en plus d'une lumière dirigée directement dans l'œil au moyen d'un miroir semi transparent. De plus, la quantité de lumière arrivant à l'œil était également contrôlée avec un filtre devant l'œil. De notre côté, la lumière est issue d'écrans géants entourant le sujet ainsi qu'une multitude de petites lumières indirectes : les projecteurs émettent leur lumière en direction des écrans. Nous n'avons pas braqué de lumière directement dans l'œil de nos sujets et nous n'avons pas utilisé de filtre pour atténuer la lumière (tout en plaçant néanmoins des hypothèses). Enfin, Rea et Ouellette filtraient la lumière arrivant à l'œil au moyen d'un filtre mais le montage ne semble pas être complètement étanche et a pu laisser passer de la lumière parasite non atténuée.

Ainsi, de part la comparaison entre nos résultats et les préditions, à notre avis, le modèle de Rea et Ouellette ne devrait pas être utilisé dans le domaine de la Réalité Virtuelle pour déterminer la performance de l'œil (par rapport à la modélisation de la vision humaine), avec pour seule base le contraste et la luminance.

## CHAPITRE 4. RÉSULTATS

---

# CONCLUSION

L'objectif de cette première expérimentation était d'alimenter la brique « contraste luminance » de notre modèle de score pour la Réalité Virtuelle. Le postulat de base était qu'il existe des modèles qui définissent une performance, entièrement basée sur le système visuelle, et qui prennent notamment en paramètre d'entrée le contraste et la luminance, mais ces modèles n'ont pas été fait pour la Réalité Virtuelle, ni même pour des écrans quelconques. On cherche alors à démontrer la transposabilité d'un de ces modèles dans notre domaine afin de pouvoir l'utiliser directement dans notre score.

La démarche était donc de reproduire le protocole expérimental ayant servi à établir le modèle initial, mais en le transposant dans un simulateur, et de quantifier la bonne corrélation entre les valeurs prévues par le modèle et celles mesures sur des sujets. Les résultats sont de ce point de vue là décevant car ils montrent un comportement radicalement différent de ce qui était prévu. Bien que l'on ait quelques hypothèses sur l'origine de cette différence de comportement, cette dernière reste encore difficile à expliquer pleinement.

On peut établir un certain nombre de choses de cette expérimentation. Premièrement, le modèle de Rea, et plus particulièrement la variante développée par Rea et Ouellette avec des temps de réaction, ne semble pas fonctionner pour la Réalité Virtuelle. Ensuite, et par extension, on ne peut pas utiliser directement le modèle de Rea comme on le souhaitait pour établir le critère de « contraste/luminance » de notre modèle de score. Cela n'implique pas néanmoins l'impossibilité de corrélérer notre critère avec un autre modèle de performance visuelle.

Se pose alors la question de la continuité à donner aux résultats de cette expérimentation. On peut choisir de continuer à approfondir la performance visuelle, en créant par exemple notre propre modèle, mais on choisit de s'intéresser à une autre brique élémentaire de notre modèle de score : la latence. Ce changement de sujet permet de travailler plusieurs aspects du modèle et de ne pas s'enfermer dans une seule branche. Le chapitre suivant concerne donc la partie expérimentale autour du traitement du critère suivant, la latence.

## CONCLUSION

---

## **Cinquième partie**

### **Expérimentation : latence**



# INTRODUCTION

qsd



CHAPITRE

1

# MESURES PRÉLIMINAIRES

qsd



## CHAPITRE

# 2

## PROTOCOLE EXPÉRIMENTAL

### 2.1 Dispositif

#### 2.1.1 Hypothèses de travail

qsd



# CONCLUSION

qsd



## CONCLUSION GÉNÉRALE

**A**venir ... Aliquam erat volutpat. Aliquam a mattis nunc, et aliquam felis. Etiam et dolor vel odio eleifend porta vel quis turpis. Praesent tincidunt aliquet egestas. Phasellus enim neque, tincidunt ut pretium at, bibendum at lectus. Cras scelerisque, tellus id euismod aliquet, ante elit lobortis velit, in sodales quam enim et sem. Nulla id auctor tellus. Duis vehicula orci ut nisi iaculis scelerisque. Phasellus vitae odio non massa ornare posuere in a felis. Morbi pretium eleifend ipsum, sit amet consequat mauris elementum ac. Aenean hendrerit viverra magna nec bibendum. Nam non ultrices mauris. Etiam dictum turpis et ipsum eleifend, commodo volutpat odio tempus. Nullam eget felis ornare, cursus metus sit amet, ornare justo. Donec sagittis ligula ut leo auctor pretium.

On peut imaginer un usage « détourné » pour déterminer la lisibilité d'un simulateur : en pondérant tous les critères liés à cette tâche (acuités, contraste et luminance, fluidité en cas de haute vitesse) et en mettant la pondération des critères « inutiles » à cette tâche à 0.

---

## CONCLUSION GÉNÉRALE

---

# **Bibliographie**



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

ANSES. 2014. *Effets Sanitaires Potentiels des Technologies Audiovisuelles en 3D Stéréoscopique*. Rapport d'Expertise Collective.

Artal, Pablo, Ferro, Manuel, Navarro, Rafael, & Miranda, Ismael. 1993. Effects of aging in retinal image quality. *JOSA A*, **10**(7), 1656–1662.

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.

Ball, Karlene K., Beard, Bettina L., Roenker, Daniel L., Miller, Richard L., & Griggs, David S. 1988. Age and visual search : expanding the useful field of view. *JOSA A*, **5**(12), 2210–2219.

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.

Beretta, Giordano. 2000. *Understanding Color*.

Bezzubik, V. V., & Belashenkov, N. R. 2015. Modeling the contrast-sensitivity function of the human visual system. *Journal of Optical Technology*, **82**(10), 711–717.

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.

Blackwell, O. Mortenson, & Blackwell, H. Richard. 1971. IERI : Visual Performance Data for 156 Normal Observers of Various Ages. *Journal of the Illuminating Engineering Society*, **1**(1), 3–13.

Bradley, A.P. 1999. A wavelet visible difference predictor. *IEEE Transactions on Image Processing*, **8**(5), 717–730.

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.

Burkhardt, J.-M. 2003. Réalité virtuelle et ergonomie : quelques apports réciproques. *Le travail humain*, **66**(1), 65–91.

Burkhardt, Jean-Marie, Lourdeaux, Domitile, & Fuchs, Philippe. 1999 (Dec.). Conception d'un système de RV pour la formation des agents de conduite aux opérations en milieu ferroviaire.

- Burt, P., & Adelson, E. 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, **31**(4), 532–540.
- Burton, Kerri B., Owsley, Cynthia, & Sloane, Michael E. 1993. Aging and neural spatial contrast sensitivity : Photopic vision. *Vision Research*, **33**(7), 939–946.
- Cook, Christopher A., Koretz, Jane F., Pfahl, Arnold, Hyun, John, & Kaufman, Paul L. 1994. Aging of the human crystalline lens and anterior segment. *Vision Research*, **34**(22), 2945–2954.
- Daly, Scott J. 1992 (Aug.). Visible differences predictor : an algorithm for the assessment of image fidelity.
- Damelincourt, Jean-Jacques, Zissis, Georges, Corbé, Christian, & Paule, Bernard. 2010. *Eclairage d'intérieur et ambiances visuelles*. Optique & Vision. Paris : Tec&Doc.
- 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.
- Deeley, Robin J., Drasdo, Neville, & Charman, W. Neil. 1991. A simple parametric model of the human ocular modulation transfer function. *Ophthalmic and Physiological Optics*, **11**(1), 91–93.
- 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.
- Dobelle, Wm. H. 2000. Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex :. *ASAIO Journal*, **46**(1), 3–9.
- Dodgson, Neil A. 2004 (May). Variation and Extrema of Human Interpupillary Distance.
- Drago, Frédéric, Martens, William L., Myszkowski, Karol, & Seidel, Hans-Peter. 2003. Perceptual Evaluation of Tone Mapping Operators. *Pages 1–1 of : ACM SIGGRAPH 2003 Sketches & Applications*. SIGGRAPH '03. New York, NY, USA : ACM.
- 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.
- Fairchild, Mark D. 1998. *Color appearance models*. First edition edn. Wiley-IS&T series in imaging science and technology. Chichester, West Sussex : Wiley.
- Fairchild, Mark D. 2005. Human Color Vision. *Pages 1–34 of : Color appearance models*, 2nd ed edn. Wiley-IS&T series in imaging science and technology. Chichester, West Sussex, England ; Hoboken, NJ : J. Wiley.

- 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.
- Fuentese, 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).
- Glassner, Andrew S. 1995. *Principles of Digital Image Synthesis*. The Morgan Kaufmann series in computer graphics and geometric modeling. San Francisco : Morgan Kaufmann.
- Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. 1991. A neurological dissociation between perceiving objects and grasping them. *Nature*, **349**(6305), 154–156.
- 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.
- Habekost, Martin. 2013. Which Color Differencing Equation Should Be Used ? *International Circular of Graphic Education and Research*, 20–33.
- Halonen, L., & Puolakka, M. 2011. CIE and Mesopic Photometry. CIE publication. Vienna : CIE Central Bureau.
- 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.
- Hunt, Robert W. G. 1995. *The Reproduction of Color*. 5th edition edn. Kingston-upon-Thames : Fountain Press.
- James, Thomas W., & Kim, Sunah. 2010. Dorsal and Ventral Cortical Pathways for Visuo-haptic Shape Integration Revealed Using fMRI. Pages 231–250 of : Kaiser, Jochen, & Naumer, Marcus Johannes (eds), *Multisensory Object Perception in the Primate Brain*. New York, NY : Springer New York.
- Judd, Deane Brewster, & Wyszecki, Günter. 1975. *Color in Business, Science, and Industry*. 3d ed edn. Wiley series in pure and applied optics. New York : Wiley.
- Kemeny, Andras. 2014. From driving simulation to virtual reality. ACM Press.
- Kilbride, Paul E., Hutman, Lucinda P., Fishman, Marlene, & Read, John S. 1986. Foveal cone pigment density difference in the aging human eye. *Vision Research*, **26**(2), 321–325.

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.

Le Grand, Yves. 1972. *Optique Physiologique : Lumière et Couleurs*. 2 edn. Vol. 2. Masson & Cie.

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.

Liou, Hwey-Lan, & Brennan, Noel A. 1997. Anatomically Accurate, Finite Model Eye for Optical Modeling. *Journal of the Optical Society of America A*, **14**(8), 1684.

Lubin, Jeffrey. 1995. A Visual Discrimination Model for Imaging System Design and Evaluation. *Pages 245–283 of : Vision Models for Target Detection and Recognition*.

Lukin, A. 2009. Improved visible differences predictor using a complex cortex transform.

Mantiuk, R., Myszkowski, K., & Seidel, H.-P. 2004. Visible difference predictor for high dynamic range images. vol. 3. IEEE.

Mantiuk, Rafal, Daly, Scott J., Myszkowski, Karol, & Seidel, Hans-Peter. 2005 (Mar.). Predicting visible differences in high dynamic range images : model and its calibration.

Mantiuk, Rafal K., & Ramponi, Giovanni. 2015. Human Vision Model Including Age Dependencies.

Mazur, Kim M., & Reising, John M. 1990. The Relative Effectiveness of Three Visual Depth Cues in a Dynamic Air Situation Display. *Proceedings of the Human Factors Society Annual Meeting*, **34**(1), 16–20.

Mehrabi, Mostafa, Peek, Edward M., Wuensche, Burkhard C., & Lutteroth, Christof. 2013. Making 3D work : a classification of visual depth cues, 3D display technologies and their applications. *Pages 91–100 of : Proceeding AUIC '13 Proceedings of the Fourteenth Australasian User Interface Conference*, vol. 139. Melbourne, Australia : ACM.

Michelson, Albert A. 1995. *Studies in optics*. Dover ed edn. New York : Dover Publications.

Mikkola, Maija, Boev, Atanas, & Gotchev, Atanas. 2010. Relative importance of depth cues on portable autostereoscopic display. ACM Press.

Mishkin, Mortimer, Ungerleider, Leslie G., & Macko, Kathleen A. 1983. Object vision and spatial vision : two cortical pathways. *Trends in Neurosciences*, **6**(Jan.), 414–417.

Moon, Parry, & Spencer, Domina Eberle. 1944. Visual Data Applied to Lighting Design. *Journal of the Optical Society of America*, **34**(10), 605.

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.
- Moreau, Guillaume, & Tisseau, Jacques. 2006. *Le Traité de la Réalité Virtuelle. Volume 3. Vol. 3.* Paris : École des mines de Paris. OCLC : 288965251.
- Naka, K. I., & Rushton, W. A. H. 1966. An attempt to analyse colour reception by electrophysiology. *The Journal of Physiology*, **185**(3), 556–586.
- Neveu, Pascaline. 2012. *Impact des Dispositifs de Visualisation en Relief sur les Composantes Oculomotrices d'Accommodation et de Vergence.* Psychologie, Université de Paris 8 - Vincennes - Saint-Denis.
- 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.
- Ogle, Kenneth N. 1968. Optics : An Introduction for Ophthalmologists. *ResearchGate*, Jan.
- Owsley, Cynthia, Sekuler, Robert, & Siemsen, Dennis. 1983. Contrast sensitivity throughout adulthood. *Vision Research*, **23**(7), 689–699.
- Patrick, John. 1992. *Training : Research and practice.* Academic Press.
- Pattanaik, Sumanta N., Ferwerda, James A., Fairchild, Mark D., & Greenberg, Donald P. 1998. A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display.
- Peli, Eli. 1990. Contrast in Complex Images. *Journal of the Optical Society of America A*, **7**(10), 2032.
- Pelli, Denis G., & Bex, Peter. 2013. Measuring contrast sensitivity. *Vision Research*, **90**(Sept.), 10–14.
- 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.
- Pokorny, Joel, Smith, Vivianne C., & Lutze, Margaret. 1987. Aging of the human lens. *Applied Optics*, **26**(8), 1437.
- Rea, M. S. 1987. Toward a Model of Visual Performance : A Review of Methodologies. *Journal of the Illuminating Engineering Society*, **16**(1), 128–142.
- Rea, M. S, & Ouellette, M. J. 1988. Visual performance using reaction times. *Lighting Research and Technology*, **20**(4), 139–153.
- 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.
- Reinhart, William F., Beaton, Robert J., & Snyder, Harry L. 1990. Comparison of depth cues for relative depth judgments. vol. 1256. International Society for Optics and Photonics.
- Robertson, A. R. 1990. Historical development of CIE recommended color difference equations. *Color Research & Application*, **15**(3), 167–170.
- Rogé, Joceline, & Pébayle, Thierry. 2009. Deterioration of the useful visual field with ageing during simulated driving in traffic and its possible consequences for road safety. *Safety Science*, **47**(9), 1271–1276.
- Rogé, Joceline, Pébayle, Thierry, Lambilliotte, Elina, Spitzensetter, Florence, Giselsbrecht, Danièle, & Muzet, Alain. 2004. Influence of age, speed and duration of monotonous driving task in traffic on the driver's useful visual field. *Vision Research*, **44**(23), 2737–2744.
- Rogers, Brian, & Graham, Maureen. 1979. Motion Parallax as an Independent Cue for Depth Perception. *Perception*, **8**(2), 125–134.
- Rose, Albert. 1948. The Sensitivity Performance of the Human Eye on an Absolute Scale\*. *Journal of the Optical Society of America*, **38**(2), 196.
- Rushmeier, H., Ward, G., Piatko, C., Sanders, P., & Rust, B. 1995. Comparing Real and Synthetic Images : Some Ideas About Metrics. 82–91.
- Sagawa, Ken, & Takahashi, Yasuro. 2001. Spectral luminous efficiency as a function of age. *Journal of the Optical Society of America A*, **18**(11), 2659.
- Schanda, János, & International Commission on Illumination (eds). 2007. *Colorimetry : Understanding the CIE System*. [Vienna, Austria] : Hoboken, N.J : CIE/Commission internationale de l'éclairage ; Wiley-Interscience.
- Sekuler, Allison B., & Bennett, Patrick J. 2000. Effects of Aging on the Useful Field of View. *Experimental Aging Research*.
- Sharma, Gaurav, & Bala, Raja. 2013. *Digital color imaging handbook*. OCLC : 856194495.
- Sharma, Gaurav, Wu, Wencheng, & Dalal, Edul N. 2005. The CIEDE2000 color-difference formula : Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, **30**(1), 21–30.
- Spencer, Greg, Shirley, Peter, Zimmerman, Kurt, & Greenberg, Donald P. 1995. Physically-based glare effects for digital images. ACM Press.
- Stoffregen, Thomas A., Bardy, Benoit G., Smart, L. J., & Pagulayan, Randy. 2003. One the Nature and Evaluation of Fidelity in Virtual Environments. In : *Virtual and Adaptive Environments*.
- Surdick, R. Troy, Davis, Elizabeth T., King, Robert A., Corso, Gregory M., Shapiro, Alexander, Hodges, Larry, & Elliot, Kelly. 1994. Relevant Cues for the Visual Perception of Depth : Is Where You See it Where it is? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **38**(19), 1305–1309.

- Surdick, R. Troy, Davis, Elizabeth T., King, Robert A., & Hodges, Larry F. 1997. The Perception of Distance in Simulated Visual Displays :A Comparison of the Effectiveness and Accuracy of Multiple Depth Cues Across Viewing Distances. *Presence : Teleoperators and Virtual Environments*, **6**(5), 513–531.
- Sève, Robert. 2009. *Science de la couleur aspects physiques et perceptifs*. Marseille : Chalagam éd. OCLC : 690444017.
- 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.
- Vos, J.J., & van den Berg, T.J. 1999. *CIE 135 / 1-6 Disability Glare*. Technical Report. CIE.
- Wandell, Brian A. 1995. *Foundations of vision*. Sunderland, Mass : Sinauer Associates.
- Ward, Greg. 1994. A Contrast-Based Scalefactor for Luminance Display. Pages 415–421 of : Heckbert, Paul S. (ed), *Graphics gems IV*. The Graphics gems series. Boston : AP Professional.
- Watson, Andrew B., & Solomon, Joshua A. 1997. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, **14**(9), 2379.
- Watson, Andrew B., & Yellott, John I. 2012. A Unified Formula For Light-adapted Pupil Size. *Investigative Ophthalmology & Visual Science*, **53**(14), 143–143.
- Wen, Senfar. 2006. Display gamut comparison with number of discernible colors. *Journal of Electronic Imaging*, **15**(4), 043001–043001–8.
- Westheimer, Gerald. 1986. The Eye as an Optical Instrument. In : *Handbook of Perception and Human Performance*, vol. 1.
- Winkler, Stefan. 1999. Issues in vision modeling for perceptual video quality assessment. *Signal Processing*, **78**(2), 231–252.
- Winkler, Stefan. 2000 (June). Quality Metric Design : a Closer Look.
- Woods, Andrew J., Docherty, Tom, & Koch, Rolf. 1993 (Sept.). Image distortions in stereoscopic video systems.
- Wyszecki, Günter, & Stiles, W. S. 2000. *Color Science : Concepts and Methods, Quantitative Data, and Formulae*. Wiley classics library ed edn. Wiley classics library. New York : John Wiley & Sons.

## **ANNEXES**



# CALCUL DE LA PERFORMANCE VISUELLE, MÉTHODE DES TEMPS DE RÉACTION

## Performance visuelle générique

Quatre paramètres d'entrée sont nécessaires :

- L'aire de la cible visuelle  $\omega$  en stéradians ( $2.0 \times 10^{-6} \leq \omega \leq 2.8 \times 10^{-3}$ ),
- Le rayon pupillaire  $r$  en millimètres obtenu avec le modèle de Weale :  $2r = 4.77 - 2.44 \tanh[0.3 \log_{10}(L_A)]$ ,
- La luminance d'adaptation  $L_A$  en candela par mètre carré ( $cd/m^2$ ),
- Les luminances relatives à la tache visuelle : luminance du fond d'affichage ( $L_b$ ) et luminance de la cible ( $L_t$ ).

### Calcul du contraste de seuil $C_{t,d}$

$$\begin{cases} A = \log_{10} [\tanh(20000 \omega)] \\ L = \log_{10} \left[ \log_{10} \left( \frac{10 I_R}{\pi} \right) \right] \\ I_R = L_A \pi r^2 \end{cases} \quad (1)$$

$$\log_{10}(C_{t,d}) = -1.36 - 0.179 A - 0.813 L + 0.226 A^2 - 0.0772 L^2 + 0.169 A L \quad (2)$$

## Calcul de la constante de semi-saturation $K$

$$\begin{cases} A* = \log_{10} [\tanh(5000 \omega)] \\ L* = \log_{10} \left[ \tanh \left( \frac{0.04 I_R}{\pi} \right) \right] \end{cases} \quad (3)$$

$$\log_{10}(K) = -1.76 - 0.175 A* - 0.0310 L* + 0.112 A*^2 + 0.171 L*^2 + 0.0622 A* L* \quad (4)$$

## Calcul de la réponse maximale $R_{max}$

$$R_{max} = 0.000196 \log_{10}(I_R) + 0.00270 \quad (5)$$

## Calcul de la performance $R$

$$\begin{cases} C_V = \frac{T|L_b - L_t|}{L_a} \\ \Delta C_d = |C_V - C_{t,d}| \end{cases} \quad (6)$$

$$R = \frac{(\Delta C_d)^{0.97}}{(\Delta C_d)^{0.97} + K^{0.97}} \quad (7)$$

## Ajout de l'influence de l'âge

Avec  $a$  l'âge de l'observateur en années,  $a \geq 20$ .

## Réduction de l'illumination rétinienne

$$\begin{cases} P = 1 - 0.017 (a - 20) \\ 2r = 4.77 - 2.44 \tanh[0.3 \log_{10}(L_a)] \end{cases} \quad (8)$$

$$I'_R = P I_R = P L_A \pi r^2 \quad (9)$$

## Réduction du contraste rétinien

$$\begin{cases} A = \log_{10} [\tanh(20000 \omega)] \\ L = \log_{10} \left[ \log_{10} \left( \frac{10 I_R}{\pi} \right) \right] \\ \epsilon = 1 + \left[ \frac{0.113}{45} (a - 20) \right] \end{cases} \quad (10)$$

$$\log_{10} \left( \frac{C'_{t,d}}{\epsilon} \right) = -1.36 - 0.179 A - 0.813 L + 0.226 A^2 - 0.0772 L^2 + 0.169 A L \quad (11)$$



# MESURES DE LUMINANCE

## Luminance de fond

TABLE 1 – Luminance globale du simulateur en fonction de la nuance de gris affichée.

Gris	Bas-gauche	Haut-gauche	Centre	Bas-droit	Haut-droit	Moyenne
0	0.07	0.07	0.06	0.07	0.07	0.07 cd/m <sup>2</sup>
16	0.13	0.13	0.12	0.13	0.13	0.13 cd/m <sup>2</sup>
32	0.40	0.40	0.39	0.43	0.42	0.41 cd/m <sup>2</sup>
48	1.02	1.01	1.00	1.08	1.04	1.03 cd/m <sup>2</sup>
64	2.06	2.02	2.01	2.22	2.11	2.09 cd/m <sup>2</sup>
80	3.68	3.65	3.56	3.96	3.82	3.73 cd/m <sup>2</sup>
96	5.86	5.55	5.65	5.97	5.90	5.79 cd/m <sup>2</sup>
112	8.44	8.26	8.25	9.06	8.74	5.88 cd/m <sup>2</sup>
128	11.40	11.30	11.30	12.40	12.20	11.72 cd/m <sup>2</sup>
144	14.17	14.53	14.83	16.07	15.87	15.09 cd/m <sup>2</sup>
160	17.73	18.07	18.50	20.17	19.93	18.88 cd/m <sup>2</sup>
176	21.80	22.03	22.77	24.87	24.63	23.22 cd/m <sup>2</sup>
192	26.03	26.37	27.30	29.47	28.93	27.62 cd/m <sup>2</sup>
208	30.27	30.57	31.77	33.80	33.60	32.00 cd/m <sup>2</sup>
224	34.23	34.70	35.93	38.50	38.00	36.27 cd/m <sup>2</sup>
240	37.97	38.47	39.63	42.80	42.13	40.20 cd/m <sup>2</sup>
255	40.53	40.97	42.53	45.53	44.53	42.82 cd/m <sup>2</sup>

## Luminance de cible

TABLE 2 – Luminance de la cible (en  $cd/m^2$ ) en fonction de sa nuance de gris et de la nuance de gris du fond.

Gris	Fond : 0	Fond : 32	Fond : 80	Fond : 128	Fond : 176	Fond : 255
0	0.06	0.22	1.73	5.51	11.10	2030
16	0.09	0.26	1.76	5.54	11.10	20.30
32	0.25	0.40	1.91	5.68	11.30	20.60
48	0.58	0.73	2.23	6.00	11.60	20.80
64	1.13	1.28	2.78	6.55	12.10	21.40
80	1.94	2.11	3.61	7.38	12.93	22.20
96	3.06	3.20	4.70	8.46	14.00	23.30
112	4.46	4.61	6.13	9.87	15.50	24.70
128	6.24	6.40	7.87	11.60	17.20	26.50
144	8.20	8.36	9.88	13.57	19.10	28.40
160	10.30	10.40	11.90	15.70	21.30	30.60
176	12.60	12.80	14.30	18.00	23.20	32.90
192	15.00	15.10	16.67	20.37	25.90	35.20
208	17.40	17.60	19.10	22.80	28.30	37.57
224	19.80	19.90	21.47	25.10	30.70	40.00
240	21.97	22.17	23.63	27.30	32.87	42.07
255	23.37	23.53	25.03	28.77	34.30	43.40

# QUESTIONNAIRES

No dossier \_\_\_\_\_

Date \_\_\_\_\_



QUESTIONNAIRE SUR LA PROPENSIO NÀ L'IMMERSION  
Laboratoire de Cyberpsychologie de l'UQO\*  
(2002)

Indiquer votre réponse en inscrivant un "X" dans la case appropriée de l'échelle en 7 points. Veuillez prendre en compte l'échelle en entier lorsque vous inscrivez vos réponses, surtout lorsque des niveaux intermédiaires sont en jeu. Par exemple, si votre réponse est "une fois ou deux", la deuxième case à partir de la gauche devrait être utilisée. Si votre réponse est "plusieurs fois mais pas extrêmement souvent", alors la sixième case (ou la deuxième à partir de la droite) devrait être utilisée.

1. Devenez-vous facilement et profondément absorbé(e) lorsque vous visionnez des films ou des télérromans?



2. Vous arrive-t-il d'être tellement absorbé(e) dans une émission de télévision ou un livre que les gens autour de vous ont de la difficulté à vous en tirer?



3. Jusqu'à quel point vous sentez-vous mentalement éveillé(e) ou vif(ve) d'esprit en ce moment même?



4. Vous arrive-t-il d'être tellement absorbé(e) dans un film que vous n'êtes pas conscient(e) des choses qui se passent autour de vous?



5. À quelle fréquence vous arrive-t-il de vous identifier intimement avec les personnages d'une histoire?



6. Vous arrive-t-il d'être tellement absorbé(e) dans un jeu vidéo que vous avez l'impression d'être à l'intérieur du jeu plutôt qu'en train de manœuvrer des manettes de jeu et de regarder un écran?



7. Comment vous sentez-vous au plan de la forme physique aujourd'hui?



8. Dans quelle mesure êtes-vous efficace pour vous couper des distractions extérieures lorsque vous êtes en train de faire quelque chose?



9. Lorsque vous assistez à un match sportif, vous arrive-t-il de devenir tellement pris(e) par le match que vous réagissez comme si vous étiez un des joueurs?



10. Vous arrive-t-il d'être tellement absorbé(e) dans des rêveries, en plein jour, que vous n'êtes pas conscient des choses qui se passent autour de vous?



11. Vous arrive-t-il d'avoir des rêves qui semblent tellement réels que vous vous sentez désorienté(e) au réveil?



12. Quand vous faites du sport, vous arrive-t-il d'être à ce point absorbé(e) que vous perdez la notion du temps?



13. Dans quelle mesure arrivez-vous à vous concentrer lors d'activités plaisantes?

PAS DU TOUT	MODÉRÉMENT BIEN	TRÈS BIEN
-------------	-----------------	-----------

14. À quelle fréquence jouez-vous à des jeux vidéos? (SOUVENT devrait correspondre à tous les jours, ou tous les deux jours, en moyenne)

JAMAIS	À L'OCCASION	SOUVENT
--------	--------------	---------

15. Vous est-il déjà arrivé(e) d'être excité(e) lorsque vous visionnez une scène de poursuite ou de combat à la télévision ou dans un film?

JAMAIS	À L'OCCASION	SOUVENT
--------	--------------	---------

16. Vous est-il déjà arrivé(e) d'être apeuré(e) par quelque chose se produisant à la télévision ou dans un film?

JAMAIS	À L'OCCASION	SOUVENT
--------	--------------	---------

17. Vous est-il déjà arrivé(e) de demeurer apeuré(e) longtemps après le visionnement d'un film d'épouvante?

JAMAIS	À L'OCCASION	SOUVENT
--------	--------------	---------

18. Vous arrive-t-il de devenir absorbé(e) à un point tel dans une activité donnée que vous en perdez la notion du temps?

JAMAIS	À L'OCCASION	SOUVENT
--------	--------------	---------

\* Inspiré de la version originale de : Witmer, B.G. & Singer. M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, 7(3), 225-240.

# **Questionnaire sur la Propension à l'Immersion (QPI):**

## Laboratoire de Cyberpsychologie de l'UQO

### **Validation :**

- 94 participants ont rempli le questionnaire avant une immersion virtuelle;
- Les analyses factorielles ont été effectuées en composantes principales et avec des rotations varimax;
- Des corrélations inter-items ont été effectuées;
- Alpha de Cronbach = .78.

### **Cotation :**

Total : items 1 à 18, sur une échelle de 1 à 7.

- Sous-échelle « Focus » : Items 1 + 2 + 3 + 8 + 13
- Sous-échelle « Implication » : Items 4 + 5 + 10 + 12 + 18
- Sous-échelle « Émotions » : Items 11 + 15 + 16 + 17
- Sous-échelle « jeu » : items 6 + 9 + 14

### **Normes :**

	<b>Moyenne</b>	<b>Écart type</b>
<b>Total</b>	64.11	13.11
<b>Sous-échelle « Focus »</b>	24.81	7.54
<b>Sous-échelle « Implication »</b>	15.33	8.67
<b>Sous-échelle « Émotions »</b>	14.25	6.70
<b>Sous-échelle « jeu »</b>	6.56	4.95

Inspiré de la version originale de : Witmer, B.G. & Singer. M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, 7(3), 225-240.



## Questionnaire sur les cybermalaises\*

Laboratoire de Cyberpsychologie de l'UQO  
(Traduit de Kennedy, R.S. et al., 1993)

Numéro \_\_\_\_\_ Date \_\_\_\_\_

Consignes : Encerclez à quel point chaque symptôme ci-dessous vous affecte présentement.

1. Inconfort général	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
2. Fatigue	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
3. Mal de tête	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
4. Fatigue des yeux	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
5. Difficulté à faire le focus	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
6. Augmentation de la salivation	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
7. Transpiration	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
8. Nausées	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
9. Difficulté à se concentrer	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
10. Impression de lourdeur dans la tête	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
11. Vision embrouillée	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
12. Étourdissement les yeux ouverts	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
13. Étourdissement les yeux fermés	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
14. *Vertiges	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
15. **Conscience de l'estomac	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>
16. Rots	<u>Pas du tout</u>	<u>Un peu</u>	<u>Modérément</u>	<u>Sévèrement</u>

\* Les vertiges sont vécus comme une perte de l'orientation par rapport à la position verticale.

\*\* L'expression « conscience de l'estomac » est habituellement utilisée pour désigner un sentiment d'inconfort sans nausée.

Dernière version : Mars 2013

# Questionnaire sur les Cybermalaises

Laboratoire de Cyberpsychologie de l'UQO

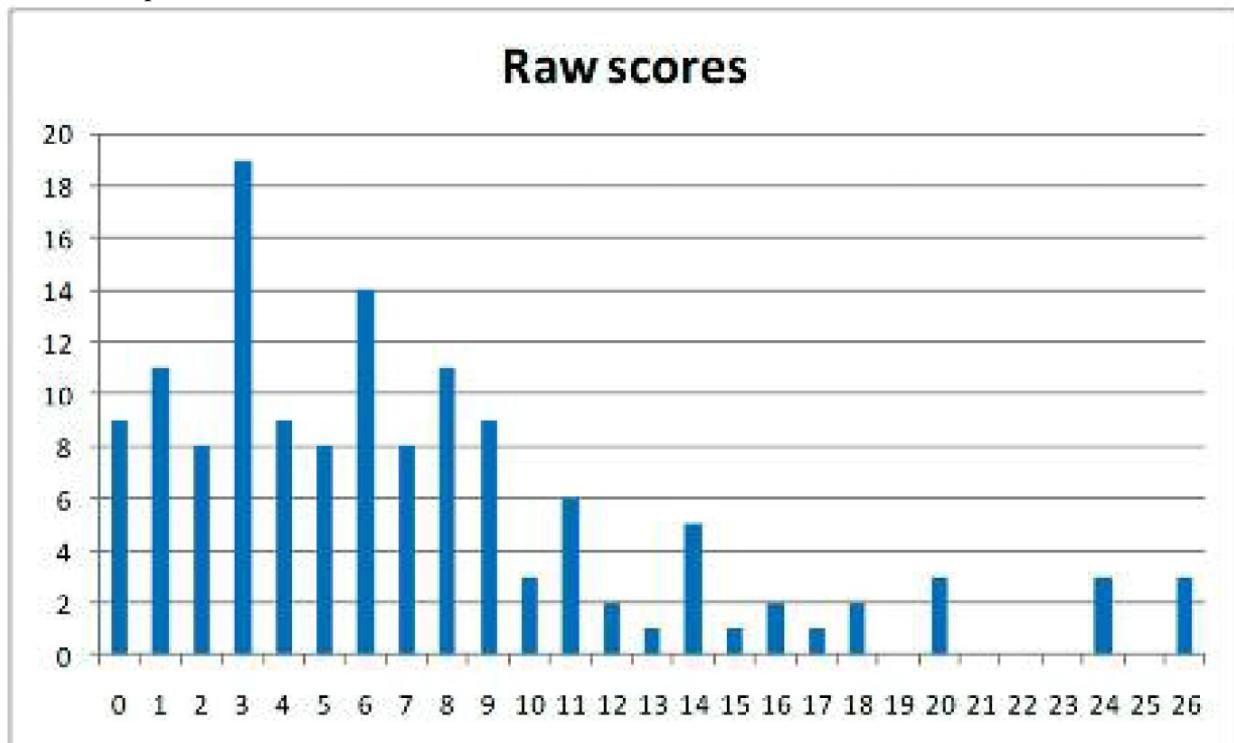
## Cotation de la version canadienne-française du Questionnaire sur les cybermalaises :

- Faire le Total des items 1 à 16 : échelle de 0 (pas du tout) à 3 (sévèrement).
  - Sous-échelle « *Nausée* » : items 1 + 6 + 7 + 8 + 12 + 13 + 14 + 15 + 16.
  - Sous-échelle « *Oculo-moteur* »: items 2 + 3 + 4 + 5 + 9 + 10 + 11.

Pour consulter les articles scientifiques sur la validation canadienne-française de ce questionnaire, veuillez-vous référer aux articles suivants :

- BOUCHARD, S., Robillard, & Renaud, P. (2007). Revising the factor structure of the Simulator Sickness Questionnaire. Acte de colloque du *Annual Review of CyberTherapy and Telemedicine*, 5, 117-122.
- BOUCHARD, S., St-Jacques, J., Renaud, P., & Wiederhold, B.K. (2009). Side effects of immersions in virtual reality for people suffering from anxiety disorders. *Journal of Cybertherapy and Rehabilitation*, 2(2), 127-137.
- BOUCHARD, S., Robillard, G., Renaud, P., & Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side-effects. *Journal of Computer and Information Technology*, 1(3), 20-32.

Basée sur les résultats de l'étude de Bouchard, St-Jacques, Renaud, & Wiederhold (2009), voici la moyenne des scores reportés dans l'étude de validation:



Note. Pour la cotation et validation de la version originale anglaise, veuillez consulter l'article suivant : Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3), 203-220.

No dossier \_\_\_\_\_

Date \_\_\_\_\_



## QUESTIONNAIRE SUR L'ÉTAT DE PRÉSENCE\*

Laboratoire de Cyberpsychologie de l'UQO  
(2002)

Décrivez votre expérience dans l'environnement en marquant d'un "X" la case appropriée de l'échelle en 7 points, et ce en accord avec le contenu de la question et les étiquettes descriptives. Veuillez prendre en compte l'échelle en entier lorsque vous inscrivez vos réponses, surtout lorsque des niveaux intermédiaires sont en jeu. Répondez aux questions indépendamment les unes des autres et dans l'ordre dans lequel ils apparaissent. **Ne sautez pas** de questions et **ne retournez pas** à une question précédente afin de modifier votre réponse.

### EN FONCTION DE L'ENVIRONNEMENT DANS LEQUEL VOUS ÉTIEZ

1. Dans quelle mesure étiez-vous capable de contrôler les événements?



2. Dans quelle mesure l'environnement était-il réactif (sensible) aux actions que vous y faisiez?



3. Dans quelle mesure vos interactions avec l'environnement vous semblaient-elles naturelles?



4. Dans quelle mesure les aspects visuels de l'environnement vous invitaient-ils à vous y impliquer?



5. Dans quelle mesure les mécanismes permettant votre mouvement dans l'environnement vous semblaient-ils naturels?



6. Dans quelle mesure vos sens étaient-ils trompés par le réalisme du mouvement des objets à travers l'espace?



7. Dans quelle mesure les expériences que vous avez vécues dans l'environnement virtuel ressemblaient-elles à celles de l'environnement réel ?



8. Étiez-vous capable d'anticiper les conséquences des mouvements que vous faisiez?



9. Jusqu'à quel point étiez-vous en mesure d'explorer activement l'environnement de façon visuelle?



10. Jusqu'à quel point la sensation de déplacement à l'intérieur de l'environnement virtuel était-elle confondante (réaliste)?



11. À quelle distance pouviez-vous examiner les objets?



12. Jusqu'à quel point pouviez-vous examiner les objets sous différents angles?



13. Jusqu'à quel point étiez-vous impliqué(e) dans l'expérience vécue dans l'environnement virtuel?



14. Jusqu'à quel point avez-vous ressenti un délai séparant vos actions de leurs conséquences?



15. À quel rythme vous êtes-vous adapté(e) à l'expérience vécue dans l'environnement virtuel?



16. En termes d'interactions et de déplacements dans l'environnement virtuel, jusqu'à quel point vous sentiez-vous compétent(e) à la fin de l'expérience?



17. Jusqu'à quel point la qualité visuelle de l'appareillage graphique vous a-t-elle incommodé(e) dans l'exécution des tâches requises?



18. Dans quelle mesure les mécanismes de contrôle de votre mouvement ont-ils interféré avec l'exécution des tâches requises?



19. Jusqu'à quel point êtes-vous parvenu(e) à vous concentrer sur les tâches requises plutôt que sur les mécanismes utilisés pour effectuer lesdites tâches?



RÉPONDEZ SI L'ENVIRONNEMENT VIRTUEL COMPRENAIT DES SONS :

20. Dans quelle mesure les aspects auditifs de l'environnement vous invitaient-ils à vous y impliquer?



21. Dans quelle mesure arriviez-vous à identifier correctement les sons produits dans l'environnement?



22. Dans quelle mesure arriviez-vous à localiser correctement les sons produits dans l'environnement?



RÉPONDEZ SI L'ENVIRONNEMENT VIRTUEL COMPRENAIT DES ASPECTS HAPTIQUE (TOUCHER) :

23. Dans quelle mesure pouvez-vous explorer activement et de façon tactile (par le toucher) l'environnement?



24. Jusqu'à quel point pouvez-vous déplacer ou manipuler les objets dans l'environnement virtuel?



Dernière version: Mars 2013

\*Version originale : Witmer, B.G. & Singer. M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, 7(3), 225-240. Structure factorielle révisée: Witmer, B.J., Jerome, C.J., & Singer, M.J. (2005). The factor structure of the Presence Questionnaire. *Presence*, 14(3) 298-312.

# Questionnaire sur l'État de Présence (QÉP)

Laboratoire de Cyberpsychologie de l'UQO

## Validation :

- 101 participants ont rempli le questionnaire suite une immersion virtuelle;
- Les analyses factorielles ont été effectuées en composantes principales et avec des rotations varimax;
- Des corrélations inter-items ont été effectuées;
- Alpha de Cronbach = .84
- Maintenant 19 items (pour les EV sans sons/toucher) et 24 items (pour les EV avec sons/toucher)

## Cotation :

Total : Items 1 à 19, sur une échelle de 1 à 7.

- Sous-échelle « Réalisme » : Items 3 + 4 + 5 + 6 + 7 + 10 + 13
- Sous-échelle « Possibilité d'agir » : Items 1 + 2 + 8 + 9
- Sous-échelle « Qualité de l'interface » : Items (tous inversés) 14 + 17 + 18
- Sous-échelle « Possibilité d'examiner » : Items 11 + 12 + 19
- Sous-échelle « Auto-évaluation de la performance » : Items 15 + 16
- Sous-échelle « Auditif\* » : Items 20 + 21 + 22
- Sous-échelle « Haptique\* » : Items 23 + 24

\* NOTE : Les sous-échelles « *haptique* » et « *auditif* » ne font pas partie de l'analyse factorielle. Les items peuvent être compilés si l'environnement virtuel contient des sons ou la possibilité de voir son corps ou de toucher des objets. Par contre, les items de la sous-échelle « *résolution* » de Witmer & Singer font maintenant partie de nos analyses.

## Normes :

	Moyenne	Écart type
<b>Total</b>	104.39	18.99
<b>Sous-échelle « Réalisme »</b>	29.45	12.04
<b>Sous-échelle « Possibilité d'agir »</b>	20.76	6.01
<b>Sous-échelle « Qualité de l'interface »</b>	15.37	5.15
<b>Sous-échelle « Possibilité d'examiner »</b>	15.38	4.90
<b>Sous-échelle « Auto-évaluation de la performance »</b>	11.00	2.87

Dernière version: Mars 2013

Version originale : Witmer, B.G. & Singer. M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, 7(3), 225-240. Structure factorielle révisée: Witmer, B.J., Jerome, C.J., & Singer, M.J. (2005). The factor structure of the Presence Questionnaire. *Presence*, 14(3) 298-312.



# PUBLICATIONS

# Realism Score for Immersive Virtual Reality Systems and Driving Simulators

Benoit Perroud<sup>1,2</sup>, Stéphane Régnier<sup>2</sup>, Andras Kemeny<sup>1,2</sup> and Frédéric Mérienne<sup>1</sup>

(1) Arts et Métiers ParisTech, CNRS, Le2i, 2 Rue T. Dumorey 71100 Chalon-sur-Saône, France. E-mail : fredéric.mérienne@ensam.eu

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

**Abstract** - Traditional 2D or 3D quality assessment methods are not sufficient to assess the realism of the outputs of a simulator/immersive virtual reality system. We propose an assessment method based on a scoring system through a new approach. The objective of this paper is to propose a score scale for any simulator or immersive display system that would represent how close to the human visual system the signals that are sent through the display are. Weighted items considered are contrast, acuity, frames per seconds, brightness, field of view and the number of color available.

**Keywords:** Virtual Reality, Realism Assessment, Objective Scoring, Human Visual System

## Introduction

Defining realism can be done by choosing one acceptations out of the different meanings that the word can have in VR [Fuc03] : realistic looking, realistic construction (models), physiologic realism (vision and perception), psychological realism (acceptation) and presence. The definition that was of interest for this study is the one based on realistic physiological parameters (i.e. physiological models of perception of the scene that are matching with the human visual system (HVS) behavior).

Presently, there already exists numerous methods for quality assessment of 2D images/video streams and 3D objective and subjective image quality assessment. Different reviews were done earlier [Moo13b, Moo13a]. But these techniques are not sufficient to deal with quantifying realism : a new method has to be designed. Quality assessment is based on the comparison of a source file with the same file that went through a transformation loop (like video encoding) [Ča04] while VR images are generated in real time.

We hence wanted to develop a score scale, a grade system, that would be transposed from physiological criteria.

## Work Hypotheses

Though the grading system is to be usable in any situation and in any simulator, several hypotheses or guiding rules were made.

First, we divide the different items of the grading system in two types : those for which a behavioral model is already known in the literature, and those for which we found a number of distinct values. The scale can be summarized this way :

- The grade of 0 is given to any value beyond which the graded item is not working, the value is not acceptable.
- The grade of 80 is given to any value which is the standard performance of a perfect observer.
- The grade of 100 is given to any value that is the maximum of the HVS, beyond which the body cannot, physically, perceive differences. All the intermediate values are not always assigned, depending on the values found in the literature.

## Scoring System Proposition

The objective of the study is to have, for any simulator or immersive display system, a score between 0 and 100 that would represent how close to the human visual system the signals that are sent through the display are. It aims to be independent of the content of the virtual scene and independent of the viewer. It is made of 7 weighted items (see as follow) and completed by non-physiological limiting factors in a second time (such as uniformity throughout the simulator). Weighting (based on use-cases) and limiting factors will be addressed later.

### Contrast

Contrast is the difference in brightness/color between the light and the dark parts of an image, of an object.

Though there exists several local or global definitions of contrast, such as Michelson, Weber or Peli, none of them are suitable for complex images [Win99b, Win99a]. We hence fall back on practical values and measurement, such as ANSI contrast (also called simultaneous contrast). The ANSI contrast shows contrast in an approximation of a running simulation. It is measured for a single

screen from a 16 black and white squares pattern. ANSI contrast must be generalized to the whole simulator and hence become System Contrast.

## Acuity (Mono)

Monoscopic acuity is the accuracy of the human eye, its definition. This concept can be directly linked with the size of the pixels in the driving simulators or immersive displays : the smaller the pixel the more accurate the image.

Usually, it is said that the HVS has an acuity between 30" of arc and 2' of arc, with a practical mean of 1' of arc [Fuc03]. Those values can be, in photopic conditions, refined per task [Gro08].

## Acuity (Stereo)

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 [Fuc03, Gro08]. The generally accepted model is  $\Delta r = 0.001r^2$ . Where  $\Delta r$  is the theoretical minimum observable difference in depth (in millimeters) and  $r$  the distance of observation (in meters).

## Frame Per Second (FPS)

First, motion perception starts at 16 FPS. Under this threshold, no movement is perceived, only a suite of distinct images. Secondly, Driscoll et al. [dri78] determined the frequency beyond which flickering would disappear. Critical flicker frequency, for any luminance and riple ratio, is of 70Hz.

Ventral and dorsal pathways is the main theory on how the brain is dealing with visual data. The two pathways are allegedly running at respectively 25 and 200Hz, which is correlated by the fact that one is fast and the other slow, and by the number of cortices the pathways run through [D'h11] ; but no real scientific proven values were found yet.

## Color Space Covering

CIE XYZ was established in 1931 as the a color space representing all the colors that a human eye can see. All the different systems use a fraction of this color space : gamut. Gamut size depends of the three primary colors chosen. Empirically, acceptation begins at Adobe RGB, which covers 52.1% of the CIE space.

## Brightness

The ANSI contrast is a useful practical definition of contrast. But ANSI is not a fully sufficient measurement : a 10000 nits white and a 10 nits black will give the same contrast as a 100 nits white and a 0.1 black.

Based on [Gro08], The eye can accept a very large radiance from  $10^{-6}$  to  $10^5 Cd/m^2$ . Plus, the stimulus of accommodation occurs only above the luminance value of  $0.01 Cd/m^2$ . As it is always harder to achieve because of hardware issues and of the surrounding lights that pollutes it, the black level is to serve as the reference for the brightness level.

## Field of View

Field of view (FOV) is defined as the portion of space that one can see at a given time, without moving the head.

Some of the most common values can be found in [Dev04]. On the azimuth angle, with both eyes, a standard human being is able to see from 170 to 190 degrees ; while on the elevation angle it is only 130 to 145 degrees. Within the azimuth angle, 120 degrees (only) are capable of binocular vision. Reading can be only achieved in a cone of 20 degrees while patterns recognition needs 40 degrees and color discrimination 60 degrees.

## Final Score Model

Finally, a final realism score model is proposed. It is obtained through the sum of all the previously seen items, here referred as  $\sigma_i$ , times the correlated coefficients, here referred as  $\lambda_i$  ; everything being divided by the sum of the coefficients to have the score value fit between 0 and 100. Score is given by the following equation :  $S = \frac{1}{\sum_{i=1}^n \lambda_i} \sum_{i=1}^n (\lambda_i \cdot \sigma_i)$ .

## Conclusion

We propose a scoring system based on a new approach. Such a score system could be useful for quantifying the correlation between a simulator and the HVS, which would lead to the ability, based on the simulators' perks, to choose on which simulator carry out an experiment and to put the results of experiments in perspective. On a more industrial point of view, this scoring system could help in building the specs of new simulators.

## References

- M. Čadík, **Human Perception and Computer Graphics**, in Czech Technical University Postgraduate Study Report, 2004.
- C. Devisme, **Optimisation du Traitement des Indices de Profondeur pour Application à la Vision Artificielle - Cas de la Disparité Horizontale**, Mémoire, 2004.
- F. D'hondt, **Emotion et Espace Visuel : Approche Neuro-magnétique, Neurosomatique et Comportementale**, Ph.D. thesis, 2011.
- The Eyes and Vision**, in W. G. Driscoll, W. Vaughan and Optical Society of America, eds., Handbook of optics, McGraw-Hill, New York, 1978, ISBN 978-0-07-047710-0.
- P. Fuchs, G. Moreau and A. Berthoz, **Le traité de la réalité virtuelle. Volume 1, Volume 1**, Les Presses de l'École des Mines, Paris, 2003, ISBN 978-2-911762-47-5.
- H. Gross, F. Blechinger, B. Achtnar and H. Gross, **Human Eye**, in Survey of Optical Instruments, no. ed. by Herbert Gross ; Vol. 4 in Handbook of Optical Systems, Wiley-VCH, Weinheim, 2008, ISBN 978-3-527-40380-6.
- A. K. Moorthy and A. C. Bovik, **A survey on 3D quality of experience and 3D quality assessment**, 86510M, 2013.
- A. K. Moorthy, C.-C. Su, A. Mittal and A. C. Bovik, **Subjective evaluation of stereoscopic image quality**, *Signal Processing : Image Communication*, vol. 28(8) : 870–883, 2013.
- S. Winkler, **Issues in vision modeling for perceptual video quality assessment**, *Signal Processing*, vol. 78(2) : 231–252, 1999.
- S. Winkler and P. Vandergheynst, **Computing isotropic local contrast from oriented pyramid decompositions**, vol. 4, 420–424, IEEE, 1999, ISBN 978-0-7803-5467-8.



## Model of realism score for immersive VR systems

Benoit Perroud <sup>a,b,\*</sup>, Stéphane Régnier <sup>b</sup>, Andras Kemeny <sup>a,b</sup>, Frédéric Mérienne <sup>a</sup>

<sup>a</sup> Le2i FRE 2005, Arts et Métiers, CNRS, Univ. Bourgogne Franche-Comté, HeSam, Institut Image 2 rue T. Dumorey, 71100 Chalon-sur-Saône, France

<sup>b</sup> VR and Immersive Simulation Center, Renault, Guyancourt, France

### ARTICLE INFO

#### Article history:

Received 12 December 2016

Received in revised form 2 June 2017

Accepted 29 August 2017

Available online xxxx

#### Keywords:

Virtual reality

Realism assessment

Objective score

Human visual system

### 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.

© 2017 Elsevier Ltd. All rights reserved.

## 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, Pattanaik, Shirley, & Greenberg,

\* Corresponding author at: VR and Immersive Simulation Center, Renault, Guyancourt, France.

E-mail addresses: [benoit.perroud@renault.com](mailto:benoit.perroud@renault.com) (B. Perroud), [stephane.regnier@renault.com](mailto:stephane.regnier@renault.com) (S. Régnier), [andras.kemeny@renault.com](mailto:andras.kemeny@renault.com) (A. Kemeny), [frederic.merienne@ensam.eu](mailto:frederic.merienne@ensam.eu) (F. Mérienne).

1996). A more complete definition can be found in [Philippe, Guillaume, and Berthoz \(2003\)](#); 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, Konijn, & Van der Veer, 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, Régnier, Kemeny, & Mérienne, 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, Su, Mittal, & Bovik, 2013](#); [Moorthy & Bovik, 2013](#)) and completed with a perceptual dimension ([Beghdadi, Larabi, Bouzerdoum, & Iftekharuddin, 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 ([Fuentese et al., 2015](#); [Gaelle et al., 2014](#)). 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  $BC^2\alpha^2 = \text{constant}$  with  $B$  the luminance of the object in footlamberts,<sup>1</sup>  $C$  the contrast of the object compared to the background (in %) and  $\alpha$  the visual angle under which the object is seen.
- The definition of an Ideal Observer ([Geisler, 2003](#)) is based on the Signal Detection Theory<sup>2</sup> and the Bayesian theory.<sup>3</sup> 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".

<sup>1</sup> NB:  $1fL \approx 3.426 \text{ cd/m}^2$ .

<sup>2</sup> Ability to discern a specific pattern between different other patterns or noise.

<sup>3</sup> Interpretation of the concept of probabilities as states of knowledge —visible or not— instead of a frequency/propensity of a phenomenon.—

aa

## 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 (Herbert, Fritz, Bertram, & Herbert, 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 or 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, 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

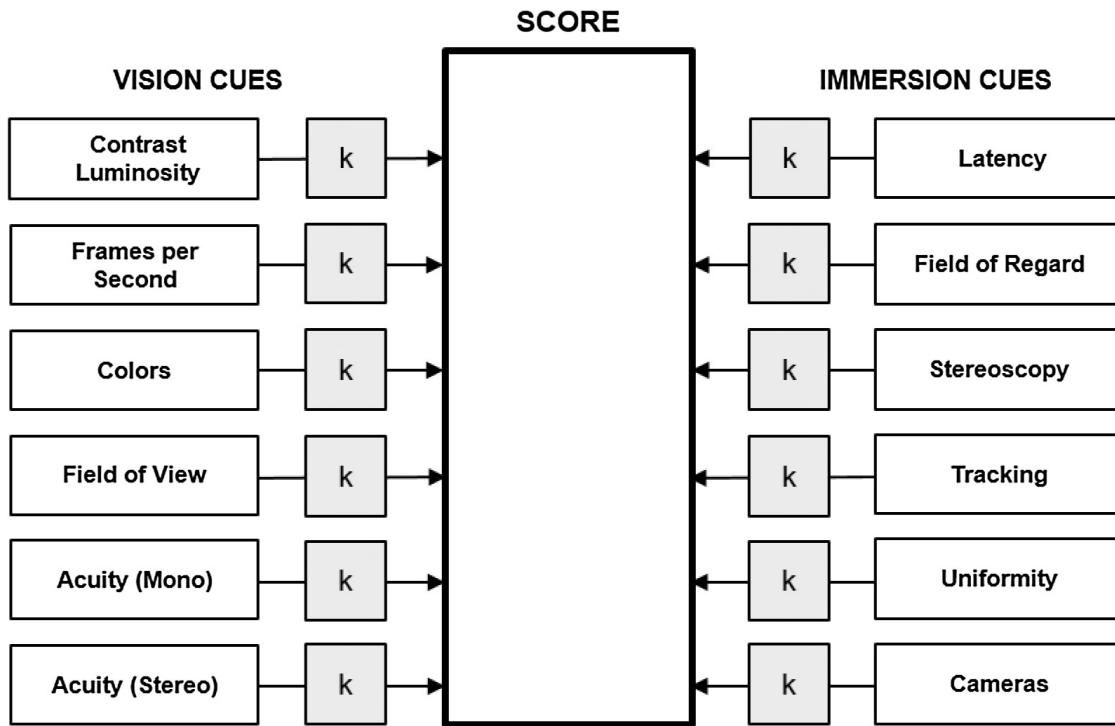
#### 4.1.1. Contrast & luminance

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<sup>4</sup> model (Blackwell, 1981) and the Rea model (Rea, 1986).

<sup>4</sup> Commission Internationale de l'Eclairage.

ab



**Fig. 1.** Overview of the proposed realism score model.

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 (luminosity 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, Connors, & Paradiso, 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 & William, 1978) aggregated the work of Landis (1954) and de Lange Dzn (1958a, 1958b) 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<sup>5</sup>  $r$ . Based on the temporal-modulation transfer-function curves, for all ripple ratios, the critical flicker frequency, at any luminance, is equal to 70 Hz.

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 to shape, color and texture. The two pathways are allegedly running at around respectively 200 Hz and 25 Hz, which is correlated by the fact that one is fast and the other slow (D'hondt, 2011), and by the number of cortices the pathways run through (D'hondt, 2011) and the own distinct latencies of the different cortices (Bullier, 2001); but no real scientific proven values were found yet. US

<sup>5</sup> Ripple ratio:  $r = \frac{\text{Amplitude Of The Fundamental Harmonic}}{\text{Mean Luminance}}$

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.<sup>6</sup>

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.):

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

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.

#### 4.1.3. Visual acuity

**4.1.3.1. 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 (Philippe et al., 2003). Those values can be, in photopic conditions, refined per task (Herbert et al., 2008) (Cf. Table 1).

Vernier resolution can only be achieved in specifics conditions and stereoscopic acuity is to be addressed with another approach that is described later on. Moreover, 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  $\alpha$  being the angle under which the pixel is seen, in *arcmins* (Eq. (2), plot given in Fig. 3.):

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

**4.1.3.2. 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 Herbert et al. (2008) and Philippe et al. (2003). The generally accepted model is as shown in Eq. (3).

$$\Delta r = 0.001r^2 \quad (3)$$

where  $\Delta r$  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 ( $\Delta v_{min}$ ) and the inter-pupillary distance (IPD).

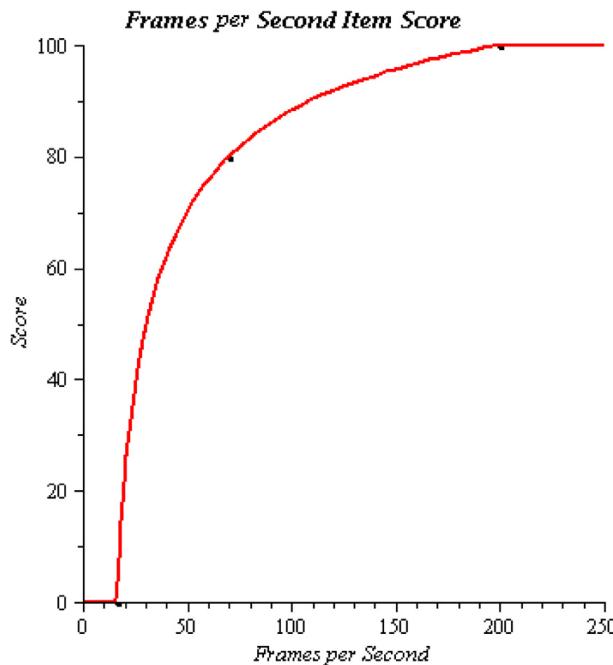
However, these two parameters can vary significantly. Inter-pupillary distance ranges from 52 to 78 mm over the population (Dodgson, 2004) while physiological threshold of vision is dependent on the luminance, especially at very low levels (Herbert et al., 2008). Plus, a general routine to measure the  $\Delta r$  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 Herbert 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.

$$\Delta v_{min} = \frac{d_{IPD} * \Delta r}{r^2} \quad (4)$$

With  $\Delta v_{min}$  the limiting angle for stereoscopic vision,  $d_{IPD}$  the inter-pupillary distance,  $\Delta r$  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 ( $\Delta v_{min}$  taken from the graph in Herbert et al. (2008), in *arcsecs*) and the actual angular resolution the VR display can achieve ( $\alpha$ , in *arcsecs*) (Eq. (5)):

<sup>6</sup> Human Eye Frames Per Second. In AMO.net America's Multimedia Online. Seen at [http://amo.net/nt/02\\_21\\_01fps.html](http://amo.net/nt/02_21_01fps.html).



**Fig. 2.** Plot of the frames per second item score function.

**Table 1**  
Acuity of the eye (Herbert 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"

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

In a CAVE application, the luminance ranges generally from  $10^{-1}$  to  $10^{+1}$  cd/m<sup>2</sup>, which gives a  $\Delta v_{\min}$  average value of 8 arc-secs.<sup>7</sup> 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, acceptance begins at Adobe RGB.

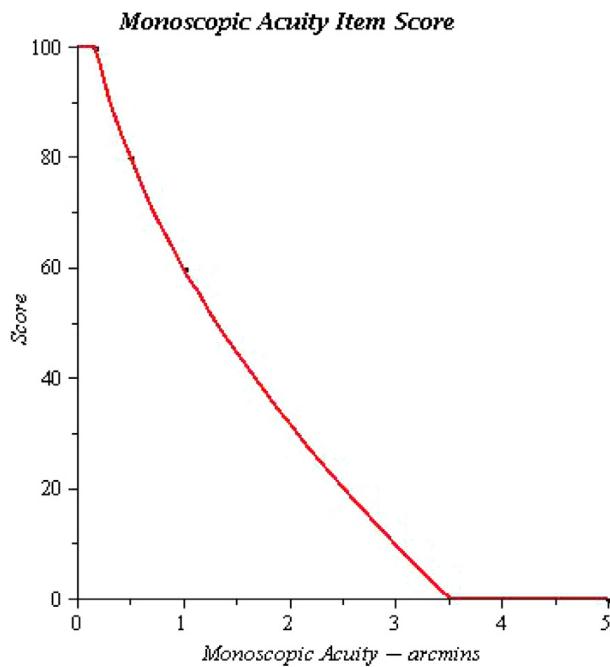
As we could not find any paper on how many colors are enough for the human visual system – even the number of discernible colors is still not precisely determined with values between 100.000 and 10 millions colors,<sup>8</sup> the item proposition is a linear function between the score and the percentage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

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

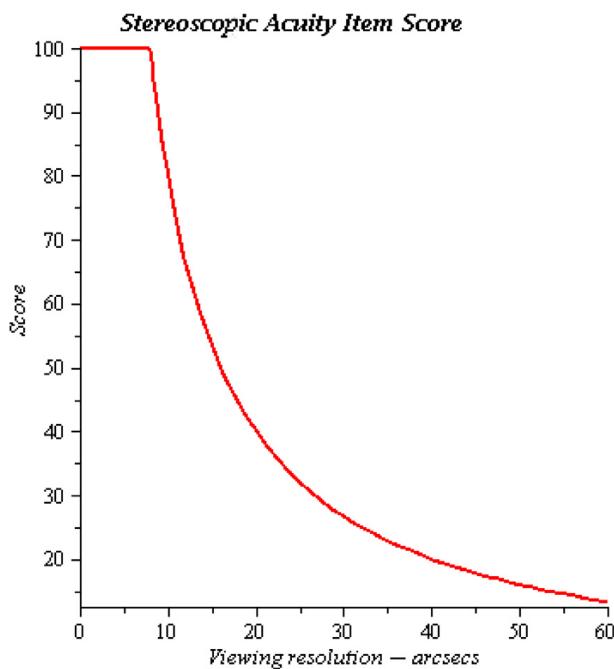
<sup>7</sup> This value was obtained by averaging the value of the limiting stereoscopic angle, in the range of luminance of the CAVE. The evolution of the limiting stereoscopic angle with respect to the luminance can be found p. 67 in Herbert et al. (2008).

<sup>8</sup> Number of Colors Distinguishable by the Human Eye. In *The Physics Factbook*. Seen at <http://hypertextbook.com/facts/2006/JenniferLeong.shtml>.

ae



**Fig. 3.** Plot of the monoscopic acuity item score function.



**Fig. 4.** Plot of the stereoscopic acuity item score function ( $\Delta v_{min} = 8$  arcsecs).

#### 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).

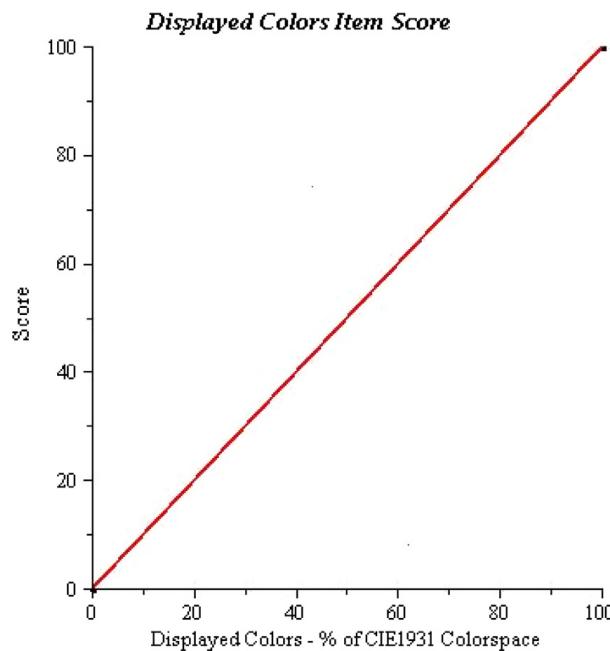
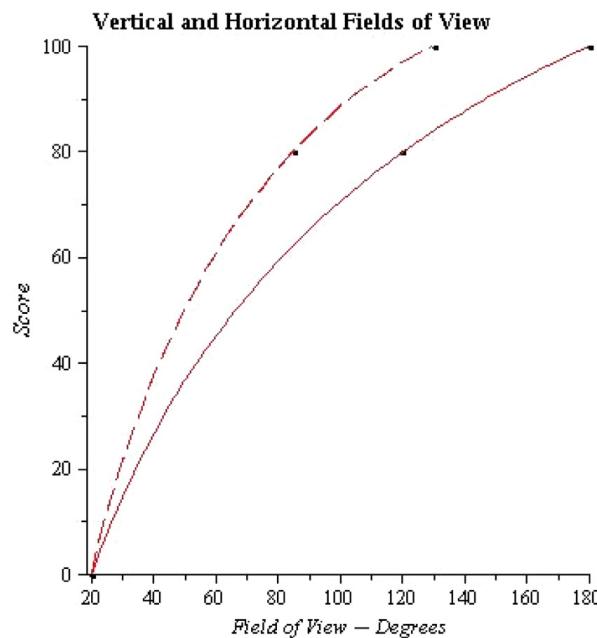
---

af

**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%

**Fig. 5.** Plot of the color item score function.**Fig. 6.** Plot of the field of view item score sub-functions  $F_h$  (solid line) and  $F_v$  (dashed line).

ag

**4.1.5.1. 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°; while on the elevation (vertical) angle it is only 130 to 145°. Within the azimuth angle, 120° (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° portion of space, at the center. Reading can only be achieved in a cone of 20° while patterns recognition needs 40° and color discrimination 60°. Local equation becomes, with  $h$  the value (in degrees) of the horizontal field of view (H-FOV) calculated in the immersive system (Eq. (7)) (see [Fig. 6](#)):

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

**4.1.5.2. Vertical axis.** Three characteristic values can be retained out of the many that one can find in the literature for the vertical field of view (V-FOV). We shall consider the values of lateral vision (130°), induced impression (85°) and lookout (20°) ([Langlois, 2013](#)). Local equation becomes, with  $v$  the value (in degrees) of the vertical field of view (Eq. (8)) (see [Fig. 6](#)):

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

**4.1.5.3. 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° on horizontal axis and 130° on the vertical one) (Eq. (9)).

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

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

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

## 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 ([Han, Xu, Tian, & Zhenjie, 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.

**4.2.2.1. Horizontal and vertical fields.** With respect to the possible movements of the head and of the eye in its orbit (up to 15°), each eye has a field of regard of: over than 200° on temporal side and around 130° on nasal side on the horizontal axis; and 310° on the vertical axis, distributed as +140 and -170° ([Philippe et al., 2003](#)). Hence, on one hand, and since visual fields overlaps, we set the maximum value of H-FOR at 360° ( $h_{max}$ ). V-FOR maximum is set at the maximum physical value, 310° ( $v_{max}$ ). 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:

ah

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

**4.2.2.2. 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° for horizontal axis and 310° for vertical axis. With the same methodology as for FOV, we obtain (Eq. (12)):

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

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

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

#### 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, Docherty, & Koch, 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 than 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 parallel. 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 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.

ai

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 SCANE-R Studio simulation. The simulation environment was a simple riviera tour with different landscapes (sea, lakes, city and mountains).

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 *womengamer*[W + G] sub-group as there was only 1 subject that fitted the category.

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 *framespersecond* 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.

#### 5.4.1. 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 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.

#### 5.4.2. 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.

**Table 3**

Dimensions of the CAVE-like immersive VR system.

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

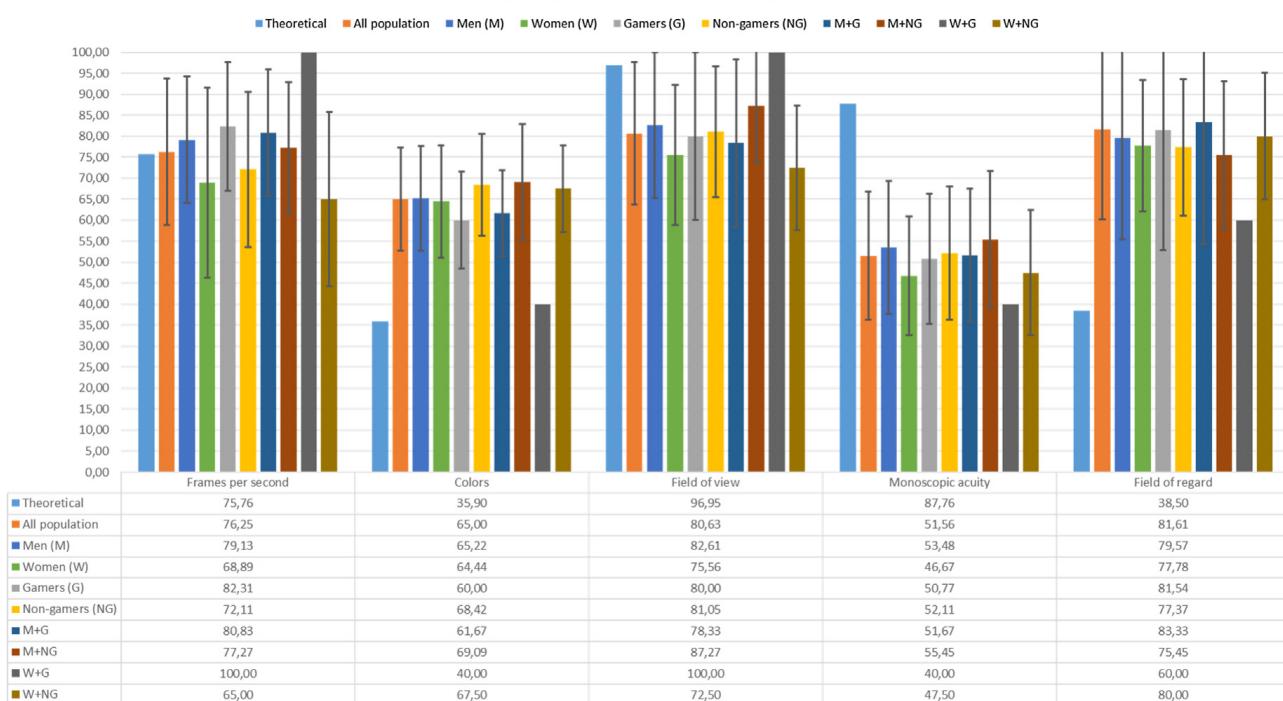
INSUFFICIENT		GOOD			MAX
1	2	3	4	5	
			X		

**Fig. 7.** 1-to-5 aided scale used to subjectively rate the different criteria of the model.**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

"Theoretical physiological model" vs "psychological acceptation"

**Fig. 8.** Results of the theoretical computed values and the subjectively rated values.

ak

## 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 (Herbert 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 through 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 acceptance 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 grades 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.

## References

- Aurat, D. (2016). *Immersion Visuelle Tridimensionnelle Haute Performance par Suivi du Mouvement des Yeux*. Paris, France: Informatique Temps Réel, Robotique & Automatique, Université de Recherche Paris Sciences et Lettres (PSL Research University).
- Bear, M. F., Connors, B. W., & Paradiso, M. 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, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2–3), 96–107.
- Burgess, A. E. (1999). The Rose model, revisited. *Journal of the Optical Society of America*, 16(3), 633–646.
- Čadík, M. (2004). Human perception and computer graphics. In *In Czech Technical University postgraduate study report*.
- Cavonius, C. R., & Robbins, D. O. (1973). Relationships between luminance and visual acuity in the rhesus monkey. *The Journal of Physiology*, 232(2), 239.
- Deering, M.F. (1998). The limits of human vision (Vol. 2).
- 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.
- Devisme, C. (2004). Optimisation du Traitement des Indices de Profondeur pour Application la Vision Artificielle – Cas de la Disparité Horizontale. Mémoire.
- D'hondt, F. (2011). *Emotion et Espace Visuel: Approche Neuromagnétique, Neurosomatique et Comportementale*. Ph.D. thesis.
- Dodgson, N. A. (2004 May). Variation and extrema of human interpupillary distance.
- Driscoll, W. G., William, V., & Optical Society of America (1978). The eyes and vision. In W. G. Driscoll, V. 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*.
- Ferwerda, J. A., Pattanaik, S. N., Shirley, P., & Greenberg, D. P. (1996). A model of visual adaptation for realistic image synthesis. ACM.
- Fuentese, S. F., Rahm, S., Wieser, K., Spillmann, J., Harders, M., Koch, P. P., et al (2015). Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. *Knee Surgery, Sports Traumatology, Arthroscopy*, 23(4), 1077–1085.
- Gaelle, F., Sonia-Yuki, S., Emmanuel, P., Lucile, V., Jean-Luc, D., & Jocelyne, T. (2014). Initial validation of a virtual-reality learning environment for prostate biopsies: Realism matters! *Journal of Endourology*, 28(4), 453–458.

al

- Geisler, W. S. (2003). Ideal observer analysis. *The Visual Neurosciences*, 10(7).
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, 58(6), 460.
- Han, F., Xu, T., Tian, C., & Zhenjie, H. (2010). *Investigation on human visual response latency*. IEEE.
- Herbert, G., Fritz, B., Bertram, A., & Herbert, G. (2008). Human eye. In G. Herbert (Ed.). *Survey of optical instruments. Handbook of optical systems* (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, A. (2009). Driving simulation for virtual testing and perception studies. In: *Proceedings of the DSC 2009 Europe*.
- Kemeny, A. (2014). *From driving simulation to virtual reality*. ACM Press.
- Landis, C. (1954). Determinants of the critical flicker-fusion threshold. *Physiological Reviews*, 34(2), 259–286.
- Langlois, S. (2013). *ADAS HMI using peripheral vision*. ACM Press.
- Millodot, M. (1969). Variation of visual acuity with contact lenses: A function of luminance. *Archives of Ophthalmology*, 82(4), 461.
- Moorthy, A. K., & Bovik, A. C. (2013 March). A survey on 3D quality of experience and 3D quality assessment.
- Moorthy, A. K., Su, C.-C., Mittal, A., & Bovik, A. C. (2013). Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8), 870–883.
- Nichols, B., & Lederman, S. J. (1980). Flicker and motion in film. In T. de Lauretis & S. Heath (Eds.), *The cinematic apparatus* (pp. 96–105). 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, B., Régnier, S., Kemeny, A., & Mérienne, F. (2016). Realism score for immersive virtual reality systems and driving simulators. In *Proceedings of the driving simulation conference DSC 2016 Europe VR* (pp. 213–215).
- Philippe, F., Guillaume, M., & Berthoz, A. (2003). *Le Traité de la Réalité Virtuelle* (Vol. 1) Paris: Les Presses de l'École des Mines.
- Rea, M. S. (1986). Toward a model of visual performance: Foundations and data. *Journal of the Illuminating Engineering Society*, 15(2), 41–57.
- Rea, M. S., & Ouellette, M. J. (1991). Relative visual performance: A basis for application. *Lighting Research and Technology*, 23(3), 135–144.
- Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *Journal of the Optical Society of America*, 38(2), 196.
- Sivak, M. (1996). The information that drivers see: Is it indeed. *Perception*, 25(9), 1081–108.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, Mass: MIT Press.
- Woods, A. J., Docherty, T., & Koch, R. (1993 September). Image distortions in stereoscopic video systems.

# Application of a Relative Visual Performance Model in a Virtual Reality Immersive System

Benoit Perroud, Stéphane Régnier, Andras Kemeny, Frédéric Mérienne

**Abstract**—Visual simulation is increasingly taking part in the creation design of new products, especially in the automotive industry. In order to be a fully trustful tool, simulation must be realistic and thus look for specifications very close to the human vision perks. As part of an evaluation process of user experience realism in a Virtual Reality system, we focus in this paper on one of the core characteristics of vision: the relationship between contrast and luminance. The experiment relates to Rea and Ouellette model and aims to validate its reaction time predictions. The performance is defined as the inverse of the reaction time. The performance hence increases when the reaction time decreases. Our results disclose a major difference in behavior compared to the theoretical predictions. Causes could come from the lighting conditions. In our opinion Rea's RVP model should not be used in the field of Virtual Reality to grade a system's performance with respect to the human vision.

**Index Terms**—Contrast, Luminance, Visual Performance, Virtual Reality

## 1 INTRODUCTION

SIMULATION, through the use of simulators and immersive display systems, becomes a more and more preponderant factor in the development of new products. In the automotive industry, it is 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 [1]. Manufacturers hence need trustful tools in order to make decisions.

The display part of a simulator must be treated with great care and the measure of its quality appears to be essential (as an example, for driving simulators, 90% of the information used by a driver are claimed to be visual, though it is yet to be fully proved numerically [2]). Our goal is to make the simulator achieve one of the five acceptations of realism [3]: the information provided to the sensory system of the body (in our case, the eyes) are the same that one could have in a real life experiment.

This paper focuses on one of the core characteristics of vision: the relationship between contrast and luminance. These two attributes are closely linked and are key factors in the trustfulness of a display. The aim is to be able to describe how much contrast and luminance one needs in Virtual Reality in order for the visual system to work as it would in real life, what is their relationship and how the

human visual system behaves at different levels of contrast and luminance.

## 2 RELATED WORKS

On the one hand, contrast can be roughly described as a ratio between different luminances. Its behavior is most of the time associated to thresholds over which a person is able to distinguish a target in a background and under which the same person cannot distinguish the target. Although the generally accepted threshold value is of 1% of the background luminance, there exists numerous methods to measure the value of the threshold and many applications [4].

On the other hand, luminance is described in photometry as the luminous intensity emerging from a surface and converging to the eye. Based on its value, it can belong to the photopic field ( $L < 0.001 \text{ cd/m}^2$ ), the scotopic field ( $L > 10 \text{ cd/m}^2$ ) or the mesopic field ( $0.001 < L < 10 \text{ cd/m}^2$ ).

Rose [5], in 1948, published one of the first works that tried to evaluate performance on a absolute scale. Rose tied contrast, luminance and size in a single formula (Eq. 1) where  $\alpha$  is the size of the object to be seen,  $L$  the luminance of the object and  $C$  the contrast of the object with respect to the background.  $k$  is a constant member that can be computed with other parameters.

$$\alpha \times L \times C^2 = k \quad (1)$$

While the Rose model gives a very good approximation of a Bayesian ideal observer, it has a limited range of validity. Hence, Burgess [6] revisited it to serve image quality purposes. Progresses were made on the model toward the limitation of the performance by the noise of the input signal.

Though, as it appears that there exists a contrast value beyond which performance diminishes, the single threshold contrast value seems not to be enough. Following this idea,

B. Perroud is with both the Le2i (FRE 2005) lab, *Arts et Métiers*, CNRS, Univ. Bourgogne Franche-Comté, HeSam, Institut Image 2 rue T. Dumorey 71100 Chalon-sur-Saône France and the VR and Immersive Simulation Center, *Renault*, Guyancourt, France

S. Régnier is with the VR and Immersive Simulation Center, *Renault*, Guyancourt, France

A. Kemeny is with both the Le2i (FRE 2005) lab, *Arts et Métiers*, CNRS, Univ. Bourgogne Franche-Comté, HeSam, Institut Image 2 rue T. Dumorey 71100 Chalon-sur-Saône France and the VR and Immersive Simulation Center, *Renault*, Guyancourt, France

F. Mérienne is with the Le2i (FRE 2005) lab, *Arts et Métiers*, CNRS, Univ. Bourgogne Franche-Comté, HeSam, Institut Image 2 rue T. Dumorey 71100 Chalon-sur-Saône France

the International Commission for Eclairage (ICE), through H. R. Blackwell [7], designed a model of visual performance. The approach is to specify a level of performance based on a visual task's lighting parameters such as its luminance and contrast and thus establish task lighting standards. The model first compute a level of visibility in which the task is performed and then rate the overall performance. The entry parameters of the model are the difficulty of the task, the age of the observer and the lighting parameters. Its range is quite consequent as it takes a luminance from 1 to 10000 cd/m<sup>2</sup> but only works for eccentricities lesser than 3 degrees. One main drawback of the model is that it should not be used in interior lighting as light is then mostly directional and specular.

However, the CIE model is only an enhanced contrast threshold model: suprathreshold values are obtained through multiplying the threshold by a constant (the visibility level). That's why Rea [8], [9] proposed a bit later its own model of visual performance (RVP) that provides results similar to other studies measuring suprathreshold behavior. Rea goals were to build a suprathreshold visual performance model that would be independent of task performance and consistent with the literature. His model is based on the compression effect that was theoretized out by Naka and Rushton in 1966 and Lipetz in 1969 as a self-shunting mechanism. The compression effect implies that at some point, despite increasing the magnitude of a stimulus, the sensation magnitude will not increase. The compression effect is described in the following equation (Eq. 2) with  $R$  the response,  $R_{max}$  the maximal response,  $I$  the stimulus intensity,  $n$  an exponent and  $K$  the stimulus intensity producing half of the maximum response.

$$\frac{R}{R_{max}} = \frac{I^n}{I^n + K^n} \quad (2)$$

In a second time, Rea and Ouellette expanded Rea's RVP model with a reaction time experiment [10]. They built a model that can predict the reaction time to performing a vision task. The difference between reaction times in given conditions of luminance and contrast and a reference reaction time can then be linked to a RVP value. Later on, Rea and Ouellette enhanced their model to add the influence of aging [11].

Last but not least, enhancement propositions of Rea's model were made by Kambisch [12] in 1991 and O'donell *et al.* 20 years later [13]. For Kambisch, though the model is a valuable tool based on robust data, the size and contrast treatments have weak theoretical bases. Kambisch replace then the sole contrast measure by three stages modeled on vision models: linear filtering, non-linear filtering and pooling stage. His model does not however consider adaptation luminance. O'donell *et al.*, on their side, showed that the addition of color in the model of RVP can significantly raise the performance level. Luminance and color are both important in the treatment process of the image by the brain. However, the importance of the color is related to the amount of contrast: the lower, the most important the color is. Under contrast values of 0.20, color can be used to help performance, while for contrast values over 0.6 the sole luminance drives the performance.

In our study, we decided to focus on Rea and Ouellette's model, without taking into account enhancement proposed by Kambisch and O'donell *et al.*. The point is to validate the reaction time predictions, through an experimentation, that can be computed through Rea and Ouellette model. The validation or invalidation would allow us to say whether the RVP model is fitted for virtual reality or not.

### 3 APPARATUS

Since we want to evaluate the validity of Rea's model in Virtual Reality (VR), we came with an experiment protocol somewhat close to his. However, some hypotheses and adaptations were necessary (*see below*).

The experiment took place in a 4 faces CAVE-like display system and involved 32 subjects: 23 males and 9 females between the age of 20 and 27 years (mean = 25, sd = 1.8). The population is described in Fig. 2. The CAVE-like display system specifications are given in the following table (Table 1). The subjects were seated in a car seat and positioned for their eyes to be 2 meters away from the front face. They were not wearing any 3D-stereoscopic glasses and the experiment was carried out with monoscopic images. They also had the instruction to keep the head still in the head-support.

Table 1  
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

The whole 4 faces were displaying a unique grey color (the background color). No other lights were lit in the room, the CAVE-like display system screens were hence the only source of light. The background color was varying between 5 values during the experiment in order to change the amount of light arriving to the eyes of the subjects. The values and their associated luminance can be found in Table 2.

Table 2  
Relation between background color and background luminance - Number of trials/contrasts per subject for each background luminance

RGB Code	Luminance	N. of trials
(0,0,0)	0.07 cd/m <sup>2</sup>	13
(32,32,32)	0.41 cd/m <sup>2</sup>	22
(80,80,80)	3.74 cd/m <sup>2</sup>	26
(128,128,128)	11.72 cd/m <sup>2</sup>	26
(176,176,176)	23.18 cd/m <sup>2</sup>	20
(255,255,255)	42.90 cd/m <sup>2</sup>	13

A stimulus shaped as a 2cm-wide disk was appearing at random timings at the center of the front face. The subjects had to press the A button on a Xbox360 controller they held in their hands as soon as they perceived the stimulus. Doing so, we could measure their reaction times. The color of the stimulus was varying over different pre-calculated levels of grey to achieve certain amounts of contrast. The experiment is hence not ecological.

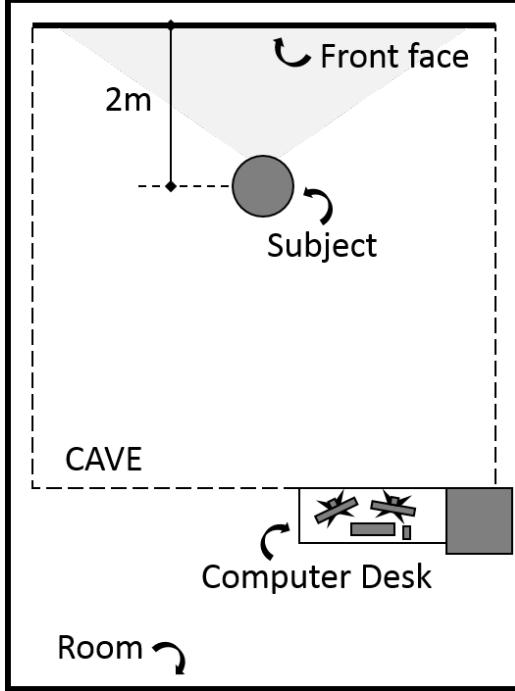


Figure 1. Top view of the implantation of the CAVE-like display system in the room.

In order to be able to predict what levels of gray should be used to display a specific amount of contrast for a given background luminance, we carried out a preliminary set of measures. First, all screens were lit in plain level of gray. Varying the level of gray, we measured the luminance emitted. We hence could plot the transfer function of RGB code toward luminance which equation was found with a least squared regression (Eq. 3). Similarly, on chosen background luminances (those used in the experiment) we displayed the target and made its level of gray vary and took measures of its luminance as well. Another equation, for the luminance of task with respect to its RGB code, was then extracted with a least squared regression (Eq. 4).  $x$  represents the RGB code value which is the same for the R, G and B components as we are only working on levels of gray. The Maple software from Maplesoft was used for the regressions. The measures were made with a Konica-Minolta Chroma Meter CS-100. All the measures and afterwards the experiment were conducted after a 1 hour warm up of the projectors as recommended by the CIE (IEC 61966-6:2005).

$$L_{bck}(x) = -65.5x^4 + 90.1x^3 + 19.7x^2 - 1.3x + 0.1 \quad (3)$$

$$L_{tgt}(x) = -26.1x^4 + 38.3x^3 + 3.7x^2 - 0.03x + 0.02 \quad (4)$$

The contrast is originally computed in Rea's experiment like in Eq. 5. We make the hypothesis that the adaptation luminance ( $L_a$ ), which is the total luminance which arrives to the eye, is equal to the the luminance from the screens (background luminance,  $L_B$ ). In other words, we say that since the screens are the only source of light in the room, the

adaptation luminance is the luminance emerging from the screens.

$$C = \frac{T|L_B - L_T|}{L_a} \quad (5)$$

Rea and Ouellette used a neural filter of transmittance  $T$  to pilot their contrast value: they would reduced the perceived contrast by densifying the neural filter in front of the eye. Plus, they used a small light directed toward the eye with a beam splitter to add on command a veiling light to reduce perceived contrast. Since we pilot our contrast directly on the screen by changing the color of the task depending of the background, we did not use any filter and can then always set the transmittance value to  $T = 1$ . Hence, with both hypotheses, we came with our own contrast equation (Eq. 6). Both equations only works when the background is more luminous than the task ( $L_B \geq L_T$ ) allowing to compute contrast values by lower value only. We hence use a slightly modified version of the equation to compute contrast values by greater value (Eq. 6).

$$C = \frac{|L_B - L_T|}{L_B} \text{ and } C = \frac{|L_B - L_T|}{L_T} \quad (6)$$

The subject was first invited to sit in the car seat and then introduced to his task (stimulus detection and reaction time measurement). The experiment was unfolding as follow: a background luminance was set. When the subject felt ready, s-he could launch the apparitions of the targets by pressing another button on the controller. When the sequence was launched, targets appeared successively with a random time between each. The target contrast relative to the background were specifically chosen to produce a desired contrast. The subject had then to wait for 3 to 5 minutes so that his/her eyes accommodate to the new luminance. The targets level of gray was chosen with great care to achieve precise amount of contrast: 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90 and 1. The contrast values were achieved, when possible, by target color brighter and darker than the background. Hence, the number of trial per luminance is different as for some background luminances some contrast values (either by brighter or darker value) were not possible. For example, when displaying a full black color (RGB Code = (0,0,0)) on the screens, it was not possible to have contrast values for the target by darker levels of gray. The order of contrast appearance was different for every subject as we used the Fisher-Yates shuffle to randomly sort it. When finishing a sequence at a given background luminance, the background luminance was changed to another one and the same process (wait for 3 to 5 minutes and then launch sequence when ready) was repeated. Similarly, the order of background luminance was randomly chosen for every subjects. The subjects got 6 non-measured trials before the beginning of the experiment to get used to it. The 6 trials were done at a luminance that would no be used during the normal experiment.

For a better understanding, the flowchart of the unfolding of the whole experiment can be found in Fig. 3. The "TARGETS" sub-flowchart that manages the apparition of the targets for a given background is detailed in Fig. 4. For each of the 6 background luminances (see Table 2), there is a

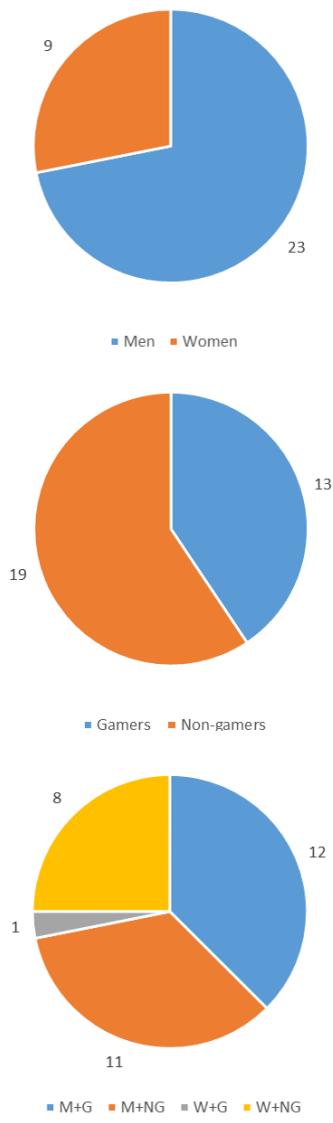


Figure 2. Distribution of the population between men (M), women (W), whether they consider themselves gamer (G) or not (NG) and the combination of the two.

different number of displayable contrasts values (see Table 2).

## 4 RESULTS

As we want to compare theoretical and measured reaction times, we first computed the predictions of Rea and Ouellette's model. These values can be found in Fig. 5. The values out of range of the model (which predicts that one should not be able to distinguish the target over the background and hence does not provide a reaction time value) are listed with an "x".

Some theoretical values such as 25432 ms or 10865 ms of predicted reaction time can seem rather important. Those come from a division in the model where the divider is very close to 0 (at least much closer than in the other cases) and hence the result races up very quickly.

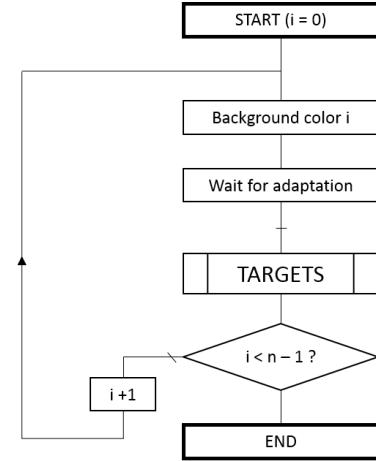


Figure 3. Flowchart of the unfolding of the experiment for a subject. The TARGETS sub-flowchart is described in an other figure.

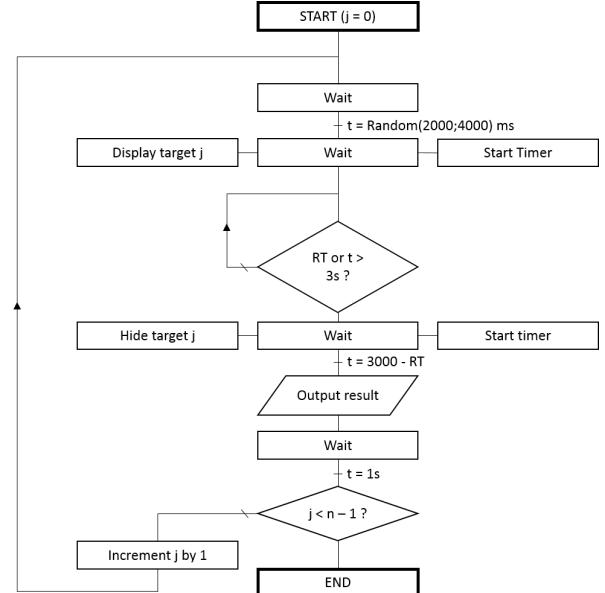


Figure 4. Flowchart of the unfolding of a sequence of targets appearances, for a given background luminance.

The averaged measured reaction times values of the 32 subjects are listed in Fig. 6 with their corresponding standard deviations. Since the population is over 30 subjects, the Pearson standard deviation was used.

Measured values range from 371 ms to 670 ms with a mean value (of the means, excluding the diverging value, *see afterwards*) of 437 ms ( $SD = 53$  ms) and a mean standard deviation of 96 ms ( $SD = 82$  ms). A particular value (2013 ms for a contrast of 0.05 and a background luminance of  $0.41\text{ cd/m}^2$ ) can be highlighted.

The performance R is defined as the inverse of the reaction time RT:  $R = 1/RT$ . It is used instead of the reaction time to characterize the performance of a subject. The performance hence increases when the reaction time decreases and decreases when the reaction time increases. Fig. 7 shows an example of the graph of both theoretical and measured performances in the reaction time experiment. The example

Theoretical predictions (in ms)						
C\L	0.07	0.41	3.74	11.72	23.18	42.9
0.05	x	x	x	x	x	x
0.1	x	x	x	x	x	x
0.15	x	x	x	x	10865	3594
0.2	x	x	25435	1924	1429	1212
0.25	x	x	1686	1041	912	839
0.3	x	x	1039	788	725	685
0.4	x	2663	696	596	568	549
0.5	x	1272	577	517	499	486
0.6	x	929	517	473	459	449
0.7	x	773	481	446	434	425
0.8	4451	684	456	426	416	409
0.9	2448	626	439	412	403	396
1	1770	585	425	402	393	387

L in cd/m<sup>2</sup>  
C clamped between 0 and 1, by definition  
x not distinguishable, below the threshold

Figure 5. Theoretical predictions for reaction times through Rea and Ouellette model.

is taken for a background luminance of 23.18 cd/m<sup>2</sup>. The x-axis represents the different values of contrast, clamped by definition between 0 and 1. The y-axis represents the performance R, inverse of the reaction time (in ms<sup>-1</sup>). The two first values of the predicted performance are below the threshold and out of range of the model. They are hence set to 0 which would be equivalent to an infinite reaction time and thus a lack of reaction.

## 5 DISCUSSION

The first outcome that strikes is while Rea's model predicts some task that should not be seen (values below the threshold and out of the range of the model), the subjects saw all the tasks that were proposed to them, at all the lighting (background luminance and task contrast) conditions.

Secondly, there exists a major difference in behavior between theoretical predictions and subjects measurements. When the performance of the first one, at a given background luminance, increases with the contrast the latter is always sort of constant. The subjects had a similar performance on all the targets. Their reaction times were somewhat identical in lighting conditions however very different. Following the predictions of the model of Rea and Ouellette our subjects should not have been that potent. They should first not be able to see the target and then, when the contrast increases, become better and better.

At the end of the day, our results do not match at all with the predictions of Rea and Ouellette's model but still converge toward it for high contrast values. We tried our best to translate Rea and Ouellette's experiment to Virtual Reality. Though there are still fundamental differences that exist and that may explain the differences between predictions and measurements.

The major difference is the quantity of light and the quantity of light sources. In the Rea and Ouellette experiment, the only sources of light were a little screen 1.68 m away from the user and a light that is directed toward the eye through a beam splitter. On top of that, the quantity of light was controlled with a neural density filter in front

Average measured reaction time (in ms)						
C\L	0.07	0.41	3.74	11.72	23.18	42.9
0.05	465	2013	670	572	500	450
0.1	474	663	450	421	434	406
0.15	474	553	439	421	434	404
0.2	468	531	418	422	407	432
0.25	465	479	415	401	414	398
0.3	475	507	423	399	412	393
0.4	443	457	451	400	406	396
0.5	440	461	418	409	394	415
0.6	440	451	405	396	403	402
0.7	466	450	416	394	408	390
0.8	432	443	402	405	394	398
0.9	464	440	404	402	400	399
1	448	426	403	401	398	371

Standard deviation of average measured reaction time						
C\L	0.07	0.41	3.74	11.72	23.18	42.9
0.05	58	429	112	314	290	155
0.1	98	447	86	66	71	50
0.15	85	286	78	85	99	48
0.2	67	323	48	87	75	88
0.25	158	63	68	50	62	92
0.3	141	215	73	53	63	72
0.4	74	59	221	59	63	52
0.5	44	78	71	62	42	67
0.6	58	66	52	52	60	81
0.7	94	80	72	49	57	48
0.8	68	67	46	80	48	67
0.9	79	61	58	61	64	77
1	152	41	52	77	64	48

L in cd/m<sup>2</sup>  
C clamped between 0 and 1, by definition

Figure 6. Average and standard deviation of all reaction times for all lighting and contrast conditions.

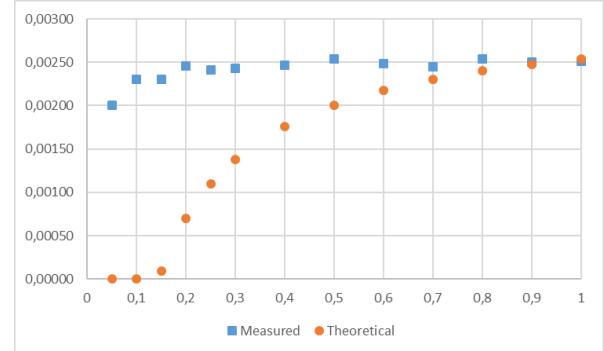


Figure 7. Graph of the predicted performances and the real measured performances in the reaction time task for a background luminance of 23.18 cd/m<sup>2</sup>, viewing distance of 2 meters and target size of 2 cm.

of the eye. On our side, the light come from giant screens surrounding the user with multiple indirect sources of light: projectors emit light towards the screens. No light-s were directly pointed toward the eye from a short distance and no neural density filter was used as we could more easily pilot the contrast and background luminance with the projectors only. Plus, Rea and Ouellette used a filter that is not directly fitted on the eye and hence let pass small amounts of light around it to the eye.

Hence, given the comparison between our results and the predictions of the model, in our opinion, Rea and Ouellette's model should not be used in Virtual Reality to determine the performance of the eye (with respect to the model of human vision) based on luminance and contrast.

## 6 CONCLUSION

The aim of the study was to be able to describe how much contrast and luminance one needs in Virtual Reality. Models already exists in the literature but none are applied to the specific field of Virtual Reality. The goal was to apply one of these models (Rea and Ouellette) in a VR context. An experiment similar to the one in Rea and Ouellette's papers was conducted in a CAVE-like immersive display system. We showed that, in our opinion, the RVP model is not fitted for Virtual Reality. Our results show a very distinct and opposite behavior from the predictions of the model. Reasons seem to come from the lighting conditions of Virtual Reality that are very different in quantity and nature from the conditions in which Rea and Ouellette conducted their own experiments.

## REFERENCES

- [1] A. Kemeny, "Driving Simulation for Virtual Testing and Perception Studies," in *Proceedings of the DSC 2009 Europe*, Monaco, 2009.
- [2] M. Sivak, "The Information That Drivers See: Is It Indeed 90% Visual ?" *Perception*, vol. 25, no. 9, pp. 1081–1089, 1996. [Online]. Available: <http://pec.sagepub.com/lookup/doi/10.1088/p251081>
- [3] P. Fuchs, G. Moreau, and A. Berthoz, *Le Traité de la Réalité Virtuelle. Volume 1*. Paris: Les Presses de l'École des Mines, 2003, vol. 1.
- [4] D. G. Pelli and P. Bex, "Measuring contrast sensitivity," *Vision Research*, vol. 90, pp. 10–14, Sep. 2013. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0042698913001132>
- [5] A. Rose, "The Sensitivity Performance of the Human Eye on an Absolute Scale\*," *Journal of the Optical Society of America*, vol. 38, no. 2, p. 196, Feb. 1948. [Online]. Available: <https://www.osapublishing.org/abstract.cfm?URI=josa-38-2-196>
- [6] A. E. Burgess, "The Rose model, revisited," *Journal of the Optical Society of America*, vol. 16, no. 3, pp. 633–646, 1999.
- [7] H. R. Blackwell, *An Analytic model for describing the influence of lighting parameters upon visual performance*, ser. Publication CIE, International Commission on Illumination, Ed. Paris, France: Bureau central de la CIE, 1981, no. no. 19/2.1-19/2.2.
- [8] M. S. Rea, "Toward a Model of Visual Performance: Foundations and Data," *Journal of the Illuminating Engineering Society*, vol. 15, no. 2, pp. 41–57, Jul. 1986. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00994480.1986.10748655>
- [9] ——, "Toward a Model of Visual Performance: A Review of Methodologies," *Journal of the Illuminating Engineering Society*, vol. 16, no. 1, pp. 128–142, Jan. 1987. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00994480.1987.10748673>
- [10] M. S. Rea and M. J. Ouellette, "Visual performance using reaction times," *Lighting Research and Technology*, vol. 20, no. 4, pp. 139–153, Jan. 1988. [Online]. Available: <http://lrt.sagepub.com/cgi/doi/10.1177/096032718802000401>
- [11] ——, "Relative visual performance: A basis for application," *Lighting Research and Technology*, vol. 23, no. 3, pp. 135–144, Sep. 1991. [Online]. Available: <http://lrt.sagepub.com/content/23/3/135>
- [12] D. Kambich, "An Alternative Relative Visual Performance Model," *Journal of the Illuminating Engineering Society*, vol. 20, no. 1, pp. 19–27, Jan. 1991. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00994480.1991.10748918>
- [13] B. O'Donell, E. Colombo, and P. Boyce, "Colour information improves relative visual performance," *Lighting Research and Technology*, vol. 43, no. 4, pp. 423–438, Dec. 2011. [Online]. Available: <http://lrt.sagepub.com/cgi/doi/10.1177/1477153511412448>

**Benoit Perroud** is a PhD Student working with both Arts et Métiers ParisTech -inside the Le2i Laboratory- and the Advanced Driving Simulation and Virtual Reality department of the Renault Group. He received a Master Degree in Engineering and a Research Master Dgreee in Digital Mockup and 3D Visualization from Arts et Métiers ParisTech in 2015. He is intesrested in realistic vision behavior in virtual reality and vision models.

**Stéphane Régnier** Stéphane Régnier is currently Expert Virtual Mockup for Engineering and Style departments at Renault. He received an engineer's degree from Institut National des Sciences Appliquées in Lyon, in 1990. He has been involved in various visualisation software developments, simulation tools, Virtual Reality since the early 90's, in various departments at Renault, R&D, Style, Engineering. He has managed several R&D teams, and has a large experience of Virtual Reality, immersive projection systems and 3D visualisation solutions.

**Andras Kemeny** is the Director of the Laboratory of Immersive Visualization, a joint Renault - Arts et Métiers ParisTech research laboratory. He has initiated and directed the development of a driving simulation software package, SCANeR Studio®, a worldwide leader and a major development and assessment driver in the loop (DIL) simulation tool for Automotive Driver Assistance Systems (ADAS), Human Machine Interfaces (HMI), autonomous and connected vehicles, deployed at a large number of automotive OEMs, suppliers and research laboratories. A. Kemeny is also Associate Professor at Arts et Métiers ParisTech and author of more than 150 scientific and technical papers and book chapters as well as several industrial patents and software copyrights.

**Frédéric Merienne** is currently professor at Arts et Métiers and director of the Image Institute, a research team of the Le2i Laboratory, since 2004. He received his PhD in 1996 at the National Polytechnical Institute of Grenoble (INPG). His research interests are focused on virtual immersion linked with engineering, cultural heritage and health applications. He is the author of many scientific papers in virtual reality and related disciplines. He is involved in different projects with industrial partners, and initiated international collaborative projects in the area of virtual reality with universities in the USA, Australia, Colombia and Malaysia.