



Titre: Ajustement d'affichage stéréoscopique par évaluation du point de
Title: regard 3D dans un environnement virtuel

Auteur: Jessica Conti
Author:

Date: 2014

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Conti, J. (2014). Ajustement d'affichage stéréoscopique par évaluation du point de
Citation: regard 3D dans un environnement virtuel [Mémoire de maîtrise, École
Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/1567/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1567/>
PolyPublie URL:

**Directeurs de
recherche:** Éric Paquette, & Benoît Ozell
Advisors:

Programme: Génie informatique
Program:

UNIVERSITÉ DE MONTRÉAL

AJUSTEMENT D’AFFICHAGE STÉRÉOSCOPIQUE PAR ÉVALUATION DU POINT
DE REGARD 3D DANS UN ENVIRONNEMENT VIRTUEL

JESSICA CONTI
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L’OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
OCTOBRE 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

AJUSTEMENT D’AFFICHAGE STÉRÉOSCOPIQUE PAR ÉVALUATION DU POINT
DE REGARD 3D DANS UN ENVIRONNEMENT VIRTUEL

présenté par : CONTI Jessica

en vue de l’obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d’examen constitué de :

Mme CHERIET Farida, Ph.D., présidente

M. OZELL Benoît, Ph.D., membre et directeur de recherche

M. PAQUETTE Eric, Ph.D., membre et codirecteur de recherche

M. RENAUD Patrice, Ph.D., membre

*À la vie,
À toutes mes lumières dans la nuit*

REMERCIEMENTS

Ce projet de recherche a pu être mené à terme grâce au soutien technique, financier et moral de personnes et d'institutions que je tiens à remercier chaleureusement.

Je souhaite tout d'abord exprimer ma profonde gratitude envers les professeurs Benoît OZELL et Eric PAQUETTE, pour m'avoir offert l'opportunité d'entreprendre cette étude et m'avoir encadrée au long de son développement. Leurs conseils, commentaires et questions furent de précieux guides pour aborder les domaines de la réalité virtuelle et de la stéréoscopie.

Je remercie également le professeur Patrice RENAUD, dont l'expertise concernant l'application de la réalité virtuelle à la psychiatrie légale m'a aidé à orienter le protocole expérimental et l'analyse des résultats afin d'en tirer des conclusions plus pertinentes.

Ma profonde reconnaissance va à Tarik BOUKHALFI pour son assistance technique précieuse au moment du développement logiciel et de la prise en main des ETG, son aide et son dévouement au cours des expérimentations, ainsi que son soutien moral sans faille.

Je tiens à exprimer toute ma gratitude envers la professeure Farida CHERIET pour avoir accepté de juger mes travaux de recherche.

Un grand merci à tous les volontaires ayant participé à l'expérimentation, pour leur disponibilité et leur sérieux. Merci également à Laurent TREMBLAY et Jérôme COLLIN, qui nous ont aimablement prêté le robot utilisé au cours des expérimentations, à Fannie CARRIER, qui a réalisé le dé-bruitage de certaines données expérimentales, ainsi qu'à tout le personnel de l'Institut Philippe-Pinel pour son amabilité.

Je souhaite remercier chaleureusement le ministère de l'Économie, de l'Innovation et des Exportations, dont le financement a permis à ce projet d'exister, ainsi que l'École Polytechnique de Montréal, l'École de technologie supérieure et l'Institut Philippe-Pinel de Montréal, dont la collaboration en est à l'origine. Ma reconnaissance sincère s'adresse également à leurs comités d'éthiques de la recherche, en particulier à mesdames Sonya AUDY, Marie-Lou MOREAU et Anne-Marie BOURRET pour leur aide et leur efficacité.

Je tiens à exprimer ma gratitude envers l'École Polytechnique (France), et en particulier la Direction des relations internationales, pour m'avoir offert la possibilité de ce double diplôme avec l'École Polytechnique de Montréal.

À titre plus personnel, je remercie enfin les personnes qui m'ont aidée et soutenue tout au long de cette maîtrise. Florian tout d'abord, pour sa gentillesse, sa patience et son soutien moral, mais aussi son expertise en Matlab et ses relectures. Renaud et Lise, pour leur présence bienveillante, ma famille et mes amis qui m'ont encouragée au cours de ces deux années, de près ou de loin.

RÉSUMÉ

Ce projet de recherche se déroule au sein d'une voûte immersive de réalité virtuelle. Un tel environnement plonge l'observateur au cœur d'un monde généré par ordinateur et projeté sur 4 faces d'un cube d'environ 3 m de côté au milieu duquel il se tient. Néanmoins, l'affichage stéréoscopique provoque généralement de l'inconfort oculaire, voire de la fatigue visuelle. Le but poursuivi était d'améliorer l'expérience immersive en tâchant de diminuer la gêne visuelle ressentie, tout en rendant le monde plus réaliste et en renforçant le sentiment de présence.

L'objectif fut donc défini de la manière suivante : concevoir et évaluer une méthode d'ajustement des paramètres de configuration du système de caméras, afin de trouver des valeurs optimales et vraisemblablement personnalisées pour chaque individu. Nos hypothèses soutenaient qu'ajuster ces valeurs en fonction de la position du point de regard permettrait de rendre la configuration plus performante. Des lunettes intégrant des caméras dirigées vers les yeux ont été utilisées comme système de suivi oculaire afin de calculer en temps réel le point de regard de l'observateur dans la scène. Cette donnée a ensuite servi à ajuster deux paramètres de la configuration du système de caméras : la distance séparant les deux caméras virtuelles utilisées pour obtenir le rendu stéréoscopique de la scène, et une profondeur de champ artificielle générée autour du point de regard. La connaissance de la position de ce dernier, et plus particulièrement des objets qui ont retenu l'attention de l'individu, fournit également des indices sur le traitement cognitif de l'information visuelle, c'est-à-dire la manière dont il a perçu et inspecté la scène.

La première étape méthodologique a consisté à développer les outils qui permettent d'intégrer le système de suivi oculaire à la voûte de réalité virtuelle de l'Institut Philippe-Pinel de Montréal, de déterminer le point de regard grâce à une méthode de *raycasting* couplée à une procédure de calibration, et enfin de modifier les paramètres de la configuration en fonction.

Subséquentement, une expérimentation a été conduite auprès de participants pour évaluer notre proposition par comparaison avec cinq autres configurations plus neutres ou classiques, intégrant un aspect éthique au projet. Les 6 configurations testées se différencient par la présence ou non d'un flou de profondeur de champ et la valeur de la distance inter-caméras (DIC) : fixe et choisie par le participant, fixe et correspondant à sa distance interpupillaire anatomique (DIP) ou variable en fonction de la position du point de regard. En dehors des méthodes de comparaison courantes répertoriées au cours de la revue de littérature, nous avons cherché une technique qui fournirait une vraie indication sur le comportement oculaire

de l’observateur au cours d’une séance d’immersion. Pour cela, nous avons mis au point un test basé sur la reproduction d’une scène réelle en virtuel.

Les 18 participants ont été recrutés sur la base du volontariat. Afin de limiter les risques de fatigue visuelle au cours de l’immersion, ceux-ci devaient être accoutumés à la réalité virtuelle et la stéréoscopie, c’est-à-dire avoir déjà joué à des jeux vidéos et assisté à des projections de films en 3D sans gêne. Ces participants ont dû effectuer différentes tâches, chacune associée à un objectif. Tout d’abord, ils devaient naviguer dans un environnement virtuel riche – un appartement comportant de nombreux objets – successivement avec les différentes configurations, et noter chacune selon le confort ressenti, le réalisme du rendu, le plaisir de navigation, la perception des distances et des profondeurs, et enfin le sentiment de présence dans le monde virtuel. Ils durent par la suite déterminer leurs limites de fusion en approchant ou reculant un objet virtuel. Celles-ci constituent une comparaison objective d’efficacité du rendu entre les différentes configurations, la diplopie générée par des difficultés de fusion des images stéréoscopiques étant l’une des sources principales d’inconfort. Finalement, les participants devaient suivre les déplacements d’un robot à la fois dans une scène réelle et dans sa reproduction la plus fidèle possible en virtuel. Les angles de vergence oculaire et les écarts angulaires à une cible disposée sur le robot constituent les indices de comparaison. La différence de variations de ces angles entre cas virtuels et réel permet de déterminer avec quelle configuration la position de l’objet, en particulier en profondeur, est rendue de la façon la plus réaliste.

Les analyses statistiques et quantitatives réalisées sur les résultats de cette expérimentation permettent de mettre en lumière les effets des paramètres testés. Concernant la distance inter-caméras, il apparaît tout d’abord qu’une DIC variable ou correspondant à la DIP est significativement plus performante que la DIC fixe en ce qui concerne les limites de fusion stéréoscopique atteintes et la correspondance des mouvements oculaires entre réel et virtuel. La distance interpupillaire anatomique conduit quant à elle à des notations subjectives plus élevées, bien que la différence avec les autres valeurs pour ce paramètre ne soit pas significative. Enfin, notre proposition de DIC variable en fonction du point de regard conduit à des comportements oculaires plus réalistes en considérant le nombre de participants ayant obtenu le résultat le plus proche du réel avec cette valeur de paramètre. C’est en revanche la DIP qui obtient les meilleures moyennes pour ce test.

En ce qui concerne le flou de profondeur de champ, les analyses révèlent que sa présence n’impacte pas significativement la capacité de fusion ou la similitude de comportement oculaire entre réel et virtuel. Elle s’est en revanche révélée significativement négative pour les notes attribuées dans le critère subjectif de perception des distances et des profondeurs, et

a également influé sur l'inspection visuelle en augmentant de manière significative la durée moyenne des fixations et saccades.

L'étude des résultats établit donc que nos propositions d'ajustement des paramètres n'ont pas amélioré significativement les performances du système de caméras et incite globalement à choisir une configuration dépourvue de flou et utilisant la distance interpupillaire anatomique. Cependant, si cette expérimentation n'a pas permis de valider nos hypothèses, elle ne les a pas non plus infirmées car certaines limitations techniques rencontrées ont plus que probablement eu une influence négative. En effet, des problèmes d'affichage intervenus en cours d'expérimentation ainsi qu'un faible nombre d'images générées par seconde, notamment en présence de flou de profondeur de champ, ont selon nous introduit des biais et conduit à des résultats qui ne sont pas imputables aux seules configurations. Nous suggérons en particulier de résoudre ces problèmes et de s'assurer, en le mesurant, que le temps de latence de mise à jour de l'affichage en fonction du point de regard n'est pas gênant pour le participant. Une seconde expérimentation permettra alors plus sûrement de conclure quant au potentiel des ajustements testés au cours de ce projet.

Cette recherche a dans un premier temps consisté à intégrer un système de suivi oculaire à la voûte de réalité virtuelle de l'Institut Philippe-Pinel de Montréal. Les applications à court terme comprennent donc la possibilité pour les cliniciens de cet institut de connaître en temps réel le point de regard de l'observateur, et ainsi l'objet de son attention, avec une bonne précision. Grâce à la procédure de calibration mise en place, la mesure d'imprécision fut réduite à 2,65 cm au cours des tests préliminaires et à 6,47 cm en moyenne sur l'ensemble des participants pour des cibles situées à 2 m, soit $0,76^\circ$ et $1,85^\circ$ d'angle visuel. Les outils d'analyse des résultats développés, extrayant des données statistiques ou des représentations graphiques du signal visuel, leur serviront également au moment de l'interprétation des comportements oculaires en matière d'attention.

Dans un second temps, nous avons conçu au cours de ce projet un cadre méthodologique novateur de comparaison des configurations stéréoscopiques. En effet, notre test utilisant la reproduction virtuelle d'une scène réelle et l'étude des mouvements de vergence pourra être employé dans d'autres recherches afin de mesurer efficacement la performance du rendu de la géométrie – positions, profondeurs, distances relatives et tailles – d'un environnement réel. Une seconde expérimentation réalisée après avoir effectué certaines corrections au niveau de l'installation de la voûte et de l'implémentation pourrait également révéler le potentiel réel de nos propositions d'ajustement.

Enfin, il est important de noter que dans le contexte d'application de la réalité virtuelle à la psychiatrie légale, toute amélioration de la configuration du système de caméras ou

du système d’affichage permet d’augmenter ses potentiels de diagnostic et de traitement, en renforçant le sentiment de présence des patients et en rapprochant les comportements exprimés au sein de la voûte de ceux dont ils feraient preuve dans le monde réel. L’intérêt est alors de valider l’utilisation d’environnements et d’avatars virtuels en lieu et place de personnes réelles, ce qui offre la possibilité de mettre en scène des situations davantage personnalisées, contrôlées et éthiquement acceptables. En mettant en lumière la configuration qui parmi les 6 testées et dans l’installation actuelle de la voûte est la plus confortable, la plus immersive et conduit en moyenne à un comportement oculaire plus naturel, notre projet permet aux cliniciens d’exploiter davantage les capacités de la réalité virtuelle.

ABSTRACT

This research project concerned a computer assisted virtual environment (CAVE). Such an environment makes the viewer dive into a computer-generated world, projected onto 4 faces of a room-sized cube in the middle of which he stands. Nevertheless, stereoscopic display often generates visual discomfort, or even visual fatigue. Our goal is to improve the immersive experience by trying to lower this visual discomfort, while making the world more realistic and improving the feeling of presence.

The objective was thus defined as follows: design and evaluate an adjustment method for the configuration settings of the cameras system, so as to find the optimal and customized values for each individual. Our assumptions stated that adjusting these values according to the gaze point position would lead to a more efficient configuration. Glasses with cameras pointing towards the eyes were used as an eye tracking system so as to calculate in real time the viewer's gaze point in the scene. This information was then used to adjust two parameters of the cameras system configuration: the distance between the two virtual cameras used for stereoscopic rendering of the scene, and an artificial depth of field generated around the gaze point. The position of the latter, and more particularly the objects that have caught the viewer's attention, also provide clues about the cognitive processing of visual information, i.e. how he perceived and inspected the scene.

The first methodological step was to develop tools in order to integrate the gaze tracking system to the CAVE of the Philippe-Pinel Institute of Montreal, to determine the gaze point through a method of raycasting coupled to a calibration procedure, and finally to change the configuration settings accordingly.

Subsequently, an experiment was conducted with participants to evaluate our proposal in comparison with five other more neutral and classic configurations, incorporating an ethical dimension to the project. The 6 tested configurations differ in the presence or absence of depth of field blur and the value of the inter-camera distance (ICD): fixed and chosen by the participant, fixed and equal to its anatomical interpupillary distance (IPD), or variable depending on the gaze point. Apart from the common comparison methods listed in the literature review, we sought a technique that would provide a true indication of the viewer's ocular behavior during an immersion session. To this aim, we developed a test based on the virtual reproduction of a real scene.

The 18 participants were recruited on a voluntary basis. To reduce the risk of eye strain during the immersion, they had to be familiar with virtual reality and stereoscopic displays,

i.e. having played video games and watched 3D movies without experiencing visual discomfort. These participants were asked to perform different tasks, each associated with a goal. First, they had to navigate through a rich virtual environment - an apartment with many objects - successively with each configuration, and give it a rating depending on the felt comfort, rendering realism, fun, distances and depths perception, and finally the feeling of presence in the virtual world. Then, they had to determine their limits of fusion by moving a virtual object closer or further. These provide an objective comparison of rendering effectiveness between the different configurations, the diplopia generated by difficulties to merge stereoscopic images being one of the main sources of visual discomfort. Finally, participants were asked to visually follow a robot in both a real scene and its as accurate as possible virtual reproduction. The angles of ocular vergence and deviations from a target located on top of the robot were used as comparison measures. The difference of angles variations between real and virtual helps determine the configuration with which the position of the object, especially the depth, is rendered in the most realistic way.

We carried out statistics and quantitative analysis on the experiment results in order to highlight the effects of the tested parameters. Concerning the inter-camera distance, it first appears that a variable ICD or one corresponding to the IPD is significantly more efficient than the fixed ICD, regarding the limits of stereoscopic fusion and the correspondence between eye movements in real and virtual. Moreover, the anatomical interpupillary distance leads to higher subjective ratings, although the difference with the other values for this parameter is not significant. Finally, our proposal of variable ICD depending on the gaze point leads to more realistic eye behavior when taking into account the number of participants who obtained the result closest to the real scene with this value of parameter. It is instead the IPD who achieves the best average for this test.

The presence of depth of field has meanwhile appeared to have no significant impact on the ability to merge stereoscopic images, or on the similarity between ocular behavior in real and virtual. However, it revealed to impact significantly negatively the scores given in the subjective criteria of distances and depths perception, and also influenced visual inspection by significantly increasing the average duration of the fixations and saccades.

Our results established that our proposals for adjusting parameters did not significantly improve the performance of the cameras system and encourages overall to select a configuration without blur and using anatomical interpupillary distance. However, if this experiment failed to prove our assumptions, it did not invalidate them because of technical limitations which had more than likely a negative influence. Indeed, display problems that occurred during the experiment as well as a small number of frames generated per second, especially

in the presence of depth of field blur, may have introduced bias and led to results that can not be attributed solely to the configuration. We suggest in particular to solve these problems and to ensure, by measuring it, that the lag in updating the display based on the gaze point is not annoying for the participant. A second experiment will then be necessary before drawing a final conclusion on the tested adjustments potential.

The first step of this research consisted in integrating an eye tracking system to the Philippe-Pinel Institute of Montreal CAVE. The short term applications therefore include the ability for the clinicians of this institute to know in real time the observer gaze point, and thus the object of his attention, with good accuracy. Thanks to our calibration procedure, measured inaccuracy was reduced to 2.65 cm over the preliminary tests and to 6.47 cm on average across all participants for targets at 2 m, i.e. 0.76° 1.85° of visual angle. The results analysis tools developed, that extract statistical data or graphical representations from the visual signal, will also be of use to them when interpreting eye behavior in terms of attention.

In a second step, we provide an innovative framework for comparing stereoscopic configurations. Indeed, our test using a virtual reproduction of a real scene and the study of vergence movements can be used in further research to effectively measure the performance of rendering the geometry – positions, depths, relative distances and sizes – of a real environment. A second experiment performed after making some corrections in the CAVE installation and in the implementation could also reveal the true potential of the adjustments we proposed.

Finally, it is important to note that in the context of virtual reality application in forensic psychiatry, any improvement in the cameras configuration system or display system increases its diagnostic and treatment potentials, by enhancing the patients' feeling of presence and by bringing the behaviors expressed in the CAVE closer to the ones they would show in the real world. The interest is then to validate the use of virtual environments and avatars instead of real people, making possible to stage more specific, controlled and ethically acceptable situations. By highlighting the configuration among the 6 tested which is, in the current CAVE installation, the most comfortable, most immersive and which leads on average to more natural ocular movements, our project enables clinicians to further exploit the virtual reality potency.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	ix
TABLE DES MATIÈRES	xii
LISTE DES TABLEAUX	xvi
LISTE DES FIGURES	xix
LISTE DES ANNEXES	xxv
LISTE DES SIGLES ET ABRÉVIATIONS	xxvi
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions et concepts de base	1
1.1.1 Vision stéréoscopique	1
1.1.2 Système de réalité virtuelle	2
1.1.3 Suivi oculaire et flou de profondeur de champ	2
1.1.4 Application de la réalité virtuelle à la psychologie et psychiatrie	3
1.2 Éléments de la problématique	3
1.2.1 Conception et évaluation de méthodes d'ajustement de l'affichage stéréoscopique	3
1.2.2 Développement d'outils pour les cliniciens	5
1.3 Objectifs de recherche	6
1.4 Plan du mémoire	7
CHAPITRE 2 REVUE DE LITTÉRATURE	8
2.1 Principes de la vision humaine	8
2.1.1 Caractéristiques rétinienne	8
2.1.2 Mouvements oculaires	9
2.1.3 Vision stéréoscopique	10

2.1.4	Fatigue et confort visuels	14
2.1.5	Introduction à l'attention visuelle	15
2.2	Suivi oculaire : traitement des données	16
2.2.1	Traitement des images des yeux	16
2.2.2	Détermination du point de regard 3D	17
2.2.3	Mesure de qualité des données	21
2.2.4	Détection des évènements oculaires	24
2.2.5	Applications des systèmes de suivi de regard	24
2.3	Systèmes stéréoscopiques	26
2.3.1	Systèmes de caméras et d'affichage	26
2.3.2	Comparaison des configurations	29
2.3.3	Comparaison des valeurs de paramètres	32
2.3.4	Critères d'évaluation de la performance	35
2.4	Flou de profondeur de champ	36
2.4.1	Techniques de génération de flou	37
2.4.2	Cas monoscopique	38
2.4.3	Cas stéréoscopique	39
2.5	Applications existantes en psychologie et psychiatrie	40
2.5.1	Suivi de regard	41
2.5.2	Impact du flou sur l'attention visuelle	43
2.5.3	Réalité virtuelle	44
2.6	Objectifs et hypothèses	47
2.6.1	Conclusion sur l'état de l'art	47
2.6.2	Rappel de l'objectif	47
2.6.3	Notre choix de paramètres	47
2.6.4	Objectifs spécifiques	49
2.6.5	Hypothèses de travail	49
CHAPITRE 3	MÉTHODOLOGIE	54
3.1	Matériel	54
3.1.1	Voûte de l'Institut Philippe-Pinel de Montréal	54
3.1.2	Voûte l'École Polytechnique de Montréal	59
3.2	Tests préliminaires	60
3.2.1	Tests des configurations du système de caméras	61
3.2.2	Intensité de la profondeur de champ	65
3.3	Calcul du point de regard 3D dans la scène virtuelle	66

3.3.1	Suivi oculaire et repérage de la tête	66
3.3.2	Calcul du point de regard	68
3.3.3	Calcul de la position des caméras	72
3.4	Expérimentation avec participants	74
3.4.1	Mise en place	74
3.4.2	Calibration	74
3.4.3	Choix des distances inter-caméras	75
3.4.4	Navigation	77
3.4.5	Limites de fusion	78
3.4.6	Comparaison entre réel et virtuel	78
3.4.7	Considérations sur les ordres	80
3.5	Participants	80
3.5.1	Risques encourus par les participants	81
CHAPITRE 4	RÉSULTATS	82
4.1	Remarques préliminaires sur les analyses	82
4.2	Choix des valeurs de paramètres	83
4.3	Notations subjectives	84
4.3.1	Ordre de navigation	87
4.3.2	Étude approfondie basée sur la DIC1	88
4.3.3	Étude approfondie basée sur les DIC3	91
4.3.4	Questionnaire de propension à l'immersion	93
4.3.5	Qualité du calcul du POR	93
4.3.6	Fixations, saccades et nombre d'objets regardés	97
4.3.7	Nombre d'objets regardés	100
4.3.8	Cartes de densité, chemin visuel et objets d'intérêt	100
4.4	Limites de fusion	103
4.4.1	Effet de la configuration stéréoscopique	103
4.4.2	Effet de la DIC	104
4.4.3	Lien entre notes subjectives et limites de fusion	105
4.4.4	Effet de l'imprécision de l'estimation du POR	106
4.5	Comparaison de mouvements oculaires	106
4.5.1	Différence de vergence entre cas réel et virtuel	107
4.5.2	Écart angulaire à la cible	112

CHAPITRE 5	DISCUSSION	117
5.1	Vérification des hypothèses	117
5.2	Groupement des participants	122
5.2.1	Port de lentilles correctrices	123
5.2.2	Genre du participant	124
5.2.3	Carrés latins	125
5.3	Autres analyses envisageables	127
5.3.1	Étude au cas par cas	127
5.3.2	Étude d'informations peu exploitées	128
5.3.3	Analyse qualitative de l'attention visuelle	128
5.4	Limitations de la solution proposée	129
5.4.1	Limitations théoriques	129
5.4.2	Matériel et implémentation	130
5.5	Synthèse quant aux objectifs	132
CHAPITRE 6	CONCLUSION	134
6.1	Synthèse des travaux	134
6.1.1	Conception de l'expérimentation	134
6.1.2	Analyse des résultats	135
6.1.3	Applications de la recherche	136
6.2	Améliorations futures	137
RÉFÉRENCES		139
ANNEXES		145

LISTE DES TABLEAUX

Tableau 2.1	Les trois configurations du système de caméras.	27
Tableau 2.2	Calcul des angles d'ouverture horizontale et verticale de la caméra virtuelle.	28
Tableau 3.1	Les 6 configurations stéréoscopiques comparées au cours de l'expérimentation.	65
Tableau 4.1	Moyennes et dénombrement des participants ayant octroyé les meilleures notes tous critères confondus.	86
Tableau 4.2	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives.	87
Tableau 4.3	ANOVA à 1 facteur contrôlé : effet de l'ordre de navigation sur les notations subjectives.	87
Tableau 4.4	Probabilités d'indépendance entre la DIC1 choisie et les notes subjectives attribuées à DIC1 et FLOUDIC1 , et variations des régressions linéaires associées.	89
Tableau 4.5	Probabilités d'indépendance entre le ratio DIP/DIC1 et les notes subjectives attribuées à DIC1 et FLOUDIC1 , et variations des régressions linéaires associées.	90
Tableau 4.6	Probabilités d'indépendance entre l'écart-type des 3 DIC1 choisies et les notes subjectives attribuées à DIC1 et FLOUDIC1 , et variations des régressions linéaires associées.	91
Tableau 4.7	Probabilités d'indépendance entre la moyenne des écarts-types des DIC3 et les notes subjectives attribuées à DIC3 et FLOUDIC3 , et variations des régressions linéaires associées.	92
Tableau 4.8	Probabilité d'indépendance entre l'écart-type inter-DIC3 et les notes subjectives attribuées à DIC3 et FLOUDIC3 , et variations des régressions linéaires associées.	93
Tableau 4.9	Probabilité d'indépendance entre le score au QPI et les notes subjectives, et variations des régressions linéaires associées.	94
Tableau 4.10	Probabilité d'indépendance entre la mesure d'imprécision et les notes moyennes, et variations des régressions linéaires associées.	95
Tableau 4.11	Probabilité d'indépendance entre la mesure de stabilité et les notes moyennes, et variations des régressions linéaires associées.	95

Tableau 4.12	Probabilités d'indépendance entre la mesure d'incertitude et la note moyenne, et variations des régressions linéaires associées.	96
Tableau 4.13	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur la quantité d'inspection visuelle. . . .	99
Tableau 4.14	Probabilités d'indépendance entre la mesure de quantité d'inspection visuelle et la note moyenne, et variations des régressions linéaires associées.	99
Tableau 4.15	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les limites de fusion.	104
Tableau 4.16	Moyennes et dénombrement pour chaque configuration des participants ayant atteint la limite la plus basse.	104
Tableau 4.17	Probabilités d'indépendance entre les limites et les notes (moyennes ou de distances), et variations des régressions linéaires associées.	106
Tableau 4.18	Probabilités d'indépendance entre la mesure d'imprécision et les limites, et variations des régressions linéaires associées.	106
Tableau 4.19	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les différences de vergence/distance et vergence en X/distance entre les cas virtuels et réel.	108
Tableau 4.20	Moyennes et dénombrement pour chaque configuration des participants ayant obtenu les meilleurs résultats en terme de vergence oculaire. . .	110
Tableau 4.21	Probabilités d'indépendance entre les différences de vergence/distance et vergence en X/distance et les notes (moyennes ou de distances), et variations des régressions linéaires associées.	111
Tableau 4.22	Probabilités d'indépendance entre la mesure d'imprécision et les différences de vergence/distance et vergence en X/distance, et variations des régressions linéaires associées.	111
Tableau 4.23	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les différences d'écart angulaire et écart angulaire en X entre les cas virtuels et réel.	114
Tableau 4.24	Moyennes et dénombrement pour chaque configuration des participants ayant obtenu les meilleurs résultats en terme d'écart angulaire à la cible.	115
Tableau 4.25	Probabilités d'indépendance entre les différences d'écarts angulaires et les notes (moyennes ou de distances), et variations des régressions linéaires associées.	115

Tableau 4.26	Probabilités d'indépendance entre la mesure d'imprécision et les différences de vergence/distance et vergence en X/distance, et variations des régressions linéaires associées.	116
Tableau 5.1	Mesures ETG d'incertitude du repérage des yeux, pour les participants avec et sans lentilles correctrices.	123
Tableau 5.2	Test de Student apparié comparant le groupe de participants portant des lentilles à celui n'en portant pas.	124
Tableau 5.3	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives et les limites de fusion, en écartant les 5 participants portant des lentilles de contact.	124
Tableau 5.4	Test de Student comparant le groupe de participants de genre féminin et de genre masculin.	125
Tableau 5.5	ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives et les limites de fusion, en écartant les 3 participantes.	125
Tableau 5.6	ANOVA à 1 et 2 facteurs contrôlés : effets de l'appartenance à un groupe et de la configuration sur les notations subjectives, les limites de fusion et les différences de vergence totale et en X.	126
Tableau A.1	Paramètres des caméras virtuelles dans le cas général.	145
Tableau A.2	Paramètre spécifique à la configuration parallèle avec troncs asymétriques.	145
Tableau A.3	Paramètres spécifiques à la configuration parallèle avec troncs asymétriques et simulation des globes oculaires.	146
Tableau A.4	Paramètres spécifiques à la configuration convergente.	147
Tableau D.1	Indications sur les critères de notation subjective.	152
Tableau D.2	Moyennes et dénombrement des participants ayant octroyé les meilleures notes, détaillés par critère.	152

LISTE DES FIGURES

Figure 2.1	Anatomie de l'œil humain.	9
Figure 2.2	Vision binoculaire.	11
Figure 2.3	Mouvements de vergence.	12
Figure 2.4	Horoptère et zone de fusion de Panum.	13
Figure 2.5	Reflets cornéens obtenus avec les ETG.	17
Figure 2.6	Point de regard défini en intersectant l'axe optique avec la scène virtuelle.	18
Figure 2.7	Point de regard défini comme milieu du plus court segment reliant les deux lignes de vue.	19
Figure 2.8	Variations de la distance interpupillaire liées aux changements de pro- fondeur du point de regard.	21
Figure 2.9	Configurations parallèle (à gauche), images translatées (au milieu) et convergente (à droite).	27
Figure 2.10	Tronc de projection d'une caméra virtuelle.	28
Figure 2.11	Parallaxe négative, nulle ou positive en fonction de la position de l'objet par rapport au plan de projection.	28
Figure 2.12	Configuration du système d'affichage.	29
Figure 2.13	Distorsion trapézoïdale.	30
Figure 2.14	Flou de profondeur de champ dépendant du point de regard. Exemples de rendus sans flou (au milieu) et avec flou (à droite).	37
Figure 2.15	Exemples de représentations graphiques des fixations : carte de densité (à gauche) et chemin de balayage (à droite).	42
Figure 3.1	Installation de la voûte de réalité virtuelle de l'Institut Philippe-Pinel.	54
Figure 3.2	Deux projecteurs par écran (en haut à gauche). Caméra utilisée par AnyBlend (en bas à gauche) pour réaliser la fusion des images (à droite).	55
Figure 3.3	Système de repérage optique à 8 caméras.	56
Figure 3.4	Captures d'écran de MotiveTracker : enregistrement du corps rigide des lunettes, représentant la tête de l'observateur.	57
Figure 3.5	Configuration de la voûte dans MiddleVR, et modification des troncs de projection en fonction de la position de l'observateur.	58
Figure 3.6	Le montage ETG et Volfoni (à gauche) et l'interface utilisateur de iViewETG (à droite).	60

Figure 3.7	Impact de la distance inter-caméras sur les images stéréoscopiques obtenues avec la configuration parallèle, troncs asymétriques (en haut) ou troncs symétriques (en bas).	63
Figure 3.8	Positions et troncs de projection des caméras virtuelles dans le cas de la configuration parallèle, troncs asymétriques et simulation des globes oculaires.	63
Figure 3.9	Exemple de cas où une image serait indéfinie pour l'un des écrans avec la configuration convergente.	64
Figure 3.10	Repère local des lunettes ETG.	68
Figure 3.11	Mesures d'imprécision, moyenne et par indice de point de calibration (de 1 à 25 positionnés tel que sur la grille), obtenues avec les différents POR (en cm).	71
Figure 3.12	Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec différentes valeurs du paramètre R.	72
Figure 3.13	Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec 5 ou 9 points de calibration.	72
Figure 3.14	Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec différentes méthodes de calibration.	73
Figure 3.15	Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues après déplacement des lunettes ou des marqueurs.	73
Figure 3.16	Les lunettes installées sur le participant (à gauche). Scène pour la calibration (à droite).	75
Figure 3.17	Scène pour le choix des DIC.	76
Figure 3.18	Détail de la scène d'appartement utilisée au cours de l'expérimentation.	77
Figure 3.19	Robot réel (en haut à gauche et au milieu) et virtuel (en haut à droite). Installation pour le test réel (en bas à gauche) et virtuel (en bas à droite).	79
Figure 4.1	DIP et choix de DIC (en cm).	83
Figure 4.2	Les 3 itérations de choix pour chacune des 4 DIC (en cm).	84
Figure 4.3	Choix de DIC en fonction de la distance de l'objet étalon.	85
Figure 4.4	Notes obtenues par chacune des configurations dans chacun des critères.	85
Figure 4.5	Notes obtenues par ordre de navigation dans chacun des critères.	87
Figure 4.6	Données quantitatives de l'inspection visuelle par configuration.	98
Figure 4.7	Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre de fixations.	100
Figure 4.8	Exemples de cartes de densité, pour les configurations DIC3 (à gauche) et FLOUDIC3 (à droite).	102

Figure 4.9	Exemples de chemins visuels.	102
Figure 4.10	Limites de fusion par configuration, pour les cas vers l'avant et vers l'arrière.	103
Figure 4.11	Limites obtenues vers l'avant et vers l'arrière en fonction de la DIC, en écartant le participant aux limites les plus importantes.	105
Figure 4.12	Vergence/distance (à gauche) et vergence en X/distance (à droite) en fonction du temps.	108
Figure 4.13	Vergence totale (en haut), vergence en X (en bas à gauche) et vergence en Y (en bas à droite) en fonction de la distance au robot.	109
Figure 4.14	Différence de vergence/distance et vergence en X/distance entre les cas virtuels et réel.	110
Figure 4.15	vergence en X en fonction de la distance au robot pour deux participants ne semblant pas percevoir correctement la profondeur avec un affichage stéréoscopique.	112
Figure 4.16	Écart angulaire (à gauche) et écart angulaire en X (à droite) en fonction du temps.	113
Figure 4.17	Différence d'écart angulaire et écart angulaire en X entre les cas virtuels et réel.	114
Figure 5.1	Mesures d'imprécision et d'instabilité pour les participants avec et sans lentilles correctrices.	123
Figure 5.2	Limites de fusion atteintes par les 6 groupes de carrés latins.	126
Figure 5.3	Notes données par les femmes et les hommes dans le critère de perception des distances.	128
Figure A.1	Tronc de projection.	145
Figure A.2	Configuration parallèle avec troncs asymétriques (à gauche) et avec simulation des globes oculaires (à droite). Les lignes en pointillés correspondent au tronc symétrique.	146
Figure A.3	Paramètres de la configuration convergente dans différents cas de figure dépendant de la position du point de regard (en haut). Exemple du tronc de projection de la caméra droite (en bas).	147
Figure B.1	Chemin de navigation (en haut) et détails de la scène d'appartement.	150
Figure C.1	Incertitudes de la détection des pupilles et des globes oculaires au cours du temps.	151
Figure D.1	Notes obtenues par DIC1 en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).	153

Figure D.2	Notes obtenues par FLOUDIC1 en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).	154
Figure D.3	Notes moyennes DIC1 et FLOUDIC1 en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).	155
Figure D.4	Notes obtenues par DIC1 en fonction du ratio DIP/DIC1 (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).	156
Figure D.5	Notes obtenues par FLOUDIC1 en fonction du ratio DIP/DIC (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).	157
Figure D.6	Notes moyennes DIC1 et FLOUDIC1 en fonction du ratio DIP/DIC (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).	158
Figure D.7	Notes obtenues par DIC1 en fonction de l'écart-type entre les 3 DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).	159
Figure D.8	Notes obtenues par FLOUDIC1 en fonction de l'écart-type entre les 3 DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).	160
Figure D.9	Notes moyennes DIC1 et FLOUDIC1 en fonction de l'écart-type entre les DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).	161
Figure D.10	Notes obtenues par DIC3 en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).	162
Figure D.11	Notes obtenues par FLOUDIC3 en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).	163
Figure D.12	Notes moyennes DIC3 et FLOUDIC3 en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).	164
Figure D.13	Notes obtenues par DIC3 en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).	165

Figure D.14	Notes obtenues par FLOUDIC3 en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).	166
Figure D.15	Notes moyennes DIC3 et FLOUDIC3 en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).	167
Figure D.16	Notes moyennes en fonction du score total obtenu au QPI (5 graphiques du haut). Liens implication-immersion, focus-immersion, émotion-immersion, émotion-plaisir, jeu-plaisir (5 graphiques du bas).	168
Figure D.17	Note moyenne en fonction de la mesure d'imprécision, en considérant l'ensemble des participants (à gauche) ou en écartant les 3 plus hautes imprécisions (à droite).	169
Figure D.18	Note moyenne en fonction de la mesure d'instabilité, en considérant l'ensemble des participants (à gauche) ou en écartant les 2 plus hautes instabilités (à droite).	169
Figure D.19	Notes moyennes en fonction de la mesure d'imprécision du point de regard (5 graphiques du haut), puis en écartant les 3 plus hautes imprécisions (5 graphiques du bas).	170
Figure D.20	Notes moyennes en fonction de la mesure d'instabilité du point de regard (5 graphiques du haut), puis en écartant les 2 plus hautes instabilités (5 graphiques du bas).	171
Figure D.21	Impact de l'incertitude du repérage des pupilles (à gauche) et des globes oculaires (à droite) sur les notes.	172
Figure D.22	Notes moyennes en fonction de l'incertitude sur la détection des pupilles.	172
Figure D.23	Notes moyennes en fonction de l'incertitude sur la détection des globes oculaires.	173
Figure D.24	Notes en fonction de l'angle moyen entre les 3 POR (à gauche) et de l'angle entre PORG et PORD (à droite).	173
Figure D.25	Notes moyennes en fonction de l'angle entre PORG et PORD.	174
Figure D.26	Notes moyennes en fonction de la moyenne des 3 angles entre les 3 POR.	174
Figure D.27	Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre de fixations.	175
Figure D.28	Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction de la durée moyenne des fixations. . .	176
Figure D.29	Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction de la durée moyenne des saccades. . .	177

Figure D.30	Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre d'objets regardés.	178
Figure E.1	Limites de fusion obtenues en fonction de la DIC, par configuration (en haut) en regroupant les configurations partageant la même DIC (en bas).	179
Figure E.2	Notes moyennes en fonction des limites de fusion (en haut), puis en écartant la limite de fusion extrême (en bas).	180
Figure E.3	Notes de perception des distances en fonction des limites de fusion (en haut), puis en écartant la limite de fusion extrême (en bas).	181
Figure E.4	Limites obtenues en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).	182
Figure F.1	Exemples de graphiques représentant la vergence en X en fonction de la distance montrant les angles supérieurs induits par les configurations DIC1 et FLOUDIC1	183
Figure F.2	Notes moyennes (en haut) et de perception des distances (en bas) en fonction de la différence de vergence/distance et vergence en X/distance par rapport à la configuration réelle.	184
Figure F.3	Différence de vergence/distance et vergence en X/distance par rapport à la configuration réelle en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).	185
Figure F.4	Exemples de graphiques représentant les écarts angulaires total et en X à la cible en fonction du temps, montrant un écart angulaire tournant autour de 0 pour la configuration réelle.	186
Figure F.5	Exemples de graphiques représentant les écarts angulaires total et en X à la cible en fonction du temps, montrant un écart angulaire différent de 0 pour la configuration réelle.	187
Figure F.6	Notes moyennes (en haut) et de perception des distances (en bas) en fonction de la différence d'écarts angulaires total et en X à la cible par rapport à la configuration réelle.	188
Figure F.7	Différence d'écarts angulaires total et en X à la cible par rapport à la configuration réelle en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).	189

LISTE DES ANNEXES

Annexe A	Notations et calculs des paramètres du système de caméras	145
Annexe B	Informations supplémentaires sur l'implémentation et l'expérimentation	148
Annexe C	Mesures de la qualité suivi par les ETG	151
Annexe D	Notations subjectives	152
Annexe E	Limites de fusion	179
Annexe F	Comparaison des mouvements oculaires	183

LISTE DES SIGLES ET ABRÉVIATIONS

EOG	Électro-oculographie
POG	Photo-oculographie
VOG	Vidéo-oculographie
PCA	<i>Principal component analysis</i>
PCCR	<i>Pupil center - corneal reflections</i>
PCD	<i>Pupil center distance</i>
PSOM	<i>Parametrized self-organizing map</i>
ROWA	<i>Recursive online weight average filter</i>
PSNR	<i>Peak signal to noise ratio</i>
DADC	<i>Dynamic attention-aware disparity control</i>
DRC	<i>Depth range control</i>
VRET	<i>Virtual reality exposure therapy</i>
CAVE	<i>Cave automatic virtual environment</i>
FPS	<i>Frames per second</i>
VRPN	<i>Virtual reality peripheral network</i>
ETG	<i>Eye Tracking Glasses</i>
SMI	<i>SensoMotoric Instruments</i>
DLL	<i>Dynamic link library</i>
SDK	<i>Software development kit</i>
POR	<i>Point of regard</i>
DIC	Distance inter-caméras
DIP	Distance interpupillaire
IR	Infrarouge
DEL	Diode électroluminescente
QPI	Questionnaire de propension à l'immersion

CHAPITRE 1

INTRODUCTION

Au cours de cette maîtrise, nous nous sommes intéressés à l'amélioration et à la comparaison des performances de configurations stéréoscopiques employées au sein d'une voûte de réalité virtuelle. Concernant le premier aspect, nous nous sommes appuyés sur le point de regard de l'observateur dans l'environnement virtuel afin d'ajuster les paramètres de la configuration, dans le but que celle-ci fournisse un visuel plus confortable et plus proche de la réalité. En ce qui concerne la comparaison, en dehors des méthodes classiques répertoriées dans la littérature telles que l'évaluation subjective par des participants ou la détermination de leurs capacités de fusion, nous avons utilisé une méthode innovante de comparaison objective des configurations virtuelles, basée sur la reproduction à l'identique d'une scène réelle et sur l'analyse des mouvements oculaires de l'observateur.

1.1 Définitions et concepts de base

Afin de saisir le cadre de cette recherche et d'en comprendre les enjeux, il est important de définir certains concepts, qui seront développés plus largement dans la revue de littérature (voir chapitre 2).

1.1.1 Vision stéréoscopique

La vision humaine se caractérise avant tout par la capacité de percevoir la profondeur de l'environnement, grâce à la combinaison effectuée par le cerveau des informations visuelles complémentaires perçues par l'œil gauche et l'œil droit : il s'agit de la vision stéréoscopique. Les indices de profondeur sont principalement les disparités, c'est-à-dire les décalages dans la position des objets, existantes entre les deux images rétiniennes en raison de la distance qui sépare les yeux. La perception de la profondeur peut également être stimulée par deux images planes, à compter que celles-ci présentent des disparités proches de celles intervenant en observant une scène réelle. Il s'agit de la stéréoscopie, définie comme un « procédé donnant l'impression de relief à partir d'un couple d'images enregistrées aux deux extrémités d'une base » (Office québécois de la langue française).

1.1.2 Système de réalité virtuelle

Le principe d'un affichage stéréoscopique en réalité virtuelle est d'acheminer ces deux images légèrement décalées d'une même scène virtuelle aux deux yeux afin de permettre à l'observateur de la percevoir en 3 dimensions. Dans un environnement virtuel immersif tel qu'une CAVE, ou voûte de réalité virtuelle, les images sont projetées sur les murs et le sol et des lunettes stéréoscopiques actives occultent alternativement chacun des deux yeux quand l'image qui est destinée à l'autre œil est affichée. La génération des images stéréoscopiques s'effectue par ordinateur, en utilisant des caméras virtuelles positionnées dans la scène aux alentours des yeux. Le calcul de chaque image dépend de paramètres propres aux caméras tels que leur position, leur orientation, ou encore leur longueur focale. Cependant, la perception de la profondeur par l'observateur pourra, elle, être influencée par des paramètres reliant les deux caméras, en particulier leur orientation l'une par rapport à l'autre et la distance qui les sépare, dont la valeur impacte directement celle des disparités entre les deux images.

L'immersion au sein d'un tel environnement ne provoque pas les mêmes ressentis ni les mêmes réponses oculaires que l'observation naturelle d'un monde en relief. Outre le fait que le réalisme des objets, des distances, des tailles et des profondeurs peut se révéler peu convaincant, et donc diminuer le sentiment de présence dans la scène virtuelle, une exposition prolongée s'avère contraignante pour les yeux et provoque parfois inconfort et fatigue visuels. Les raisons sont principalement le conflit accommodation-convergence, des disparités trop importantes ou non réalistes, et enfin l'absence courante de flou dans la scène. De nombreuses recherches se sont penchées sur la manière d'améliorer l'expérience de l'observateur, en particulier son confort visuel, son impression de réalisme, son plaisir, sa perception des distances et des profondeurs et son sentiment d'immersion.

1.1.3 Suivi oculaire et flou de profondeur de champ

Parmi les solutions et outils proposés dans ce but, nous nous sommes particulièrement intéressés à la détermination du point de regard à l'aide d'un système de suivi oculaire, ainsi qu'à la génération de flou de profondeur de champ, dont nous ébauchons ici brièvement les concepts.

Le suivi oculaire, ou oculométrie, est défini comme un « ensemble de techniques consistant à capter les mouvements de l'œil, notamment à l'aide d'une caméra infrarouge, ce qui permet de déterminer les endroits où l'utilisateur pose son regard » (Office québécois de la langue française). Il s'agit la plupart du temps de filmer les yeux de l'individu, et d'extraire des images enregistrées les informations de direction du regard. Combinées à la position et à l'orientation de sa tête, il est alors possible de déterminer à chaque instant quelle région de

la scène réelle ou virtuelle il regardait.

Le flou de profondeur de champ correspond quant à lui à un traitement infographique des images stéréoscopiques au cours de leur génération par le pipeline graphique, afin de simuler le flou qui intervient dans la vision naturelle en raison de l'accommodation du cristallin de l'œil à la profondeur regardée dans l'espace. À partir d'un rendu entièrement net de l'image captée par la caméra virtuelle, le flou est généralement obtenu en combinant la couleur d'un pixel à un échantillonnage des couleurs de ses pixels voisins.

La détermination du point de regard a déjà été testée dans certaines recherches afin d'ajuster les paramètres stéréoscopiques (Celikcan et al., 2013, Milgram and Krüger, 1992), et la présence de flou s'est révélée permettre la diminution de l'inconfort visuel (Lambooij et al., 2007, Blum et al., 2010).

1.1.4 Application de la réalité virtuelle à la psychologie et psychiatrie

Enfin, afin d'appréhender l'étendue des domaines d'applications de notre étude, il est important de souligner le contexte de sa réalisation en collaboration avec l'Institut Philippe-Pinel de Montréal. Les environnements immersifs de réalité virtuelle représentent en effet des outils d'aide thérapeutique pour le diagnostic et le traitement de pathologies psychologiques et psychiatriques. Les patients sont alors immergés au sein de mondes virtuels dont les paramètres sont parfaitement contrôlés et définis afin de correspondre au mieux à leurs besoins. Ces environnements permettent d'améliorer l'efficacité de certains exercices, réalisés auparavant uniquement en entretien avec un clinicien, grâce à un sentiment de présence accru par l'immersion, mais également de développer de nouveaux traitements mettant en scène des situations qui ne pourraient pas être reproduites en réalité.

Cette approche d'aide thérapeutique en réalité virtuelle a déjà été testée afin de traiter certaines phobies et son efficacité en fait un sujet d'étude prometteur et novateur. Ce projet aura des retombées pratiques à court terme, étant donné que notre intégration du système de suivi oculaire à la voûte de réalité virtuelle de l'Institut Philippe-Pinel de Montréal et nos résultats sur la configuration la plus performante seront utilisés dans les expérimentations menées auprès des patients.

1.2 Éléments de la problématique

1.2.1 Conception et évaluation de méthodes d'ajustement de l'affichage stéréoscopique

Comme expliqué ci-dessus, le calcul des deux images stéréoscopiques dépend de paramètres qui peuvent influencer sur le confort ressenti en immersion et sur l'impression de réalisme

du rendu. Deux des plus importants, c'est-à-dire ayant un impact fort sur les critères cités et ayant été déjà étudiés dans la littérature, sont la distance qui sépare les caméras virtuelles et la génération artificielle d'un flou de profondeur de champ (ainsi que ses caractéristiques, telles que son intensité). En effectuant une revue de la littérature, il nous est apparu qu'il n'existe pas de valeur optimale pour ces paramètres, l'ajustement de ceux-ci pouvant par exemple dépendre de la physiologie du participant, en particulier son écart interpupillaire, et de ses préférences. Cependant, certaines approches restaient selon nous peu ou pas explorées, en particulier celles utilisant le suivi de regard dans une voûte de réalité virtuelle, et représentaient des potentialités intéressantes d'amélioration de la performance de l'affichage stéréoscopique.

Approche envisagée

Nous avons donc décidé de concevoir une méthode d'ajustement des deux paramètres précités, en fonction de choix personnels de l'individu et de l'estimation en temps réel de son point de regard dans le monde virtuel, dans le but d'optimiser son expérience d'immersion. Nous proposons à l'observateur de choisir sa DIC, celle-ci restant soit fixe tout au long de l'immersion, soit interpolée entre plusieurs choix réalisés pour des distances différentes en fonction de la distance entre son point de regard et le centre de ses yeux. Cette version interpolée n'a à notre connaissance jamais été testée, et pourrait selon nous apporter une solution à la déformation de la perception des distances et des profondeurs ressentie en immersion qui a, elle, fait l'objet de plusieurs recherches. Elle nécessite néanmoins la détermination de la position du point de regard, à notre connaissance rarement employée dans une voûte de réalité virtuelle, information nécessaire également à la simulation du flou de profondeur de champ. Nos hypothèses premières étaient donc que de permettre à l'observateur de choisir sa DIC et de l'ajuster en fonction de la distance au point de regard permettrait d'obtenir une perception de la profondeur plus réaliste. Nous supposons également que l'ajout d'un flou de profondeur de champ augmenterait le réalisme et l'intervalle de fusion de l'observateur, tout en diminuant son inconfort visuel. Nous nous basions enfin sur l'hypothèse que le repérage effectué par le suivi oculaire utilisé, en l'occurrence les Eye Tracking Glasses développées par SMI, serait suffisamment rapide et précis pour disposer d'une bonne estimation de la position du point de regard et permettre d'ajuster les paramètres sans latence.

Évolution de la réalisation

La revue de littérature ainsi que des tests préliminaires nous ont permis tout d'abord d'écarter certaines configurations et de sélectionner les plus intéressantes et innovantes. Après

avoir développé les outils permettant de calculer le point de regard en utilisant les données fournies par les ETG, et d'ajuster les paramètres de l'affichage stéréoscopique en fonction, il nous semble important de mettre à l'épreuve ces configurations au cours d'une expérimentation faisant intervenir des participants externes. Nous souhaitons les comparer à des configurations plus neutres ou classiques, en particulier celles prenant comme DIC la distance interpupillaire anatomique des individus. Les critères de comparaison sélectionnés, subjectifs et objectifs, nous guideront afin de déterminer le protocole expérimental. Plus précisément, des questionnaires permettront de recueillir les appréciations subjectives des participants. Pour les critères objectifs, ils devront accomplir certaines tâches : déplacer un objet afin de déterminer leurs limites de fusion, et suivre du regard une cible mouvante à la fois en situation réelle et virtuelle.

L'intervention de participants nécessite de prendre en compte l'aspect éthique de notre expérimentation. En particulier, il faut préciser que les risques induits par l'immersion dans une voûte de réalité virtuelle ne sont pas plus importants que ceux découlant d'une exposition à une télévision 3D ou au visionnement d'un film en 3D. La gêne qui peut être ressentie lors de l'utilisation prolongée des lunettes stéréoscopiques se doit tout de même d'être limitée en incluant des temps de pause au protocole expérimental.

1.2.2 Développement d'outils pour les cliniciens

Le deuxième élément important de la problématique derrière ce projet concerne l'application de la réalité virtuelle à la psychologie et à la psychiatrie. En effet, dans ce contexte, le suivi de regard n'est pas seulement utilisé comme un facteur d'ajustement des paramètres stéréoscopiques, mais recèle de nombreuses informations sur l'état de l'observateur, son attention visuelle et son exploration de la scène. Les cliniciens utilisent la connaissance du point de regard, ainsi que certaines données statistiques extraites à partir de celle-ci, pour le diagnostic et le traitement de certaines pathologies. Les données couramment utilisées sont les suivantes :

- Mesure de proxémie : distance spatiale entre les différents objets et sous-objets identifiés.
- Mesures de vergence : valeur de l'angle de vergence binoculaire.
- Mesures de ligne de visée : distance angulaire ou spatiale à la cible, c'est-à-dire entre le point de regard de l'observateur et les objets et sous-objets identifiés.
- Analyses statistiques : nombre de fixations et leur durée moyenne, nombre de saccades, leur durée moyenne et leur distance angulaire moyenne. Nombre de fois que le point de regard s'est porté sur un objet ou sous-objet identifié, ainsi que la durée moyenne de scrutation.
- Représentations graphiques : cartes de densités pour visualiser la répartition des points

de regard dans une scène donnée, chemins visuels reliant les points de regard successifs.

Il s'agissait donc de développer des outils capables d'extraire ces informations statistiques, au regard des objets et sous-objets identifiés *a priori* ou *a posteriori* par les cliniciens. Ces outils leur faciliteront ainsi l'étude du comportement visuel des patients au cours des expérimentations subséquentes menées, et nous permettront également d'analyser plus en détail les réponses oculaires des participants de notre expérimentation face aux différentes configurations présentées.

À noter par ailleurs que la configuration ressortant comme la meilleure à l'issue de cette expérimentation sera utilisée comme configuration de base dans la voûte de réalité virtuelle de l'Institut Philippe-Pinel, afin de diminuer la gêne ressentie par les patients et d'augmenter la durée des séances en immersion, mais également d'accroître le sentiment de présence et d'améliorer la perception réaliste de la scène virtuelle.

1.3 Objectifs de recherche

L'objectif général de ce projet est de concevoir et d'évaluer en termes de confort et d'efficacité une approche ajustant l'affichage stéréoscopique en fonction des mouvements oculaires. Il s'agit en particulier de développer un logiciel permettant de simuler et de comparer plusieurs conditions d'affichage stéréoscopique basées sur la détermination précise du point de regard du participant dans un environnement de réalité virtuelle. Les configurations testées différeront les unes des autres par la distance inter-caméras et la présence ou non de flou de profondeur de champ. Afin de déterminer quelle est la configuration optimale, une expérimentation sera menée auprès de participants, dont les sous-objectifs sont :

- Comparer subjectivement les différentes configurations stéréoscopiques, en termes de confort, réalisme, plaisir, perception des distances et des profondeurs et sentiment d'immersion, afin de déterminer la(les) configuration(s) la(les) plus efficace(s) et générant le moins d'inconfort visuel.
- Comparer objectivement les différentes configurations en déterminant les limites de fusion pour chacune, c'est-à-dire les distances à partir desquelles les participants ne pourront plus fusionner les images stéréoscopiques d'un objet.
- Comparer objectivement la perception des distances et des profondeurs entre scène réelle et scène virtuelle, en mesurant la vergence oculaire et l'écart angulaire à la cible afin de déterminer quelle(s) configuration(s) provoque(nt) les mêmes mouvements oculaires que la scène réelle.

Un objectif annexe de ce projet est de fournir aux cliniciens de l'Institut Philippe-Pinel de Montréal des outils d'analyse du comportement visuel de l'individu en environnement de

réalité virtuelle.

1.4 Plan du mémoire

Ce mémoire est organisé en quatre parties, définies comme suit.

Le chapitre 2 consiste en une revue de littérature qui présente les concepts théoriques et regroupe de nombreuses études menées en lien avec tous les aspects du travail de recherche. En effet, le sujet de cette maîtrise se situe à l'interface entre plusieurs domaines scientifiques. Il comprend en premier lieu une forte composante informatique puisqu'il touche à la réalité virtuelle, la génération de scènes par ordinateurs, l'affichage stéréoscopique et le suivi de regard. Cependant, nous avons également dû considérer des aspects biologiques, en particulier concernant l'anatomie de l'œil et la dynamique des mouvements oculaires, et aborder les applications de la réalité virtuelle en psychologie et psychiatrie. Ce chapitre se clôture par nos objectifs et hypothèses.

Le chapitre 3 présente les aspects méthodologiques de ce projet de recherche. Il retrace les étapes de notre cheminement de pensée, depuis certains tests préliminaires jusqu'à la conception du protocole expérimental final, en passant par le développement de l'application permettant de déterminer le point de regard et d'ajuster l'affichage. Il s'attarde sur les aspects techniques du projet, des technologies employées jusqu'au code développé.

Le chapitre 4 présente les résultats obtenus au cours de l'expérimentation menée auprès de 18 participants. Les résultats sont présentés, détaillés et commentés dans l'ordre de déroulement de celle-ci, à savoir les notations subjectives en premier, puis les indications objectives de performance des configurations : intervalles des limites de fusion et comparaison des angles de vergence oculaire et des écarts angulaires à la cible entre scène réelle et scène virtuelle.

Certaines considérations quant à ces résultats sont discutées dans le chapitre 5, qui présente également les conclusions tirées par rapport aux hypothèses émises.

CHAPITRE 2

REVUE DE LITTÉRATURE

2.1 Principes de la vision humaine

Afin de comprendre les enjeux et l'intérêt de suivre le regard d'un individu, il est important de se pencher plus en détail sur le fonctionnement de la vision humaine. Cette partie aborde le sujet à la fois d'un point de vue physiologique – en expliquant les mécanismes intervenant dans les mouvements oculaires – et d'un point de vue psychologique – en traitant de l'attention visuelle.

2.1.1 Caractéristiques rétinienne

La membrane tapissant le fond de l'œil, la rétine, est dotée de nombreux photorécepteurs, les cônes et bâtonnets, qui convertissent les informations lumineuses en impulsions électriques acheminées au cerveau par le nerf optique et interprétées en images.

La densité de ces cellules, et donc l'acuité spatiale visuelle, varie en fonction de la zone rétinienne. Ainsi la fovéa compte 100 000 cônes pour un diamètre de 1500 μm , tandis que la macula en compte 650 000 pour un diamètre de 5000 μm . Les bâtonnets, eux, sont présents en plus grand nombre sur la rétine périphérique tandis que la concentration en cônes diminue fortement lorsqu'on s'éloigne de la fovéa (Ganong, 2005). Les mouvements oculaires qui seront détaillés en section 2.1.2 nous permettent de projeter des zones particulières de notre champ de vision sur la fovéa, afin de les voir en haute résolution. À noter que celle-ci n'est pas située dans le prolongement exact des centres de la pupille et de l'œil, comme on peut le voir sur la figure 2.1, et qu'ainsi axe optique et axe visuel ne sont pas confondus.

Notre champ visuel total est d'environ 180° horizontalement et 130° verticalement. Étant donné qu'1° d'angle correspond approximativement à 300 μm de distance sur la rétine (DeValois and DeValois, 1988), la vision fovéale couvre environ 5° du champ visuel, avec au centre 2° d'acuité maximale. Elle ne permet donc de scruter qu'une toute petite portion du champ visuel, par exemple 3 % de la taille d'un large écran (53,34 cm¹) observé à environ 60 cm² de distance (Duchowski, 2007). La zone projetée sur la macula est, elle, d'environ 16,7°, tandis que le champ visuel « utile » s'étend jusqu'à approximativement 30°. Au-delà, nous ne distinguons presque plus les objets, mais pouvons toujours percevoir les mouvements.

1. 21 po.

2. 23,62 po.

Outre l'acuité spatiale, les caractéristiques de la rétine influent sur un autre aspect important de la vision humaine, la vision temporelle (Duchowski, 2007). Elle repose sur deux phénomènes distincts : la persistance rétinienne et le phénomène phi (Gregory, 1997), le premier faisant référence à la fréquence d'échantillonnage de l'information visuelle par la rétine, le second à sa tolérance pour des discontinuités modérées dans la position d'un objet en mouvement. Grâce à cela, un individu percevra le mouvement présent dans une succession d'images fixes – à la télévision, au cinéma, ou dans une voûte de réalité virtuelle – de la même façon qu'il le ferait dans la réalité.

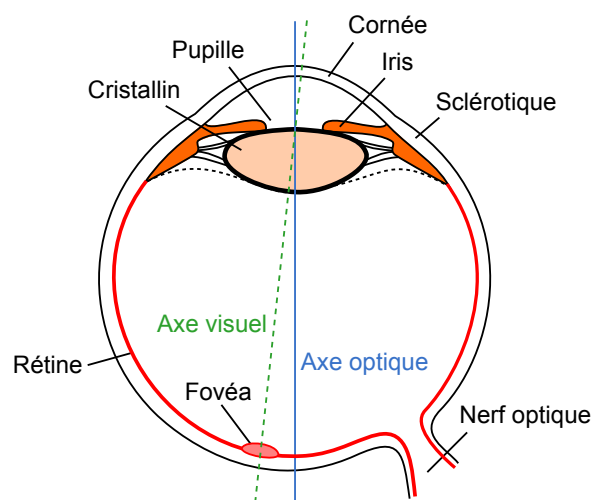


Figure 2.1 Anatomie de l'œil humain.

2.1.2 Mouvements oculaires

Fixations

Les fixations stabilisent la rétine sur un objet immobile. Elles sont caractérisées par des mouvements miniatures involontaires, les micro saccades, ayant une amplitude variant de 1' à 2' et permettant à l'image projetée sur la rétine de rester nette. Nous consacrons 90 % du temps de vision à fixer des éléments de notre environnement visuel (Duchowski, 2007).

Saccades

Les saccades sont des mouvements rapides, volontaires ou réflexes, que nous effectuons à raison de 3 à 5 fois par seconde (Lorenceanu, 2012) afin de projeter de nouvelles zones de l'environnement visuel sur la fovéa. Le temps de déplacement de l'œil est d'environ 10 à 100 ms, temps pendant lequel la personne est « aveugle » (Shebilske and Fisher, 1983) en raison

du taux de rafraîchissement rétinien. Les saccades sont supposées balistiques, c'est-à-dire que la destination est préprogrammée, ce qui signifie qu'une fois que le mouvement a été calculé jusqu'à la prochaine zone de fixation, celui-ci ne peut pas être altéré (Duchowski, 2007).

Mouvements de poursuite douce

Les mouvements de poursuite douce interviennent lorsque nous suivons un objet qui se déplace, comme une voiture ou une personne qui marche. Dépendamment de la vitesse de la cible, les yeux sont capables de s'y adapter de manière à ce que l'image ne s'écarte pas de la fovéa (Duchowski, 2007), résultant en un mouvement continu et fluide, c'est-à-dire sans saccades.

Ces différents mouvements peuvent être interprétés en terme d'attention visuelle, comme expliqué en section 2.1.5.

2.1.3 Vision stéréoscopique

À partir des deux images 2D formées sur les rétines, illustrées en figure 2.2, le cerveau est capable d'évaluer la profondeur de notre environnement visuel : il s'agit de la vision stéréoscopique, ou binoculaire (DeAngelis, 2000). Pour cela, il s'appuie sur les disparités horizontales (et verticales dans une moindre mesure), c'est-à-dire les différences de position entre des points correspondants sur les images rétiniennes causées par l'espacement entre les deux yeux. Ces disparités suffisent à elles seules à percevoir la profondeur, néanmoins d'autres indices monoculaires ou binoculaires (parallaxe, occlusions, tailles et déplacements relatifs, perspective, ombres, etc.) peuvent les compléter (DeAngelis, 2000, Lipton, 1997). En réalité virtuelle, la disparité horizontale est l'angle entre l'œil gauche, l'objet virtuel et l'œil droit. Elle est liée à la distance sur l'écran entre un point de l'image gauche et son correspondant dans l'image droite, ainsi qu'à la position de l'observateur (Leroy et al., 2012).

Les disparités ne sont souvent appréhendées consciemment par l'individu que dans les cas extrêmes, lorsqu'il regarde au loin tout en plaçant son pouce très près de son visage par exemple. Les deux images du pouce observées correspondent à la projection de ce dernier sur l'une et l'autre des rétines (Blum et al., 2010). C'est en particulier dans cet espace proche que les disparités verticales interviennent. Elles sont dues au fait qu'un objet excentré par rapport au plan médian de la tête apparaît plus grand à l'œil dont il est le plus proche. Pouvant atteindre 12 % pour un objet à une distance frontale de 20 cm et 2 % à 140 cm, elles ne servent que d'indicateurs supplémentaires qui modifient la perception des formes, tailles et distances induites par les disparités horizontales (Garding et al., 1995, Perez-Bayas, 2001).

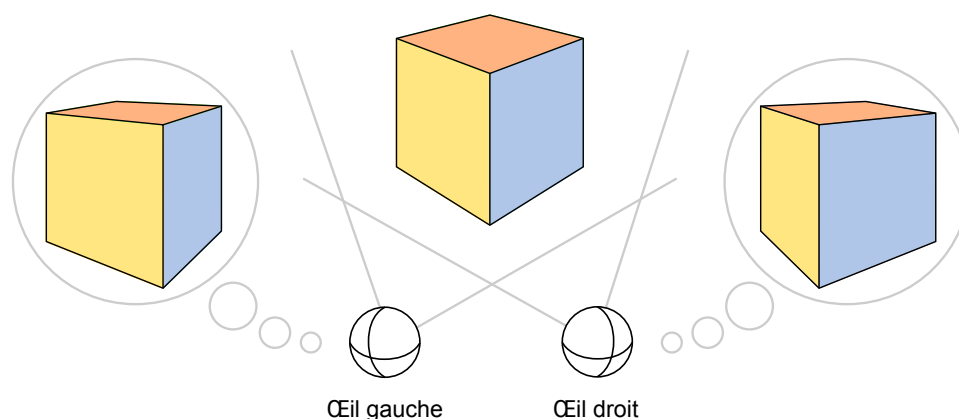


Figure 2.2 Vision binoculaire.

Distance interpupillaire

La distance interpupillaire réfère de manière ambiguë dans la littérature à l'écart entre les centres des deux pupilles lorsque les yeux fixent un objet à l'infini (Office québécois de la langue française, 2002a), ou à celui-ci en prenant en compte la convergence oculaire. Dans le reste de ce mémoire, c'est la première définition qui est employée, sauf précision contraire. Pour la vaste majorité des adultes, elle se situe entre 50 mm et 70 mm, avec une moyenne et une médiane d'approximativement 63 mm (Dodgson, 2004, Lambooi et al., 2007).

Convergence et accommodation

Les différentes catégories de mouvements oculaires détaillées en section 2.1.2 s'accompagnent toutes de mouvements de vergence et du phénomène d'accommodation. Un mouvement de vergence correspond à la rotation des deux yeux dans des directions opposées afin de localiser la région d'intérêt sur la fovéa. Il s'agit d'un procédé principalement actionné par la disparité (Takeda et al., 1999). Près de 70% de l'intervalle de vergence normale d'un individu est utilisé dans un rayon d'1 m. L'angle de vergence, soit l'angle existant entre les deux directions de vue, varie d'environ 14° lorsque le point de regard passe de l'infini à la distance de convergence confortable la plus proche située aux alentours de 25 cm, et de 36° quand il passe au point le plus proche sur lequel les yeux peuvent converger (Daugherty et al., 2010). La figure 2.3 illustre les variations de cet angle liées aux changements de profondeur du point de regard.

Le phénomène d'accommodation réfère quant à lui à l'étirement plus ou moins important du cristallin permettant au focus de s'effectuer à la bonne profondeur de champ, procédé

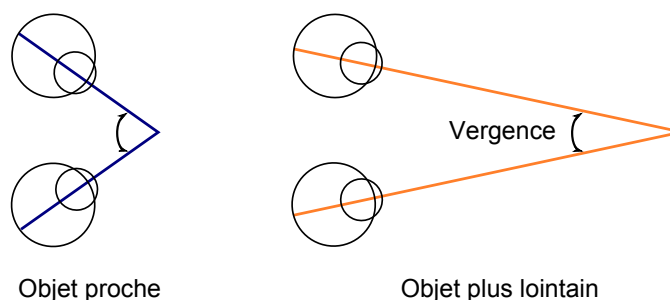


Figure 2.3 Mouvements de vergence.

actionné par le flou (Takeda et al., 1999). En raison du diamètre de la pupille, les objets qui sont à une profondeur différente du point fixé (et en dehors de l’horoptère, voir la section 2.1.3) apparaîtront floutés. Pour reprendre l’exemple précédent où l’on place son pouce près du visage tout en regardant au loin, les deux images du pouce seront perçues floues (Blum et al., 2010). À ce flou de profondeur de champ s’additionne celui de vision périphérique, dû uniquement aux différences de densités des capteurs de la rétine vues en section 2.1.1.

Bien que l’interconnexion de ces deux phénomènes soit assez complexe³, dans des conditions normales de vue l’accommodation et la vergence varient de manière synchrone et en fonction de la distance à l’objet d’attention (Wann et al., 1995). Elles sont intrinsèquement et réflexivement liées (Rushton and Riddell, 1999). *A contrario*, dans un environnement de réalité virtuelle, les yeux doivent maintenir l’accommodation au niveau du dispositif d’affichage, tandis que la présence d’indices de disparité nécessite des mouvements de vergence (Wann et al., 1995), ce qui conduit au conflit accommodation-convergence.

Horoptère et fusion stéréoscopique

L’horoptère, ou cercle de Vieth-Müller (DeAngelis, 2000), présenté en figure 2.4, désigne pour un degré de convergence donné la ligne qui connecte tous les points de l’espace se projetant identiquement sur les rétines gauche et droite (Perez-Bayas, 2001). Une disparité croisée (parallaxe négative) caractérise alors les objets situés devant cette ligne, une disparité décroisée (parallaxe positive) ceux situés derrière (Lambooij et al., 2007).

Zone de fusion de Panum La zone de fusion de Panum est une extension en profondeur de l’horoptère, où la petite quantité de disparité binoculaire rend encore possible la fusion

3. Une réaction d’accommodation en l’absence de convergence, par exemple lorsqu’un œil est occulté, peut induire une réponse de convergence concurrente, la convergence accommodative (AC en anglais). De même, une réponse de convergence va provoquer une réponse accommodative concurrente appelée accommodation de convergence (CA en anglais). Les ratios AC/A et CA/C sont des indices des degrés de convergence accommodative et d’accommodation de convergence (Fukushima et al., 2009).

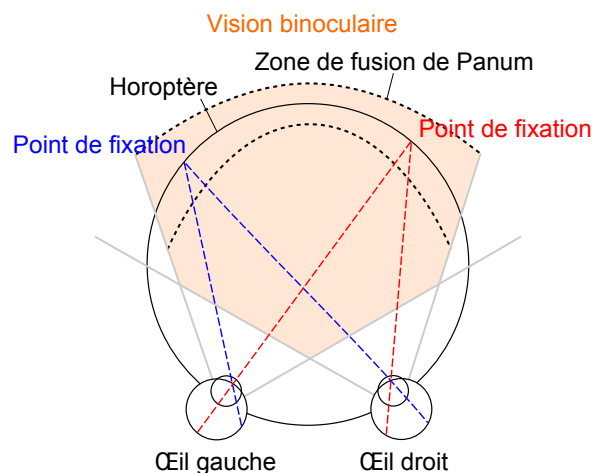


Figure 2.4 Horoptère et zone de fusion de Panum.

des deux images sans mouvements de vergence (mais au prix de quelques efforts qui peuvent susciter de la fatigue visuelle, Leroy et al., 2012). Au-delà des limites de cette zone, les deux images ne peuvent plus être fusionnées et sont perçues séparément : il s'agit du phénomène de diplopie (Perez-Bayas, 2001). Les angles maximums de disparité varient d'un individu à l'autre et peuvent être extrêmement petits au regard du champ visuel humain. Yeh and Silverstein (1990) ont mesuré, en fonction de conditions plus ou moins contraignantes, des angles de $27'$ à $4,93^\circ$ en croisé, et de $24'$ à $1,57^\circ$ en décroisé. Ils varient également en fonction de l'excentricité par rapport à la fovéa : Lambooij et al. (2009a) mentionnent $0,1^\circ$ au niveau de celle-ci, $0,33^\circ$ à une excentricité de 6° et $0,66^\circ$ à une excentricité de 12° .

Distance limite de fusion Il existe une limite de distance en dessous de laquelle, même situé sur l'horoptère, un objet ne pourra être fusionné. En rapprochant par exemple son pouce de son visage, un individu finira fatalement par expérimenter de la diplopie. En réalité virtuelle, ce cas correspond au moment où une parallaxe excessive conduit à une image qui apparaît floue et qu'il est impossible de fusionner, probablement en raison de la dissociation trop importante entre accommodation et convergence (Woods et al., 1993).

Jones et al. (2001) mentionnent une expérimentation menée afin de déterminer les limites de fusion, dans laquelle les participants étaient placés à 0,7 m d'un écran LCD autostéréoscopique de $35,05\text{ cm}^4$. Pour la condition expérimentale la moins contraignante, les limites au-delà de l'écran étaient généralement comprises entre 2 m et 20 m, celles devant l'écran entre 0,3 m et 0,5 m. L'expérimentation de Woods et al. (1993) consistait quant à elle à afficher un tore de 4 cm de diamètre selon deux protocoles :

4. 13,8 po.

- Parallaxe horizontale croissante : le tore est affiché au niveau de la surface de l'écran, puis la parallaxe est augmentée graduellement de manière croisée (derrière l'écran) ou décroisée (devant l'écran) jusqu'à la perte de fusion de l'observateur.
- Parallaxe horizontale décroissante : le tore est affiché avec une parallaxe croisée ou décroisée de la taille de la largeur de l'écran, progressivement réduite jusqu'à ce que l'observateur puisse fusionner les images.

Les participants, au nombre de 10, s'asseyaient approximativement à 0,8 m du moniteur de 40,64 cm⁵, chaque mesure étant réalisée au moins 3 fois. Les résultats révèlent une grande variabilité inter participants, et suggèrent que l'intervalle de profondeur augmente avec une exposition plus importante aux dispositifs stéréoscopiques. Dans le but de présenter une scène stéréoscopique à un maximum d'individus, les auteurs suggèrent donc de minimiser l'intervalle de profondeur totale de celle-ci et de placer les objets d'intérêt proches de l'écran. Ils préconisent également de minimiser la parallaxe verticale afin d'augmenter le confort, leur expérimentation avec parallaxe verticale croissante ayant indiqué que celle-ci devait rester inférieure à 7 mm pour que la fusion soit toujours réalisable.

2.1.4 Fatigue et confort visuels

Les termes de fatigue et de confort visuels sont souvent employés sans distinction dans la littérature. Pourtant, la notion de pression oculaire impliquée par le premier est mesurable en tant que diminution de la performance du système visuel, tandis que le confort est, lui, subjectif (Lambooij et al., 2009b).

L'exposition, en particulier prolongée, à un système stéréoscopique est une source possible de ces maux. Les symptômes les plus courants de fatigue vont alors d'une vision double à des maux de tête persistants, en passant par une sensation de lourdeur des yeux (Leroy et al., 2012). Lambooij et al. (2007) citent comme causes classiques une parallaxe binoculaire excessive, en particulier lorsqu'elle est source de diplopie (Blum et al., 2010), et le conflit accommodation-convergence (défini dans la section 2.1.3). Selon les auteurs, ces deux facteurs apparaissent cependant être « d'importance mineure lorsque les valeurs de disparité ne dépassent pas une certaine limite d'angle visuel, qui fournit tout de même un intervalle de profondeur satisfaisant »⁶. Dans cet intervalle, les causes avancées sont alors principalement des mouvements rapides en profondeur du point de regard – conduisant à une demande excessive de la relation accommodation-convergence – et des intensités de flou peu naturelles, trop importantes ou trop faibles – pouvant causer une perception non naturelle et ambiguë de la profondeur. Une image entièrement nette peut par exemple réduire la zone de fusion et

5. 16 po.

6. Notre traduction.

susciter des conflits d'indices de profondeur (Lambooij et al., 2007).

En raison du fait que l'accommodation s'effectue toujours au niveau de l'écran, il semble qu'il soit impossible de prévenir l'inconfort visuel lorsque l'environnement virtuel présenté à l'observateur s'étend sur un large intervalle de profondeur. Cependant, il est possible de le diminuer en jouant sur les paramètres stéréoscopiques et le flou de profondeur de champ.

2.1.5 Introduction à l'attention visuelle

« La vision humaine est un processus fragmenté qui requiert l'intégration perceptuelle de petites régions afin de construire une représentation cohérente de l'ensemble » ⁷ (Duchowski, 2007). Étant donné que les yeux sont des canaux sensoriels qui disposent d'une capacité limitée, appréhender la totalité du champ visuel nécessite de laisser vagabonder le point de regard. Lors de cette inspection, qui n'est pas complètement aléatoire mais partiellement déterministe (Doll et al., 1993), certaines régions attirent l'attention. Elles sont initialement perçues de manière parafovéale, et guident la direction de la prochaine saccade afin de pouvoir identifier sémantiquement l'objet présent dans cette région. C'est donc l'attention, que le grand dictionnaire terminologique définit comme la « concentration de l'esprit, parmi un ensemble de stimulus, sur ceux qui présentent un aspect intéressant ou marquant [...] » (Office québécois de la langue française, 2002b), qui sert de filtre de sélection des informations visuelles.

Les mouvements oculaires décrits plus haut (voir la section 2.1.2) peuvent alors être interprétés en termes d'attention visuelle : les fixations correspondraient naturellement au désir de maintenir son regard sur un objet d'intérêt et les saccades au changement volontaire de point de focus de l'attention (Buswell, 1935, Yarbus, 1967). Suivre et enregistrer le point de regard d'un individu permet donc de savoir ce qu'il a trouvé intéressant dans une scène, et même d'obtenir des informations sur la façon dont il l'a perçue (Duchowski, 2007). À noter cependant l'existence de l'attention para fovéale (la concentration consciente et volontaire sur un objet présent en vision périphérique sans déplacer les yeux vers cet objet), ou le fait qu'il est possible de « regarder dans le vide » (fixer une région de l'environnement visuel tout en pensant à autre chose). Malgré tout, ces cas demeurent rares lors de l'inspection visuelle d'une scène et le point de regard peut être supposé correspondre globalement au point d'attention.

7. Notre traduction.

2.2 Suivi oculaire : traitement des données

Il existe aujourd'hui de nombreux appareils et méthodes permettant d'effectuer du suivi oculaire : électro-oculographie, lentilles de contact et fluxmètre, photo-oculographie et vidéo-oculographie, etc., les plus précis étant en général les plus intrusifs (Duchowski, 2007). Cette partie examine le traitement des données acquises par ces dispositifs, depuis la détermination du point de regard 3D jusqu'à l'interprétation du signal en mouvements oculaires.

2.2.1 Traitement des images des yeux

Les techniques de photo-oculographie et vidéo-oculographie, qui consistent à photographier ou filmer les yeux, nécessitent l'application d'algorithmes de traitement afin de localiser ces derniers dans l'image puis d'extraire les informations nécessaires à la détermination de la direction de vue.

Concernant le premier point, une méthode courante consiste à parcourir l'image entière avec une fenêtre rectangulaire dont la taille correspond à celle de la région d'un œil. La sous-image extraite est projetée dans l'espace des « yeux propres⁸ », obtenu en réalisant une analyse en composantes principales. En anglais *principal component analysis* sur une base de données d'images d'œil. La sous-image retenue est celle dont la mesure de distance avec sa projection est la plus petite (Talmi and Liu, 1999).

La détermination de la direction de vue s'effectue quant à elle typiquement avec la méthode dite « Centre de la pupille - reflets cornéens⁹ », basée sur la relation géométrique entre le centre de la pupille et la première image de Purkinje¹⁰ (Kwon et al., 2006, Talmi and Liu, 1999), dont un exemple est donné en figure 2.5. Kwon et al. supposent qu'il existe une relation linéaire joignant la distance entre ces éléments et le point de regard. À l'aide d'une caméra monoculaire, leur système estime par traitement d'image la position de la pupille et des reflets cornéens produits par deux diodes électroluminescentes infrarouge. Les valeurs de distances entre ces points, couplées à une étape de calibration pour chaque individu où ces informations sont enregistrées alors qu'il regarde les coins supérieur gauche et inférieur droit d'un écran, permettent de déterminer le point de regard (x,y) pour chaque œil sur l'écran. La précision atteinte, évaluée sur 4 participants à qui il était demandé d'observer 16 points, est à peu près uniformément distribuée entre 0,4 et 0,5° (aucune n'étant supérieure à 0,6°).

Une autre méthode consiste à mesurer la proportion entre l'iris, sombre, et la sclérotique, plus claire (voir la figure 2.1). Le modèle utilisé pour l'iris est généralement un cercle de

8. En anglais *eigeneye*.

9. En anglais *pupil center - corneal reflections*.

10. Les images de Purkinje sont les réflexions de la lumière sur les différentes couches structurales des yeux, la première image correspondant à la réflexion cornéenne.

rayon et de position variable, tandis que les bords de la sclérotique sont modélisés par des arcs d'ellipse. La variation du ratio des aires à gauche et à droite de l'iris indique alors des changements horizontaux de direction du regard (Talmi and Liu, 1999).

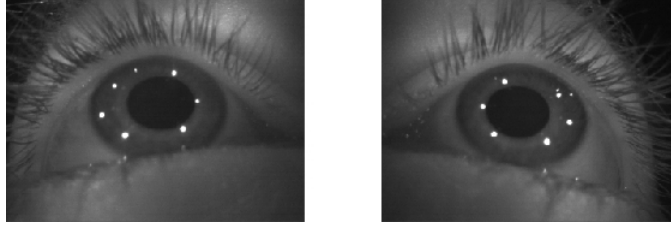


Figure 2.5 Reflets cornéens obtenus avec les ETG.

2.2.2 Détermination du point de regard 3D

Les techniques décrites ci-dessus sont des préalables à l'obtention du point de regard 3D : à partir des données de position 2D du regard sur l'écran ou des vecteurs de direction de vue calculés, il reste à intégrer l'information de profondeur afin de déterminer où l'observateur a porté son attention dans la scène virtuelle en 3 dimensions.

Profondeur des pixels

Il s'agit pour cette méthode de déduire la profondeur à l'aide de celles des pixels de l'image, stockées par exemple dans un tampon de profondeur¹¹ (Blum et al., 2010). Hillaire et al. (2008a) ont utilisé un système de suivi fournissant le point de regard 2D sur l'écran, c'est-à-dire les coordonnées (x, y) du pixel correspondant, avec cependant une précision supérieure à la taille d'un pixel. La profondeur du point de regard 3D est alors obtenue en calculant la moyenne pondérée, en fonction de la valeur sémantique et de l'éloignement du centre, des profondeurs des pixels d'une zone carrée centrée sur (x, y) . Le vecteur de direction de vue, exprimé dans le repère de la caméra à l'aide de ces coordonnées, est ensuite transformé dans le repère de l'environnement virtuel. La position du point de regard p dans ce repère est alors obtenue grâce à l'équation suivante :

$$p = c + vxd \quad (2.1)$$

où c la position de la caméra, v le vecteur vue dans le repère virtuel et d la profondeur calculée précédemment.

11. En anglais *Z-buffer*.

Géométrie de la scène : *raycasting*

Ici, le point de regard 3D est obtenu en lançant le rayon correspondant à un des axes optiques de l'individu dans la scène et en calculant son intersection avec celle-ci comme dans une méthode de *raycasting*. Pour ce faire, le vecteur de direction de vue ainsi que la position de sa base doivent être exprimés dans le repère monde, en combinant par exemple le suivi oculaire avec le repérage de la tête de l'individu et une procédure de calibration. Le postulat est ensuite que le premier objet intersecté par l'axe ainsi défini est l'objet d'attention, le point de regard étant situé au niveau de l'intersection tel que l'illustre la figure 2.6.

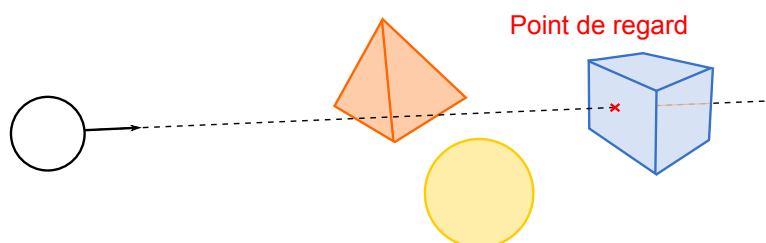


Figure 2.6 Point de regard défini en intersectant l'axe optique avec la scène virtuelle.

Pfeiffer (2012) soulignent que cette méthode présente l'avantage de pouvoir être utilisée non seulement dans un environnement virtuel mais également réel, à condition d'avoir un modèle géométrique précis de la scène environnante. De plus, un suivi monoculaire est suffisant, préférentiellement de l'œil dominant, et les estimations de point de regard obtenues se révèlent assez précises. Néanmoins, ils notent certaines limites. En particulier, le postulat considérant que le premier objet touché est l'objet d'attention ne prend pas en compte les fenêtres ou les miroirs. De plus, en lieu et place d'un rayon qui se révèle trop restrictif en raison de la possibilité d'attention para fovéale, un cône devrait être utilisé pour représenter la zone d'attention de l'individu, son angle d'ouverture correspondant par exemple à l'angle de haute acuité visuelle. D'autres cas problématiques sont ceux d'un objet dont la taille est plus petite que la précision de l'estimation (comme la pointe d'une aiguille), ou d'un objet ciblé par l'un des yeux mais occulté par un autre élément à la vue de l'autre.

Les approches précédentes peuvent être réalisées avec des systèmes de suivi monoculaires, tandis que celles détaillées ci-dessous nécessitent des systèmes binoculaires car elles s'appuient sur la relation entre les deux vecteurs de direction ou entre les positions des deux pupilles. Il est important de noter que des expérimentations ont prouvé l'existence de mouvements de vergence appropriés en réponse à des stimuli stéréoscopiques, malgré le conflit accommodation-convergence, rendant ces méthodes légitimes (Duchowski et al., 2011, Dagher et al., 2010, Wann et al., 1995).

Intersection des lignes de vue

Cette approche, qui requiert de connaître les vecteurs de direction des deux yeux, consiste à considérer que le point de regard 3D est situé à l'intersection des axes optiques, ou plus vraisemblablement s'ils ne sont pas concourants au milieu du segment de longueur minimale qui les relie, comme on peut le voir en figure 2.7. Pfeiffer (2012) qualifie cette méthode d'holistique. Elle estime le point de regard en utilisant uniquement des informations sur l'observateur et dépasse ainsi certaines des limitations de la méthode précédente. Néanmoins, les auteurs soulignent qu'elle requiert des efforts techniques plus importants, par exemple l'utilisation d'un système de suivi binoculaire ou éventuellement de disposer les points de calibration sur un volume plutôt qu'un plan, faisant ainsi accroître leur nombre de 9 à 27, de 16 à 48 ou plus encore. De plus, les erreurs et le bruit présents dans les mesures de vecteur de direction, qui impactent fortement l'estimation du point de regard, peuvent conduire à une mauvaise précision (Essig et al., 2006).

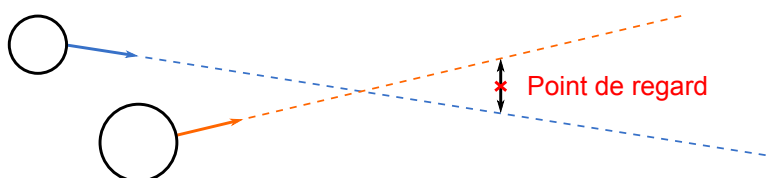


Figure 2.7 Point de regard défini comme milieu du plus court segment reliant les deux lignes de vue.

Afin d'améliorer cette dernière, Essig et al. (2006) ont mis au point une technique de calibration basée sur un réseau de neurones (dont les détails sont donnés en section 2.2.3). Les 8 participants de leur expérimentation devaient ensuite regarder 16 marqueurs placés sur 4 plans d'un cube, situés à 11 cm, 3,67 cm, -3,67 cm et -11 cm de l'écran. C'est la distance entre la position du point de regard calculé et la position exacte du marqueur qui servait de mesure de précision. Ils ont obtenu une erreur moyenne totale en x, y et z de 2,78 cm, contre 6,18 cm avec l'autre méthode de calibration testée, en notant dans les deux cas une erreur significativement plus importante pour la coordonnée z (en moyenne 4,16 cm contre 0,97 cm et 1,03 cm en x et y).

Angle de vergence

Cette méthode détermine puis utilise la relation entre l'angle de vergence et la profondeur du point de regard. En effet, la variation de cette profondeur s'accompagne de mouvements de convergence oculaire, donc de l'altération de l'angle entre les deux vecteurs de direction.

Duchowski et al. (2011) ont mesuré, à l'aide d'un système de suivi oculaire, les réponses de vergence d'un individu tandis qu'il observait en monoscopique ou stéréoscopique des objets virtuels situés à différentes profondeurs. Chaque ligne de la grille 5 x 5 de cubes de 2 cm de côté utilisée était affichée à 12 cm de la précédente, la première étant située à 30 cm devant l'écran. La comparaison des mesures empiriques de profondeur du point de regard, fournies par le système de suivi à partir de la vergence, et les profondeurs des cibles dans le repère réel ont montré que les premières correspondaient bien aux secondes dans le cas stéréoscopique. Duchowski et al. (2011) affirment « même que l'étendue de la vergence correspond à la profondeur de la cible regardée ». Cependant, ils ont dû pour arriver à ce résultat lisser leurs estimations grâce à un filtre de Butterworth de degré 6, puis ajuster la courbe entière à celle des profondeurs affichées à l'aide d'une méthode des moindres carrés, afin de contrer le haut niveau de bruit ainsi que les sous-estimations ou surestimations de la profondeur par l'observateur. En particulier, ils ont remarqué un décalage significatif entre ces deux courbes et ce, même dans le cas monoscopique : la profondeur moyenne estimée n'était pas située au niveau de l'écran mais oscillait entre 10 et 20 cm devant celui-ci.

Les résultats obtenus par Daugherty et al. (2010) lors d'une expérimentation similaire sont également mitigés. Pourtant, et afin de simplifier leur étude, ils ont considéré l'écart interpupillaire des participants et leur distance à l'écran fixes (respectivement 6,3 cm et 50 cm), et se sont limités à une vergence symétrique. Grâce à ces assertions, et en utilisant un système de suivi fournissant les coordonnées des points de regard gauche et droit sur l'écran, ils calculaient la disparité angulaire, c'est-à-dire la variation dans l'angle de vergence lorsque le point de regard passe de la profondeur de l'écran à une autre. Après un lissage nécessaire des données oculaires, les moyennes temporelles de disparités angulaires pour les 12 participants durant l'observation d'un disque apparaissant à 3 profondeurs distinctes étaient respectivement de 0,82, 0,26 et 0,31. Bien que la différence soit significative entre les disparités mesurées pour le plan de devant et celles des deux autres, ce n'est pas le cas entre le plan moyen et l'arrière-plan. De plus, les mesures empiriques ont montré que la supposition d'une disparité nulle au niveau de l'écran n'était pas toujours vérifiée, ou éventuellement cachée par le bruit dans les données de suivi.

Distance interpupillaire

Il s'agit pour cette méthode d'estimer la profondeur du point de regard 3D en évaluant la distance interpupillaire¹², qui augmente ou diminue en fonction que l'objet d'attention s'éloigne ou se rapproche des yeux comme on peut le voir en figure 2.8. Kwon et al. ont testé cette approche. La première étape de leur algorithme consiste à inférer la droite de

12. En anglais *pupil center distance*.

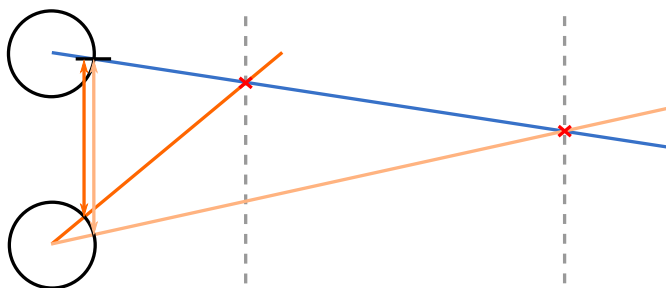


Figure 2.8 Variations de la distance interpupillaire liées aux changements de profondeur du point de regard.

direction de vue à partir du point de regard 2D obtenu avec une méthode PCCR, la seconde à déterminer la profondeur d le long de cette droite grâce à la valeur de DIP. Ils remarquent avant tout que la fonction des variations théoriques $f(\text{DIP}) = d$, obtenue expérimentalement en utilisant une simulation simple du globe oculaire, n'est pas linéaire. Ils ont appliqué leur méthode aux interactions regard-dépendants avec des objets d'un environnement virtuel. Une calibration permettait de définir la fonction f pour chacun des 3 participants. Ils devaient ensuite sélectionner, en les fixant, des fléchettes affichées sur un écran stéréoscopique à barrière de parallaxe. Celles-ci étaient disposées aléatoirement dans un espace virtuel divisé en 5 plans parallèles à l'écran (placés à des profondeurs de 500 mm, 630 mm, 840 mm, 1270 mm et 2600 mm et scindés en 4 x 3 régions). La précision de leur système fut déterminée par le pourcentage de réussite de sélection, soit 97%, 93% et 93% sur 30 essais. Il n'a cependant été testé que pour une partition discrète de l'espace virtuel, tout comme la technique précédente basée sur les angles de vergence. Par exemple, la DIP du premier participant était comprise entre 750 et 752 pour le plan à 500 mm, entre 753 et 755 pour celui à 630 mm, etc.

2.2.3 Mesure de qualité des données

Les grandeurs usuelles de comparaison des systèmes de suivi oculaires et des techniques de détermination du point de regard 3D sont la précision, qui réfère à la différence entre la mesure et la donnée réelle, par exemple de la position du point de regard, et la stabilité qui correspond au bruit que l'on observe dans les mesures alors que le regard demeure parfaitement fixe. Il est possible d'améliorer les performances d'un système à ces deux niveaux, en particulier en effectuant une procédure de calibration pour le premier et en lissant les données pour le second. Holmqvist et al. (2012) soulignent qu'actuellement la manière la plus commune de rendre compte de la précision d'un système de suivi de regard est de se référer aux spécifications données par le constructeur, bien qu'il soit possible d'effectuer ses propres mesures.

Procédure de calibration

La calibration des systèmes de suivi oculaire est une étape essentielle afin de s'adapter aux différences entre les participants ou entre les sessions d'expérimentations (par exemple au niveau de la position relative de la tête par rapport à la caméra). Elle consiste la plupart du temps à demander à l'individu d'observer successivement n points de calibration disposés sur une grille 3D rectiligne dans l'espace réel (Talmi and Liu, 1999, Essig et al., 2006, Kwon et al., 2006). Soit $P = p_i$ l'ensemble des vecteurs mesurés par le système, $Q = q_i$ l'ensemble des vecteurs « corrects » correspondants, il s'agit de construire la fonction d'interpolation $f : \mathbb{R}^3 \mapsto \mathbb{R}^3$, telle que $f(p_i) = q_i \forall i = 1 \dots n$ (Zachmann et al., 2000). Une fois définie, l'application de cette fonction sur un nouveau vecteur mesuré p fournit le vecteur q corrigé correspondant.

Dans le cas de la méthode basée sur les multi-quadriques de Hardy (Zachmann et al., 2000), utilisée par exemple pour la calibration des systèmes de repérage électromagnétique dans les voûtes de réalité virtuelle, la forme générale de la fonction est

$$f(P) = \sum A_i \omega_i(P), P, A_i \in \mathbb{R}^3 \text{ avec } \omega_i(P) = \frac{1}{\sqrt{(P - P_i)^2 + R^2}}, R > 0$$

Il s'agit d'une méthode globale, c'est-à-dire qu'elle utilise pour la correction d'une donnée l'ensemble des couples (p_i, q_i) enregistrés. Les coefficients A_i sont déterminés en résolvant, selon chacune des coordonnées x, y et z, les systèmes d'équations linéaires résultant de l'expression de $f : \sum_{j=1}^n A_j \omega_j(p_i) = q_i$

Concernant le paramètre R^2 , qui influe sur le lissage de la fonction d'interpolation, bien qu'il n'existe pas de valeur « magique » ou de formule simple et robuste pour le déterminer, les valeurs optimales obtenues expérimentalement sont $10 \leq R^2 \leq 1000$ pour un ensemble de 100 à 200 mesures.

Essig et al. (2006) ont quant à eux mis au point une technique de calibration basée sur des anaglyphes et l'utilisation d'une PSOM, un type de réseau de neurones artificiel. Les participants étaient invités à observer 27 marqueurs situés sur un cube 3 x 3 x 3 dont les plans, parallèles à l'écran, étaient positionnés à 8 cm devant celui-ci, à son niveau, ou à 8 cm derrière. Les enregistrements effectués ne servaient alors pas à déterminer une fonction d'interpolation mais à entraîner une PSOM, chargée ensuite de corriger les vecteurs mesurés dans l'ensemble du volume de stimuli. Comme évoqué en section 2.2.2, ils ont ainsi amélioré la précision de leur système, passant d'une erreur moyenne de 6,18 cm avec une calibration géométrique classique à 2,78 cm, cette diminution étant notablement supérieure pour la coordonnée en z et pour l'arrière-plan.

Débruitage des données

Une étape de prétraitement des données brutes consistant à réduire le bruit peut s'avérer nécessaire afin d'en extraire les informations substantielles, par exemple identifier les mouvements oculaires (voir la section 2.2.4). Celui-ci provient de périodes où les caractéristiques de l'œil utilisées (telles que la pupille et le reflet cornéen avec une méthode PCCR) n'ont pu être détectées et suivies, par exemple lorsque « les lunettes, les lentilles de contact, les cils ou les clignements de paupières empêchent la caméra de capturer clairement l'œil »¹³ (Holmqvist et al., 2012).

Le débruitage du signal consiste la plupart du temps à lui appliquer un filtre afin de le lisser. Duchowski et al. (2011) ont prétraité leurs données empiriques de profondeur du point de regard en éliminant tout d'abord les mesures aberrantes, situées au-delà de 2 fois la valeur de l'écart-type calculé avec tous les points de mesure pour une même profondeur de cible. Ils ont ensuite appliqué un filtre passe-bas de Butterworth de degré 6 afin d'augmenter la stabilité de leur estimation de profondeur sans latence ou perte significative de précision. Hillaire et al. (2008a) ont également utilisé un filtre passe-bas, avec une fréquence seuil de 15 Hz, afin d'éliminer les tressaillements haute fréquence dans les coordonnées du point de regard 2D fournies par leur système de suivi.

Les méthodes de la médiane et de la moyenne consistent quant à elles à remplacer un enregistrement par l'une ou l'autre de ces valeurs calculée sur une fenêtre temporelle, par exemple de 10 images (Daugherty et al., 2010). Cette dernière est généralement centrée sur l'enregistrement, plutôt que située en amont, ce qui permet de réduire le temps de latence introduit et d'augmenter la réactivité du signal lissé aux mouvements oculaires réellement effectués, mais nécessite la connaissance des enregistrements suivants.

Le filtre ROWA peut *a contrario* être utilisé « en ligne » car il détecte et remplace les données bruitées en s'appuyant uniquement sur les mesures précédentes. Comparé à d'autres filtres image et audio existants, en utilisant la mesure de déviation angulaire entre une cible et le point de regard associé, il obtient l'une des meilleures performances en termes de PSNR moyen ; 84,41 dB contre 69,67 dB pour un filtre médian (Chartier and Renaud, 2008). En revanche, si le bruit s'avère trop important, ce filtre retournera après un certain temps une valeur constante.

Enfin, afin de gérer une mauvaise qualité du système de suivi de regard et de maximiser la validité des résultats, il peut s'avérer nécessaire d'exclure certaines mesures. Les critères cités par (Holmqvist et al., 2012) comprennent le pourcentage de valeurs nulles dans les données brutes, une faible précision ou encore un nombre élevé d'évènements ayant une vitesse supérieure à 800°/s.

13. Notre traduction.

2.2.4 Détection des évènements oculaires

Une fois la position 3D du point de regard obtenue, des analyses quantitatives permettent d'étudier plus en profondeur le comportement oculaire de l'individu. C'est en particulier la séquence de positions qui est étudiée, généralement afin de la fractionner automatiquement et si possible de manière non ambiguë en saccades et fixations. Ces dernières s'avèrent particulièrement intéressantes car elles indiquent au mieux les localisations de l'attention de l'observateur.

Duchowski et al. (2002) mentionnent trois approches principales pour détecter ces évènements, deux d'entre elles utilisant l'écart angulaire entre deux positions du point de regard. Pour la première, cet écart est utilisé conjointement à des seuils, temporel et spatial, afin de repérer les portions stationnaires du signal. Le regard doit par exemple demeurer dans une zone de 5° durant plus de 300 ms pour que cette partie du signal soit étiquetée comme fixation.

La seconde se base sur la vitesse du mouvement oculaire, l'écart angulaire étant alors divisé par le temps. Dès lors que la vélocité du signal dépasse un certain seuil, une saccade commence. Les fixations sont alors repérées implicitement, comme les portions de signal entre deux saccades. Les seuils utilisés pour ces méthodes peuvent être déterminés empiriquement, par exemple en se basant sur le fait que la plupart des saccades qui se produisent naturellement ont une amplitude de moins de 15° ou 20° . Il faut cependant tenir compte de la qualité des données de suivi, (Holmqvist et al., 2012) ayant par exemple remarqué que, pour un algorithme de détection basé sur la vitesse avec un seuil adaptatif, les fixations deviennent moins nombreuses et durent plus longtemps lorsque la précision diminue ou lorsque certaines mesures sont perdues. À noter enfin que ces méthodes ne prennent généralement pas en compte les mouvements de poursuite douce, la première les classant vraisemblablement comme des saccades, la seconde comme des fixations.

La troisième approche consiste à projeter les données brutes de position du point de regard dans la scène virtuelle, et à réaliser la recherche de mouvements oculaires au sein de cet espace 3D. Les fixations sont alors associées à des régions d'intérêt de l'environnement virtuel, ce qui permet par exemple de repérer une saccade à la bordure entre deux objets, ce que les approches citées précédemment ne détectent pas si l'angle est très petit.

2.2.5 Applications des systèmes de suivi de regard

Applications diagnostiques

Dans le cas des applications diagnostiques, les mouvements oculaires cités ci-dessus sont utilisés comme indicateurs passifs du regard (Duchowski, 2007). Ils sont simplement enregis-

trés, afin d'être étudiés *a posteriori* et de fournir des preuves quantitatives et objectives du schéma d'attention visuel de l'observateur. Le dispositif de suivi est préférablement dissimulé autant que possible afin de ne pas influencer l'individu. Les mesures statistiques typiques, détaillées en section 2.5.1, incluent entre autres le nombre de fixations et le temps total de fixation sur un objet ou une région d'intérêt.

Applications interactives

Les applications interactives adaptent quant à elles leur réponse en fonction du point de regard de l'observateur, ce qui permet une interaction en temps réel avec celui-ci. L'archétype de ce type d'application est celui où le point de regard remplace le curseur d'une souris afin de désigner ou de cliquer sur des régions de l'environnement : il s'agit d'applications sélectives. Un seuil de durée sur les fixations détectées en temps réel distinguerait par exemple les clics. La connaissance du point de regard peut également être utilisée en rétroaction afin de modifier les paramètres d'affichage stéréoscopique, dans des applications dites dépendantes du regard¹⁴. Par exemple, en raison de la diminution de la densité de capteurs sur la rétine autour de la fovéa, les détails en périphérie du point de regard peuvent être amoindris sans que l'observateur détecte une différence avec l'image d'origine (Reingold et al., 2003). Diminuer la résolution permettrait alors de minimiser la bande passante requise ou d'accélérer le rendu, ce qui se révèle parfois nécessaire dans le cas d'environnements virtuels complexes (McCormick et al., 1997).

La plus grande difficulté à laquelle il faut faire face est alors le temps de latence. En effet, sans pouvoir prévoir où le regard va se porter à l'instant suivant, la plupart des systèmes d'affichage dépendant du regard présenteront un décalage entre le moment où les yeux se déplacent et celui où l'affichage est mis à jour. Ce délai sera proportionnel à la fois au temps de mesure du regard (typiquement 16 ms pour des dispositifs de suivi qui utilisent de la vidéo 60 Hz) et au temps nécessaire par la suite pour calculer le rendu et l'afficher (typiquement 33 ms pour un système qui affiche 30 trames par seconde, Duchowski, 2007). L'utilisateur pourra percevoir des temps de latence à partir de 20 ms, ceux-ci ne se révélant dérangeants que pour des durées plus longues.

Après avoir introduit le principe de la stéréoscopie en environnement immersif, les deux parties suivantes détaillent l'utilisation interactive du point de regard pour ajuster certains paramètres d'affichage. La partie suivante reviendra sur l'aspect diagnostique en abordant les applications dans le domaine de la psychologie et psychiatrie.

14. En anglais *gaze-dependent*.

2.3 Systèmes stéréoscopiques

Les systèmes d’affichage stéréoscopique utilisent le principe de vision binoculaire afin de donner à l’observateur une impression de profondeur alors que les images projetées sont en 2D. Pour cela, deux vues légèrement décalées de la même scène sont générées par ordinateur. Il s’agit ensuite d’acheminer vers chacun des yeux l’image qui lui est destinée, par exemple en utilisant deux écrans distincts dans le cas d’un visiocasque où à l’aide de lunettes de vision stéréoscopique. Dans le cas d’un système passif, l’image affichée comporte les informations gauche et droite, séparées par les lunettes grâce au fait que la lumière émise est polarisée différemment pour chacune. Les lunettes actives comportent, elles, une couche de cristaux liquide qui occulte à haute fréquence l’œil gauche ou l’œil droit en synchronisation avec l’affichage.

Une voûte de réalité virtuelle, ou CAVE, est un système stéréoscopique immersif comportant plusieurs écrans qui entourent l’observateur. Les images y sont projetées de façon à donner l’illusion d’un seul grand écran. Elles sont calculées de manière à préserver une parallaxe cohérente pour l’observateur tandis qu’il se déplace et modifie son point de vue sur l’environnement, afin d’augmenter son sentiment de présence et d’immersion. Pour cela, la position et l’orientation de sa tête au sein de la voûte sont détectées en temps réel par un système de repérage, ainsi qu’éventuellement d’autres appareils tels qu’un gant permettant d’attraper ou de déplacer des objets. Ce type d’environnement a été présenté pour la première fois lors de la conférence SIGGRAPH’92 (École polytechnique de Montréal, 2010-2014)

La génération des images stéréoscopiques s’effectue à l’aide de deux caméras virtuelles légèrement espacées. Le calcul dépend de paramètres tels que l’écart entre ces caméras, leur distance focale, la présence ou non d’un flou de profondeur de champ, etc. Pour la plupart, il n’existe pas de valeur idéale universelle ; ils doivent alors être fixés en fonction de l’observateur, ou modifiés en temps réel pour s’adapter à son point de regard par exemple. Cette section détaille chacun de ces aspects, et mentionne certaines mesures de performance permettant de comparer les configurations.

2.3.1 Systèmes de caméras et d’affichage

Configurations des caméras

Les trois configurations référencées dans la littérature (Woods et al., 1993, Allison, 2004, Stelmach et al., 2003) et récapitulées dans le tableau 2.1 diffèrent par l’orientation des caméras l’une par rapport à l’autre, comme illustré sur la figure 2.9, ainsi que par la symétrie ou l’asymétrie de leurs troncs de projection. Lorsque les caméras sont parallèles, elles sont orientées perpendiculairement à l’écran, celui-ci se retrouvant donc parallèle à leurs plans de

projection. La translation des images pour la seconde configuration s'effectue horizontalement dans ce plan.

Dénomination de la configuration	Orientation	Tronc de projection
Parallèle ¹⁵	Parallèles	Asymétriques
Images translatées ¹⁶	Parallèles	Symétriques
Convergente	Convergentes	-

Tableau 2.1 Les trois configurations du système de caméras.

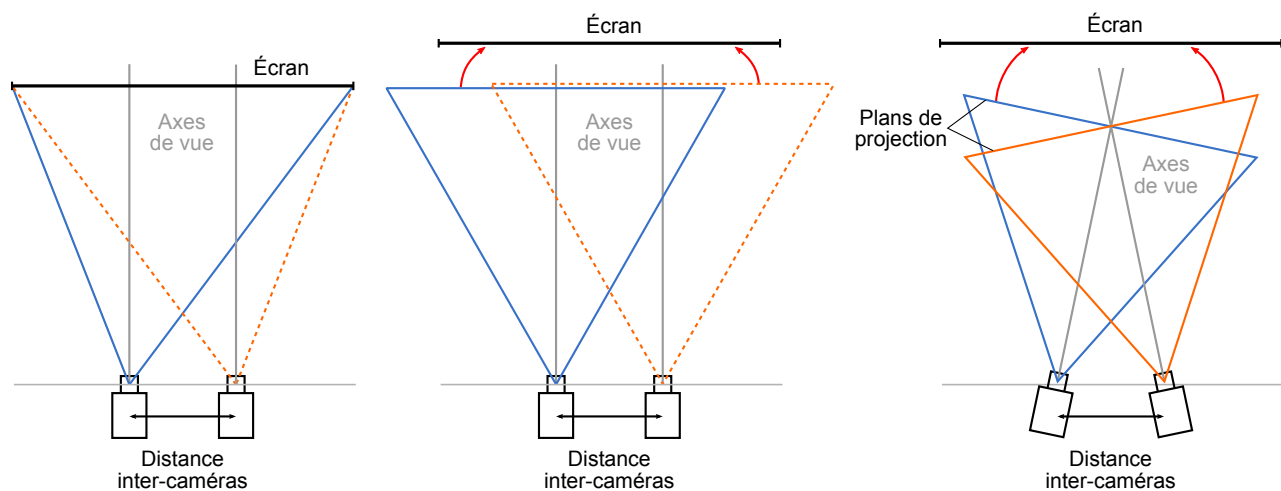


Figure 2.9 Configurations parallèle (à gauche), images translatées (au milieu) et convergente (à droite).

Paramètres des systèmes

La figure 2.10 récapitule les propriétés qui définissent une caméra virtuelle et dont les valeurs diffèrent selon la configuration. Le terme de distance focale (Bourke, 1999a) désigne ici la distance au plan de projection, soit le plan qui intersecte le tronc perpendiculairement à la direction de la caméra et sur lequel les objets de la scène virtuelle sont projetés afin d'obtenir l'image. La valeur de la distance focale est liée avec celles de la hauteur et largeur de ce plan, ainsi que des angles d'ouverture horizontale et verticale de la caméra par les relations décrites dans le tableau 2.2.

Habituellement, lorsque les caméras sont parallèles, elles partagent le même plan de projection. Dans le cas des troncs symétriques, les intersections avec les troncs sont confondues,

$\alpha_{horizontal}$	Cas symétrique	$2 \times \arctan(\frac{largeur}{2 \times distance focale})$
	Cas asymétrique	$\arctan(\frac{largeur + DIC}{2 \times distance focale}) + \arctan(\frac{largeur - DIC}{2 \times distance focale})$
$\alpha_{vertical}$		$2 \times \arctan(\frac{hauteur}{2 \times distance focale})$

Tableau 2.2 Calcul des angles d'ouverture horizontale et verticale de la caméra virtuelle.

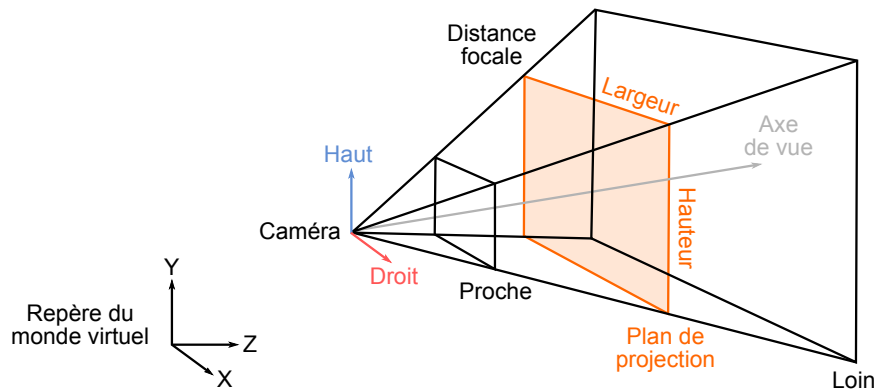


Figure 2.10 Tronc de projection d'une caméra virtuelle.

entre elles et avec l'écran. La disparité entre les projections sera croisée pour des objets situés devant ce plan, décroisée pour ceux situés derrière, comme on peut le voir en figure 2.11. À ces paramètres, indépendants pour chaque caméra, s'ajoutent ceux qui les relient : la distance inter-caméras, et pour le cas convergent, la distance de convergence à laquelle leurs axes de vue s'intersectent (Woods et al., 1993).

Le système d'affichage est déterminé, lui, par l'écart interpupillaire de l'observateur, sa distance à l'écran, et les dimensions de ce dernier (Woods et al., 1993), tel que l'illustre la figure 2.12.

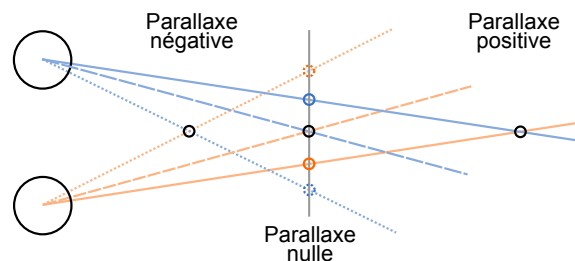


Figure 2.11 Parallaxe négative, nulle ou positive en fonction de la position de l'objet par rapport au plan de projection.

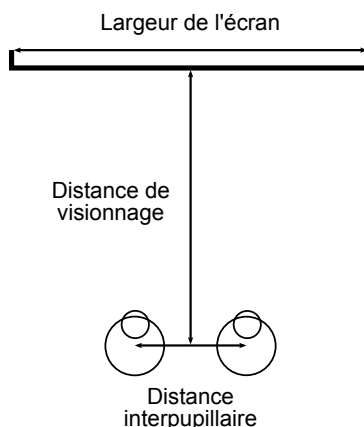


Figure 2.12 Configuration du système d’affichage.

2.3.2 Comparaison des configurations

La question de la configuration la plus performante et des valeurs de paramètres les plus adéquates existe depuis longtemps dans la littérature sur l’imagerie stéréoscopique (Allison, 2004) et divise les chercheurs. Cette section revient sur les différents arguments avancés en faveur et en défaveur des différentes configurations, et réunit certaines expérimentations effectuées afin d’optimiser le réalisme du rendu et de réduire la fatigue visuelle.

Disparités et distorsions

Le terme de distorsion désigne la manière dont les images stéréoscopiques d’une scène diffèrent d’une vue réelle de celle-ci (Woods et al., 1993). Dans leur article, Woods et al. (1993) reviennent sur un certain nombre de distorsions, telles que la non-linéarité de profondeur, les distorsions en cisailles, la non-correspondance entre les dimensions (x et y contre z) lors de la mise à l’échelle d’une image ou la courbure des plans de profondeurs (spécifique à la configuration convergente). La distorsion en cisaille intervient par exemple lorsque l’observateur se déplace parallèlement à l’écran : les points à différentes profondeurs dévient de différentes quantités, ce qui apparaît comme un cisaillement (Jones et al., 2001). Mais la distorsion la plus notoire est celle dite trapézoïdale¹⁷, spécifique à la configuration convergente, qui intervient en raison du fait que le plan du dispositif d’affichage n’est pas parallèle aux plans de projection des caméras. Cette distorsion amplifie les disparités horizontale et verticale, provoquant des déformations indésirables de la profondeur stéréoscopique ainsi que de l’inconfort visuel. C’est essentiellement l’amplitude excessive de la parallaxe verticale qui est source de problèmes. Celle-ci augmente avec la distance inter-caméras, ou lorsque la dis-

17. En anglais *keystone*.

tance de convergence ou la distance focale diminuent. Ainsi, pour une lentille d'une longueur focale de 3,5 mm, une distance de convergence égale à 1 m et distance inter-caméras de 75 mm, la parallaxe verticale sera de 8,2 mm dans le coin d'un écran de 40,64 cm¹⁸ (Woods et al., 1993). Elle augmente également en s'éloignant de la droite d'intersection des plans de projection vers les coins, comme on peut le voir en figure 2.13, s'additionnant à la parallaxe rétinienne naturelle due à l'éloignement de l'objet d'attention par rapport au plan médian de la tête (voir la section 2.1.3). L'intervalle de fusion vertical étant connu pour être plus petit que l'intervalle horizontal (Allison, 2004), cette parallaxe peut alors provoquer des difficultés de fusion et de l'inconfort visuel (Stelmach et al., 2003). Cependant, Allison prédit que les disparités verticales sont mieux tolérées avec de larges écrans.

À noter enfin que les distorsions horizontale et verticale ont des effets contraires sur la courbure des plans de profondeur, la première causant une déformation convexe de l'espace stéréoscopique – un objet situé à une distance fixe de l'observateur apparaissant plus loin lorsqu'il est placé dans un coin de l'image plutôt qu'au centre de celle-ci (Woods et al., 1993), la seconde une déformation concave – un gradient de disparité plus important dans les coins indiquant un objet plus proche. Ces deux distorsions dues à la configuration convergente agissent donc de manière à s'annuler l'une l'autre (Allison, 2004).

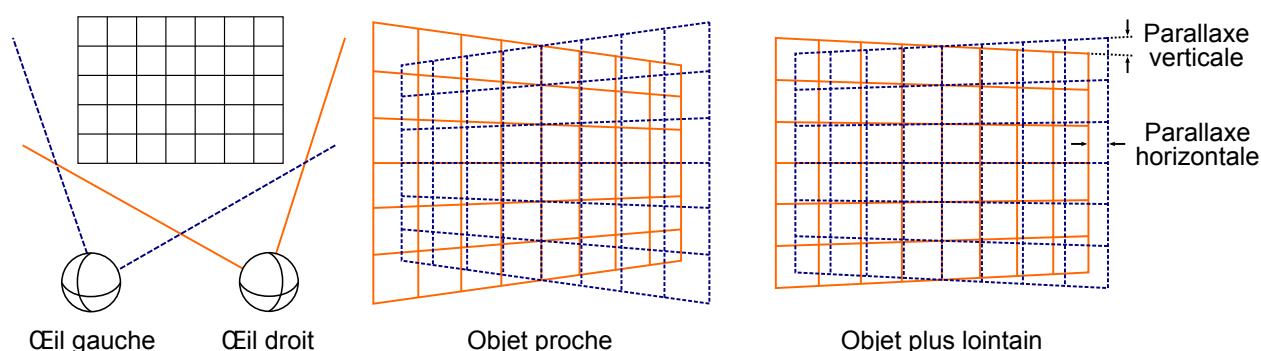


Figure 2.13 Distorsion trapézoïdale.

Avantages et désavantages des configurations

Certaines expérimentations concluent sur les avantages subjectifs et objectifs de la configuration parallèle. IJsselsteijn et al. (2000) ont étudié les effets de trois paramètres des systèmes de caméras stéréoscopiques (distance inter-caméras, distance de convergence et longueur focale des caméras réelles) ainsi que de la durée d'affichage sur le jugement de 12 observateurs

18. 16 po.

quant à la qualité de la profondeur et son naturel. Les résultats ont révélé une préférence des participants pour la configuration où la distance de convergence et la longueur focale étaient les plus importantes ($C=2,60$ m, $F=20$ mm), en comparaison des autres configurations ($C=1,30$ m, $F=10$ mm,) et ($C=1,30$ m, $F=20$ mm), ce qui fournit une base subjective en faveur de la configuration parallèle.

Woods et al. (1993) ont quant à eux utilisé un critère plus objectif : la déformation de l'espace stéréoscopique lors de la variation, dans le cas parallèle ou convergent, des paramètres des systèmes de caméras et d'affichage (distance de convergence, distance inter-caméras, angle d'ouverture horizontale, distance de visualisation, largeur de l'écran et distance interpupillaire). À noter que la distance de convergence pour la configuration stéréoscopique est modifiée en translatant les capteurs des caméras, c'est-à-dire en faisant varier la géométrie de leurs troncs de projection asymétriques. Ils ont développé un programme permettant de visualiser la transformation des coordonnées de l'espace objet à l'espace image sur une grille rectiligne composée de carrés de 10 cm de côté. Ils recommandent suite aux résultats obtenus d'utiliser une configuration avec des caméras parallèles plutôt que convergentes, en raison des distorsions trapézoïdales et de la courbure des plans de profondeur expliquée plus haut en section 2.3.2.

À l'inverse, d'autres arguments penchent en faveur des configurations convergente et images translatées. Allison (2004) souligne pour cette première tout d'abord sa relative simplicité conceptuelle étant donné qu'elle reproduit les mouvements rotatifs des yeux que les humains effectuent naturellement afin d'aligner les axes oculaires sur la cible d'intérêt. Ensuite, elle pallie l'impossibilité en configuration parallèle d'explorer la région proche située entre les caméras, étendant ainsi les limites de l'espace de travail du système. Enfin, les images produites dans le cas parallèle peuvent être inconfortables lorsque les objets de l'environnement sont disposés près des caméras, les disparités horizontales étant alors très importantes (Stelmach et al., 2003). Les configurations images translatées et convergente permettent alors, en modifiant la convergence ou la translation des images, d'ajuster la disparité des objets en les « approchant de l'écran » afin d'améliorer le confort en réduisant le conflit accommodation-convergence. Cependant, dans ce cas, les profondeurs de l'environnement virtuel seront modifiées, et la recherche de confort conduit à ne plus faire correspondre les « profondeurs perçues » des objets aux « profondeurs réelles ».

Stelmach et al. (2003) ont comparé les trois configurations au cours de quatre expérimentations menées sur respectivement 20, 10, 18 et 18 sujets. La première utilisait des images réelles filmées par des caméras espacées de 6,1 cm et de convergence variable (60 cm, 120 cm, 180 cm, 240 cm et infini, soit caméras parallèles). Pour la seconde, les images étaient

générées par ordinateur. Les questionnaires subjectifs remplis par les participants ont révélé que la configuration convergente pouvait améliorer le confort visuel par rapport à la configuration parallèle, selon les auteurs en raison du fait qu'elle diminue la magnitude absolue des disparités dans l'image. Les deux expériences suivantes avaient pour but de comparer les configurations translatée et convergente, en particulier afin d'étudier l'effet des distorsions trapézoïdales sur le confort visuel. Les résultats ont indiqué que ces configurations étaient perceptuellement équivalentes, et que « la présence de disparités verticales n'avait pas d'effet négatif notable sur la notation du confort visuel ou sur la profondeur apparente »¹⁹, tout du moins lorsqu'elles ne sont pas trop importantes. En effet, au cours de la quatrième et dernière expérience, l'effet trapézoïdal fut introduit artificiellement à partir d'images obtenues avec des caméras parallèles, la disparité maximale passant de 2.9 % de la hauteur de l'image (convergence à 60 cm ou translation des images à 94 pixels) à 6.7 %. Ils ont alors observé une baisse des notations de confort. De plus, bien que les résultats soient en faveur des configurations convergentes et translatées pour un degré de convergence modéré, un effet secondaire de ces deux configurations est qu'elles changent la distribution des disparités dans l'image stéréoscopique, et donc la perception globale de la profondeur, en rapprochant le point de disparité nulle des caméras.

2.3.3 Comparaison des valeurs de paramètres

Le choix de configuration des caméras n'est pas le seul facteur de confort ou de réalisme. En particulier, « la représentation correcte des indices de disparité dépend de manière critique de la séparation entre les deux images, de la distance interpupillaire de l'observateur et de sa distance au plan de l'image »¹⁹ (Wann et al., 1995). Cette section détaille le rôle et l'interconnexion de ces deux premiers paramètres.

Distance inter-caméras

L'impact de la DIC sur la performance d'un système stéréoscopique, quelle que soit sa configuration de caméras, a été grandement étudié dans la littérature. Sa valeur influe directement sur l'amplitude des disparités entre les images gauche et droite, donc sur la quantité de profondeur perçue (Jones et al., 2001) mais aussi le confort visuel.

Son incidence est intrinsèquement liée à la valeur de son pendant dans le système d'affichage : la distance interpupillaire. Par exemple, pour une disparité et une distance de visionnage données, les individus dont la DIP est plus petite perçoivent plus de profondeur et atteignent en conséquence plus rapidement les limites de fusion (Lambooij et al., 2007).

19. Notre traduction.

Aussi les valeurs de DIC testées incluent généralement une valeur proche de la DIP anatomique, voir la modélisation exacte de la séparation entre les yeux de l'observateur. Au cours de leur expérimentation citée précédemment (voir la section 2.3.2), IJsselsteijn et al. (2000) ont testé 7 valeurs de DIC : 0, 4, 8, 12, 16, 20 et 24 cm. Les notations de qualité et de naturel de la profondeur ont révélé que les 12 participants préféraient un affichage stéréoscopique (augmentation des notes entre 0 et 4 cm) à condition que les disparités restent dans des limites naturelles (chute des notes après 8 cm). Ce résultat était particulièrement notable pour les valeurs des deux autres paramètres de l'expérimentation générant les disparités les plus importantes ($C = 1.30$ m, $F = 20$ mm). Best (1996) ont quant à eux comparé la DIP anatomique de chacun de leurs 12 participants à 3 autres valeurs : 5,0 cm, 6,3 cm et 7,4 cm, les extrêmes correspondant aux réglages minimum et maximum de DIP du visiocasque utilisé, la valeur centrale à la moyenne adulte de DIP. Après une immersion de 15 minutes durant laquelle ils durent estimer la taille de deux objets 2D dans un environnement virtuel, les participants étaient invités à remplir un sondage subjectif sur l'inconfort visuel ressenti. Bien que les jugements sur la taille des objets ne fussent pas affectés par les différences de réglage de DIP du visiocasque, les sujets ont reporté un inconfort significativement plus important dans les conditions extrêmes, préférant une fois encore des valeurs de DIC proches de la DIP naturelle.

Cependant, Wann et al. (1995) ont remarqué qu'en permettant à leurs participants de régler la DIC durant le visionnage d'une scène de réalité virtuelle, ceux-ci faisaient en sorte que les images fusionnent là où leurs yeux convergeaient (c'est-à-dire que la profondeur qu'ils choisissaient pour les objets ne dépendait que de celle qu'ils estimaient et non de celle que les auteurs avaient donnée en créant la scène virtuelle). De plus, la distance inter-caméras choisie était systématiquement plus petite que la distance interpupillaire anatomique. Les auteurs soutiennent donc qu'il faut choisir une distance inter-caméras en fonction de l'intervalle de profondeur que l'observateur aimerait échantillonner, car il n'existe pas de distance qui réduira l'effort de fusion sur l'intervalle total. Selon eux, « à une distance de travail proche, la DIC qui réduit l'effort de fusion est significativement plus petite que la DIP de l'individu »²⁰. Jones et al. (2001) préconise également de choisir une distance inter-caméras *ad hoc*, après plusieurs essais, et ne convenant alors qu'à un individu en particulier. Wann et al. (1995) soulignent enfin qu'il faudrait prendre en compte la vergence, qui introduit de petits changements dans la DIP étant donné que les centres de rotation oculaire sont situés 5-6 mm derrière le point nodal du système optique (Bennett and Francis, 1962). Celle-ci est par exemple réduite d'1 mm entre une distance de convergence infinie et de 30 cm (calcul qui tient uniquement compte de la rotation des yeux). Même s'il ne s'agit que d'un petit

20. Notre traduction.

pourcentage, il introduira un biais spatial et l’observateur percevra différemment les profondeurs. Utiliser une DIC fixe est donc équivalent à supposer que la vergence de l’individu reste constante (Wann et al., 1995), une solution étant alors de modifier en temps réel la valeur de ce paramètre.

Modification en temps réel des paramètres

L’ajustement en temps réel de certains paramètres stéréoscopiques, en fonction de la position de l’observateur ou de la tâche qu’il a à accomplir par exemple, peut accroître la performance d’une configuration en jouant en particulier sur la perception des profondeurs.

Les résultats de l’expérimentation de Milgram and Krüger (1992) ont révélé que la modification de la distance inter-caméras avait un impact significatif sur le jugement des distances et des profondeurs, dans le cas d’une configuration convergente (distance de convergence à 1,782 m). En effet, au cours de deux expérimentations menées respectivement auprès de 12 et 10 individus, l’augmentation de DIC conduisait à une surestimation des distances, sa diminution à une sous-estimation, les trois écarts testés étant 80 mm, 160 mm et 240 mm. Des retours à court terme, sous la forme d’indications ou d’indices de profondeur ajoutés à la scène, permettaient cependant de corriger les erreurs de jugement. Les auteurs soutiennent que l’ajustement en temps réel de ce paramètre dans le cas d’une tâche de télémanipulation pourrait maintenir une configuration théoriquement « optimale » en fonction des conditions d’observation, permettant par exemple à l’opérateur, en diminuant la valeur de DIC, d’augmenter son intervalle de profondeurs fusionnables au prix d’une faible résolution de niveau de profondeurs. Cela le forcera cependant à s’adapter continuellement aux changements dans la relation entre les distances inter objets perçues et réelles, en particulier en raison du fait que les indices de taille restent inchangés lorsque ceux des profondeurs changent.

Jones et al. (2001) décrivent quant à eux une nouvelle méthode de calcul en temps réel de la position des caméras virtuelles, pour une configuration parallèle, minimisant la distorsion en cisailles et la variation de distribution des profondeurs. Concernant la première, qui se produit lorsque l’observateur se déplace parallèlement à l’écran, les auteurs se sont désolidarisés de l’idée courante voulant que les caméras suivent au maximum le mouvement des yeux, en utilisant plutôt une relation de proportionnalité. Ainsi, lorsque les yeux se translatent d’une fois la DIP, les caméras sont déplacées d’une fois la valeur de la DIC, qui est en l’occurrence plus petite. La variation de distribution des profondeurs intervient en revanche lorsque l’observateur se déplace perpendiculairement à l’écran. Afin de la minimiser, ils utilisent une méthode DRC (Celikcan et al., 2013) pour maintenir fixes les limites minimum et maximum de profondeur. Il s’agit de récupérer les points le plus proche et le plus lointain après ajustement des caméras en fonction du suivi de la position de l’observateur, et de calculer la DIC de

manière à ce que la distance entre ces deux points reste fixe et que la distance à l’observateur n’ait varié que de son déplacement.

Enfin, Celikcan et al. (2013) introduisent une nouvelle approche de contrôle concerté de la DIC et de la distance focale, en fonction de l’intervalle de profondeur confortable de l’observateur et de son attention, dans le but d’améliorer le confort et la perception de la profondeur. Leur méthode, DADC, comprend une phase de calibration afin de déterminer les extremums de disparité propre à l’individu, c’est-à-dire l’intervalle dans lequel cette disparité n’est ni trop importante – afin de limiter les contraintes imposées aux yeux et la diplopie – ni trop basse – ce qui conduirait à une impression faible de profondeur. En ce qui concerne l’attention, le point de regard de l’observateur n’a pas été déterminé à l’aide d’un système de suivi, mais considéré systématiquement comme dirigé vers le centre de l’écran. Ainsi, pour chaque image de la séquence stéréoscopique, l’analyse de l’intervalle de profondeur qu’elle affiche et de son contenu sémantique conduit à la construction d’un problème d’optimisation dont la résolution fournit les paramètres de distances inter-caméras et focale. Les termes d’énergie utilisés pénalisent la distance au plan de convergence des éléments qui ont un score d’importance élevé et une petite distance au centre de l’attention de l’utilisateur, et favorisent l’intervalle total de disparité de la scène et donc la profondeur totale perçue. Les paramètres obtenus préviendront donc la violation du conflit accommodation convergence en approchant le plan de convergence de la région qui attire l’attention de l’observateur, tout en produisant un rendu riche au niveau des profondeurs. Une contrainte de cohérence temporelle permet que les paramètres changent de manière régulière entre les différentes images. Leur méthode a été comparée en termes de qualité, profondeur et confort à la méthode DRC citée ci-dessus ainsi qu’à une méthode naïve, dans une étude réalisée auprès de 15 participants. Elle obtient de meilleures moyennes dans tous les critères, en particulier celui de la qualité, ce qui laisse supposer qu’adapter les paramètres de distance inter-caméras et de distance focale en fonction de l’attention du participant peut améliorer la performance de la configuration stéréoscopique.

2.3.4 Critères d’évaluation de la performance

Cette section résume les différents critères de comparaison des configurations stéréoscopiques utilisés dans les recherches citées tout au long de la revue de littérature.

Notations subjectives

Selon Allison (2004), il est important de noter que les effets de distorsions discutés précédemment sont prédits géométriquement, « mais ne correspondent pas toujours à la perception

humaine »²¹. Aussi est-il essentiel de s'intéresser aux préférences subjectives des individus pour le choix d'une configuration et de ses paramètres. En particulier, de nombreux chercheurs ont demandé à leurs participants de noter les différentes conditions expérimentales sur le critère subjectif de confort visuel. À celui-ci s'ajoutent les critères de qualité, réalisme, plaisir ou amusement, perception des distances et des profondeurs, et enfin sentiment d'immersion (Hillaire et al., 2008b, Celikcan et al., 2013). Pour recueillir les notations, Celikcan et al. (2013) ont par exemple utilisé une échelle de Likert à 5-points allant de mauvais à excellent, tout comme Blum et al. (2010) qui laissa cependant la liberté à ses participants de placer leur marque à n'importe quel endroit puis a converti les résultats mesurés vers une échelle de 0 à 100.

Critères objectifs

La capacité de fusion revient également à plusieurs reprises comme critère de comparaison, en particulier en raison du fait que les difficultés de fusion peuvent provoquer de la fatigue et de l'inconfort visuels. Les valeurs limites de fusion, en particulier dans l'espace proche de l'observateur, peuvent alors représenter un critère objectif donnant des indications sur l'intervalle de profondeur total de travail possible avec les différentes configurations.

Étant donné que les stimuli stéréoscopiques induisent des réponses oculaires similaires à celles provoquées par des stimuli réels (Duchowski et al., 2011, Daugherty et al., 2010, Wann et al., 1995), la similarité de ces réponses, en particulier les angles de vergence, avec le cas réel peut servir de mesure objective à la performance d'une configuration. Il semble en effet raisonnable de considérer que plus le rendu des profondeurs en utilisant un assortiment de valeurs de paramètres sera réaliste, plus les réponses oculaires engendrées se rapprocheront des réponses réelles.

2.4 Flou de profondeur de champ

Dans la vision naturelle d'un monde en relief, la profondeur de champ correspond à l'intervalle de distances proches de celle du point de regard où les objets sont perçus nettement (Hillaire et al., 2008b). Les éléments situés en avant et en arrière de cet intervalle apparaissant de plus en plus flous à mesure qu'ils s'en éloignent, en raison des propriétés physiologiques de l'œil vues en section 2.1.3. Ce flou est un des indices utilisés par le cerveau pour percevoir la profondeur.

En revanche, généralement dans le cas de systèmes immersifs, les images stéréoscopiques sont générées sans flou. Selon Wann et al. (1995), il est alors impossible de rendre de manière

21. Notre traduction.

véridique un large intervalle de profondeurs. En effet, l’accommodation s’effectuant à la profondeur de l’écran, tous les objets de la scène seront perçus nets quelle que soit leur position par rapport au point de regard, ce qui perturbe la perception de la profondeur. De plus, sans ce flou, les images virtuelles risquent de paraître « trop parfaites », synthétiques (Hillaire et al., 2008b), et donc de restreindre le sentiment d’immersion.

Ce choix peut être dû à la volonté de permettre à l’observateur de regarder où il le souhaite, ce qui implique d’afficher la scène entièrement nette si l’objet de son attention n’est pas connu. En revanche, la détermination du point de regard 3D peut servir à générer artificiellement le flou de profondeur de champ adéquat, tel que le présente la figure 2.14.



Figure 2.14 Flou de profondeur de champ dépendant du point de regard. Exemples de rendus sans flou (au milieu) et avec flou (à droite).

2.4.1 Techniques de génération de flou

Riguer et al. (2003) proposent deux techniques de génération de flou, en deux et cinq passes, qui utilisent la programmabilité du pipeline graphique à l’aide de nuanceurs. Pour la première, un rendu hors écran de la scène est tout d’abord réalisé. La différence entre la profondeur du point de regard et la profondeur de chaque pixel est utilisée pour calculer un coefficient de flou. Durant la seconde passe, les pixels sont floutés en combinant leur couleur à celles de pixels voisins, échantillonnés à l’aide d’un filtre appelé cercle de confusion (par exemple reproduisant une distribution en disques de Poisson). La taille de ce noyau, et donc la distance d’échantillonnage, augmente en fonction du coefficient de flou. Un test de profondeur durant la phase de mélange prévient les artefacts de bavure des couleurs²².

La deuxième technique fonctionne de la même manière en ce qui concerne la première

22. En anglais *color leaking*.

passé, mais utilise ensuite pour l'échantillonnage des pixels voisins un noyau gaussien au lieu de simuler le cercle de confusion d'une lentille de caméra physique.

2.4.2 Cas monoscopique

Hillaire et al. (2008b) se sont penchés sur l'effet de l'addition d'un flou dynamique de profondeur de champ à un jeu de tir à la première personne sur les performances du joueur et son expérience de jeu. Ils ont décrit une nouvelle méthode de génération de flou basée sur la première technique de Riguer et al. (2003) citée ci-dessus, mais dont le noyau d'échantillonnage est pivoté autour de son centre d'un angle aléatoire pour chaque pixel, afin d'éviter l'apparition d'artefacts fantômes et de les remplacer par un bruit haute-fréquence. Ils adjoignent un flou de vision périphérique, simulé indépendamment en diminuant progressivement la netteté de l'image depuis le centre, bien que celui-ci intervienne continuellement dans la vision humaine étant donné qu'il est lié aux différences de densité des récepteurs sur la rétine. Les auteurs s'attendaient à ce que ce dernier améliore l'expérience sensorielle des utilisateurs, mais les incite également à regarder au centre de l'écran. En effet, ils n'ont pas eu recours à un système de suivi oculaire, mais considéré que le joueur maintenait son regard dans une zone rectangulaire centrée dans l'écran que les sujets d'une première expérience observaient 93 % du temps. La profondeur de champ est calculée dans cette zone d'autofocus comme la moyenne pondérée des profondeurs des pixels. Les résultats de leur expérimentation ont révélé un effet négatif significatif du flou sur les performances, avec une diminution de la précision, du nombre d'adversaires tués et une augmentation du nombre de morts du joueur. Pour ce qui est de l'appréciation subjective, les opinions étaient très partagées. Sur 30 participants, 11 ont trouvé que le flou augmentait le réalisme (contre 13 non et 6 sans préférence), 9 ont trouvé qu'il augmentait le plaisir de jeu (contre 10 non et 11 sans préférence), 10 ont trouvé que la perception des distances et profondeurs était meilleure (contre 14 non et 6 sans préférence), et 11 ont trouvé que le sentiment de présence était accru (contre 10 non et 9 sans préférence). Ceux qui préféraient jouer avec le flou ont souligné que le jeu était « beaucoup plus réaliste »²³ avec une « immersion plus importante »²³. Parmi les autres, certains ont signalé que celui-ci était ennuyeux, voire inconfortable et fatigant (à l'origine de maux de tête), et parfois « trop fort »²³. Selon les auteurs, le flou de profondeur de champ se révèle dérangent essentiellement lorsque les joueurs souhaitent regarder en dehors de la zone d'autofocus. La détermination du point de regard permettrait alors de contrer ce biais.

Une seconde expérimentation de Hillaire et al. (2008a) corrobore cette conjecture. Cette fois, le point de regard est calculé à l'aide d'un système de suivi oculaire, les trois conditions de navigation comparées étant :

23. Notre traduction.

- Condition de contrôle (scène sans profondeur de champ).
- Profondeur de champ (calculée en utilisant une zone d'autofocus située au centre de l'écran).
- Profondeur de champ suivant le regard (calculée en utilisant le dispositif de suivi).

L'étude des préférences subjectives réalisée auprès de 8 hommes révèle une forte influence de l'utilisation du suivi de regard quant à l'ajout de flou de profondeur de champ. Comme dans l'article précédent (Hillaire et al., 2008b), aucun des deux premiers modes de navigation ne semble se distinguer l'un par rapport à l'autre, tandis que le troisième est significativement préféré en terme de plaisir et de sentiment d'immersion. Les participants ont indiqué qu'ils percevaient mieux la profondeur de la scène virtuelle avec cette condition, l'un d'eux ayant ajouté qu'il avait l'impression de « regarder un écran autostéréoscopique »²⁴. Cela confirme donc que le flou « peut être utilisé pour transmettre des informations additionnelles concernant les relations spatiales entre les objets, comme le font les ombres »²⁴.

2.4.3 Cas stéréoscopique

Les résultats précédents montrent que l'ajout d'un flou de profondeur de champ peut accroître le plaisir et le sentiment d'immersion des individus, tout du moins lorsqu'il est calculé en adéquation avec la localisation de leur regard. Dans le cas stéréoscopique, le flou peut également permettre de réduire l'inconfort visuel en agissant sur les limites de fusion.

En effet, la diplopie représente avec le conflit accommodation-convergence l'une des deux sources connues d'inconfort en immersion, tel que vu en section 2.1.4. Elle se manifeste pour un individu en dehors de ses limites de fusion, c'est-à-dire de son intervalle de disparités acceptables. Or, le contenu des images stéréoscopiques en fréquence spatiale (c'est-à-dire le « nombre de cycles d'une lumière périodique ou d'une oscillation de couleurs dans un degré d'angle visuel »²⁴.) influe sur les limites de fusion (Leroy et al., 2012), ces dernières augmentant lorsque la fréquence spatiale diminue (Lambooij et al., 2007). Flouter artificiellement les images avec un flou de profondeur de champ, en agissant sur leur contenu fréquentiel, peut donc augmenter l'intervalle de fusion et par là même le confort visuel (Lambooij et al., 2007, Blum et al., 2010).

Nagata (1996) a mené une expérience auprès de 3 sujets, en leur présentant un objet en premier plan devant un fond constitué de boîtes noires et blanches. Une version de ce fond était floutée, l'autre non. Les résultats ont révélé que l'intervalle de profondeurs pour l'objet de premier plan dans lequel les deux images stéréoscopiques pouvaient être fusionnées augmentait dans la version floutée. Leroy et al. (2012) ont quant à eux mis au point un algorithme qui supprime les hautes fréquences spatiales, en moyennant les textures (ce qui revient à les

24. Notre traduction.

flouter) des objets présents dans les zones de l'image à forte disparité horizontale (c'est-à-dire des objets distants de l'écran), afin de maintenir un confort visuel acceptable et de réduire la pression oculaire. Leur protocole inclut des mesures objectives (perte de l'aisance d'accommodation, punctum maximum et acuité stéréoscopique), ainsi que subjectives concernant la pression oculaire ressentie. Les 20 sujets furent soumis à 2 immersions de 30 minutes, avec ou sans traitement d'image. Les résultats ont montré que l'environnement modifié réduisait significativement la fatigue visuelle (diminution de la dégradation des trois paramètres objectifs), ce qui tend à prouver que leur méthode permet objectivement de réduire cette dernière. De plus, les sujets ont dans l'ensemble expérimenté moins de pression oculaire tandis qu'ils effectuaient la tâche définie dans le protocole expérimental, bien que la précision de celle-ci n'ait pas été améliorée.

Blum et al. (2010) ont également testé l'ajout d'une profondeur de champ artificielle dans le but de réduire l'inconfort dû à la diplopie. Leur expérimentation utilisait, conjointement à un visiocasque et un moniteur stéréoscopique, un dispositif de suivi oculaire qui estimait le point de regard 3D à partir de la profondeur des pixels (voir la section 2.2.2). L'implémentation du flou était similaire à la première méthode décrite en section 2.4.1. Les notes attribuées par les 18 participants révèlent que la présence de flou accroît significativement la qualité, qui inclut ici le critère de confort. Cependant, les auteurs ont réalisé que certains sujets n'expérimentaient pas de diplopie, même en l'absence de flou, et n'ont donc pas pu tirer de conclusion sur son effet réel sur la zone de fusion. Par contre, ceux qui n'arrivaient effectivement pas à fusionner les images avaient une préférence pour la version floutée. Enfin, pour trois participants, le flou n'était préférable que lorsqu'il affectait le fond tandis qu'ils observaient le premier plan. Dans le cas inverse, c'est la scène entièrement nette qui remportait leurs suffrages. Ce résultat laisse supposer qu'un système qui floute le fond lorsque l'utilisateur regarde des objets proches mais n'affecte pas le premier plan lorsqu'il projette son regard au loin pourrait s'avérer plus performant.

2.5 Applications existantes en psychologie et psychiatrie

Cette maîtrise est réalisée en partenariat avec l'Institut Philippe-Pinel de Montréal, et s'intègre dans son projet d'utilisation de la réalité virtuelle à des fins cliniques. La voûte de réalité virtuelle dont l'Institut a fait l'acquisition, et dont l'installation a été finalisée en septembre 2013, est vouée à être employée pour l'évaluation et le diagnostique, mais aussi le traitement des patients. Cette section revient sur diverses expérimentations ayant eu recours au suivi de regard, au flou ou à la réalité virtuelle, menées dans les domaines de la psychologie et de la psychiatrie.

2.5.1 Suivi de regard

Mesures oculaires significatives

Une première étape à l'analyse de l'attention de l'observateur consiste à mettre en relation la position 3D du point de regard avec la scène, afin de relier les mouvements oculaires aux différents éléments et régions d'intérêt de l'environnement. Des données statistiques peuvent alors être extraites du signal de suivi de regard, préalablement fractionné en fixations et saccades grâce aux méthodes vues en section 2.2.4. Henderson and Hollingworth (1998) se sont intéressés à la façon dont les caractéristiques sémantiques des régions de la scène influencent les mouvements oculaires, et soulignent à cette occasion l'importance de déterminer où les fixations tendent à se concentrer et combien de temps elles restent localisées à un emplacement particulier.

Pour une région spécifique, plusieurs métriques peuvent être utilisées en fonction du niveau d'analyse. Dans un premier temps, il est intéressant d'examiner le nombre et la durée totale des fixations sur cette région durant l'immersion, ces deux valeurs étant corrélées (Duchowski, 2007). À un niveau d'analyse plus fin, les mesures communes incluent les durées des premier et second passages du regard, un passage correspond au moment entre l'entrée et la sortie du point de regard de cette région. Selon Henderson (1992) ces mesures globales des temps de fixation traduisent plutôt certains processus cognitifs ayant lieu après l'identification. Il leur préfère la durée de première fixation, soit le temps entre le moment où l'observateur pose ses yeux pour la première fois sur un objet jusqu'à ce qu'il les dirige vers un autre endroit, y compris un autre endroit de l'objet.

En complément de ces valeurs, une représentation graphique de la séquence de fixations et saccades peut s'avérer pertinente afin d'en faciliter l'interprétation. Les cartes de densité et chemins de balayage²⁵ (Pfeiffer, 2012), illustrés en figure 2.15, permettent par exemple de visualiser dans l'environnement virtuel respectivement la position des fixations et leur succession temporelle.

Perception de scènes

La mesure du regard est à la base d'un grand nombre d'études liées aux neurosciences et à la psychologie (Duchowski, 2007), en particulier afin de comprendre les processus cognitifs qui sous-tendent l'attention visuelle. Ces projets explorent la manière dont les humains perçoivent une scène, une œuvre d'art, ou encore utilisent les informations visuelles pour effectuer des tâches du quotidien. Comme expliqué en section 2.2.5, les données fournies par les dispositifs de suivi offrent des mesures non intrusives et en temps réel du traitement de

25. En anglais *heat maps* et *scan paths*.

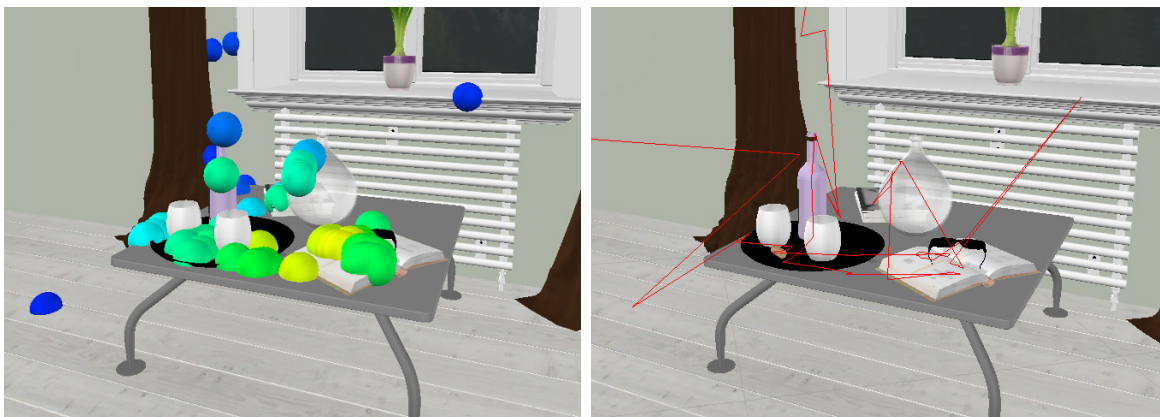


Figure 2.15 Exemples de représentations graphiques des fixations : carte de densité (à gauche) et chemin de balayage (à droite).

ces informations visuelles. L'étude des mouvements oculaires permet donc de répondre à des questions critiques concernant l'acquisition et le traitement d'informations, la perception et la cognition (Henderson and Hollingworth, 1998, Henderson and Hollingworth, 1998).

Contrairement à la lecture où le schéma est souvent prédéfini (de gauche à droite et de haut en bas pour un texte en français par exemple), l'exploration d'une scène varie de manière importante entre les individus (Duchowski, 2007). Cependant, il est possible d'affirmer que la plus grande partie de l'information globale est généralement extraite lors des quelques premières fixations (Rayner and Pollatsek, 1992), puis suivie d'une inspection plus fine des détails. Par ailleurs, en raison de la capacité limitée des récepteurs visuels vue en section 2.1.5, le cerveau doit combiner un ensemble d'instantanés fovéaux haute résolution en une scène cohérente : il s'agit de l'intégration de scène.

Dans le cas d'une image, ce processus partiellement déterministe est relié au contenu sémantique. Dans son étude sur la perception de l'Art, Buswell (1935) a réalisé la première exploration systématique des positions de fixation, en présentant 55 photos d'œuvres d'art à 200 participants. Il a remarqué par exemple que les participants avaient tendance à se concentrer sur les personnages plutôt que sur le décor en examinant « Un dimanche après-midi à l'île de la Grande Jatte » de Georges Seurat. Ces données fournissent la première confirmation que le schéma d'observation d'une scène complexe est lié à son information sémantique, et par extension à notre traitement perceptuel et cognitif de celle-ci. À noter cependant que le comportement oculaire varie en fonction de la tâche donnée à l'individu (Yarbus, 1967).

Diagnostic de pathologies

Les systèmes de suivi de regard peuvent être envisagés à des fins diagnostiques dans le domaine de la psychiatrie ou pour certaines maladies neurodégénératives.

Diefendorf and Dodge (1908) ont par exemple étudié les schémas de poursuite douce de personnes souffrant de schizophrénie, de psychoses maniacodépressives ainsi que d'autres pathologies (épilepsie, neurosyphilis, etc.). Il en est ressorti une forte corrélation entre des schémas de poursuite douce non réguliers et les patients atteints de schizophrénie.

Il a également été établi que les mouvements oculaires des individus souffrant de maladies neurodégénératives, telles que la maladie de Parkinson, étaient anormaux. Dans leur recherche, Iijima et al. (2003) ont utilisé un visiocasque couplé à du suivi oculaire afin de diagnostiquer cette pathologie. Les tests consistaient à suivre une cible virtuelle du regard. Les résultats ont confirmé que les individus sains généraient des mouvements doux et continus durant presque toute la durée de la tâche, tandis que les mouvements des patients atteints de la maladie de Parkinson étaient irréguliers, comportant des pauses dues à l'akinésie provoquée par la maladie et des saccades utilisées pour compenser le retard résultant. La forme du graphe de mouvements oculaires, les erreurs de position des yeux ainsi que le nombre de saccades peuvent dépendre de la maladie neurodégénérative, et donc être utilisés comme indicateurs afin de la repérer, mais aussi de l'identifier (Iijima et al., 2003).

La connaissance en temps réel de l'élément de la scène sur lequel le patient fixe son attention est une donnée qui intéresse particulièrement les cliniciens de l'Institut Philippe-Pinel de Montréal, pour diagnostiquer mais aussi traiter certaines déviations comportementales. Dans ce contexte, ils pourraient alors désirer orienter le point de regard du patient vers un autre élément, par exemple en ayant recours à du flou.

2.5.2 Impact du flou sur l'attention visuelle

Comme expliqué en section 2.4, les images stéréoscopiques générées en temps réel pour les systèmes immersifs sont généralement dépourvues de flou. Cependant, dans la création de contenu stéréoscopique pour le cinéma par exemple, le flou est utilisé sur certains plans de la scène afin de guider l'attention de l'observateur vers les éléments restés nets. Il ne s'agit pas ici de générer un flou de profondeur de champ cohérent avec la profondeur du point de regard l'observateur, mais de l'utiliser comme un outil afin de concentrer l'attention de celui-ci sur un objet particulier, ou au contraire l'en distraire, comme Hillaire et al. (2008b) le proposent dans la conception de jeux vidéos. Cette influence volontaire doit cependant être dissociée de celle, involontaire, inhérente aux conditions expérimentales – en particulier le temps de latence dans la mise à jour du flou – qui peut aussi biaiser le comportement de l'individu.

De plus, lorsque c'est le premier plan qui est flouté, ce procédé peut conduire à un conflit lié au fait que l'attention d'un individu est attirée naturellement à la fois par les régions nettes et par les objets proches de lui (Huynh-Thu et al., 2013).

Huynh-Thu et al. (2013) ont étudié l'impact de l'insertion d'un premier plan flou ou net sur le déploiement de l'attention visuelle de 24 participants vis-à-vis d'un contenu monoscopique ou stéréoscopique. Ils se sont particulièrement intéressés à la localisation spatiale des fixations obtenues grâce à un système de suivi oculaire. Les résultats dans le cas monoscopique indiquent que lorsque le premier plan est flouté l'attention des observateurs se dirige en dehors de cette zone. En revanche dans le cas stéréoscopique, les observateurs dirigeaient leur attention vers les objets de premier plan même lorsqu'ils étaient flous, néanmoins avec une densité plus faible et pour des durées plus courtes que dans le cas non flouté. Cependant, les auteurs soulignent des résultats dépendant à la fois de l'observateur et du contenu (le test ayant été réalisé avec 2 scènes différentes).

2.5.3 Réalité virtuelle

« Appliquée convenablement, la réalité virtuelle détient un fort potentiel quant à l'approfondissement de la compréhension et des traitements de nombreuses psychoses » (Freeman, 2008). Ses avantages comprennent la possibilité de contrôler totalement les stimuli et les variables de l'environnement, afin de faire correspondre ce dernier aux besoins et aux attentes des patients et des thérapeutes pour un traitement individualisé (Strickland, 1997). Elle permet également de mettre en scène des situations qui ne seraient pas reproductibles en réalité, car dangereuses ou inacceptables, par exemple dans le cas de l'étude des déviances comportementales sexuelles à tendances pédophiles.

Par ailleurs, le potentiel des applications cliniques de la réalité virtuelle peut être accru en la combinant à des outils de mesures de comportement dynamiques (déplacements, mouvements oculaires) afin d'observer l'évolution et les modifications de comportement au cours du traitement, et d'en tirer des conclusions sur son efficacité (Renaud et al., 2002a). Un environnement stéréoscopique se révèle également un plus, un nombre important d'études ayant montré que les images stéréoscopiques avaient un impact psychologique plus fort sur les sujets car elles intensifiaient leur sentiment de présence (IJsselsteijn et al., 2000).

Estimation des symptômes

La détermination des symptômes est généralement réalisée par le biais d'un entretien avec l'individu, en revenant sur les souvenirs de la semaine ou du mois précédent. *A contrario*, par la confrontation à des situations ou des stimuli précis (et contrôlés), la réalité virtuelle rend

possible l'observation des symptômes en situation, ainsi qu'une estimation de leur fréquence et intensité réelles (Freeman, 2008).

L'exemple du cas de la paranoïa permet de prendre l'entière mesure de l'apport de la réalité virtuelle. En effet, un questionnaire ou une entrevue avec le clinicien ne permettent pas de distinguer les pensées paranoïaques « fondées », c'est-à-dire s'appuyant sur un stimulus effectif, de celles relevant de la psychose. En revanche, lorsqu'une situation sociale neutre est présentée en utilisant la réalité virtuelle, toutes les pensées paranoïaques survenant peuvent alors être considérées comme non fondées. De plus, contrairement à une situation réelle, l'individu ne peut pas agir de manière à produire des réactions hostiles de la part des avatars (Freeman, 2008).

Dans l'expérimentation menée par Freeman et al. (2008), l'environnement virtuel utilisé était un trajet de 5 minutes entre deux stations du métro londonien. Les commentaires oraux recueillis auprès des 200 participants non cliniques ont été classés comme positifs, neutres ou paranoïaques (par exemple « Un individu était agressif - son intention était de m'intimider et de me mettre mal à l'aise »²⁶). Les résultats indiquent qu'environ 40% des participants souffraient de pensées paranoïaques, attribuant de manière significative des personnalités ou des états mentaux aux personnages générés par ordinateur. Ceux souffrant d'un trait paranoïaque supérieur ont expérimenté un plus haut degré de pensées de persécution, ce qui valide la méthodologie. Selon les auteurs, cette étude est la démonstration la moins ambiguë des pensées paranoïaques expérimentées par la population générale. Au cours d'un récent travail dans ce laboratoire, des individus qui souffraient d'hallucinations auditives ont même rapporté avoir entendu des voix dans le métro virtuel (Freeman, 2008).

Renaud et al. (2002a) présentent un autre exemple d'estimation de symptômes par le biais de la réalité virtuelle, cette fois-ci dans le cas d'une phobie. À l'aide d'un visiocasque et d'un système de repérage des déplacements, ils ont suivi et comparé le comportement de 24 individus souffrant ou non d'arachnophobie en présence d'araignées virtuelles ou de cibles neutres. Les participants avaient pour tâche de suivre visuellement l'objet qui leur était présenté. Les auteurs ont observé des signes de comportement d'évitement en réponse à l'araignée virtuelle pour certains sujets arachnophobes. Ils ont par ailleurs noté que l'état d'anxiété des participants était corrélé positivement avec leur sentiment de présence.

Renaud et al. (2002b) ont également testé la réalité virtuelle pour l'évaluation de la préférence sexuelle des individus, en mesurant le point de regard du patient, son état affectif et son sentiment de présence après une interaction avec un modèle virtuel nu.

26. Notre traduction

Développement de traitements

Bien qu'originellement la réalité virtuelle n'ait été utilisée que pour traiter de simples phobies, en particulier en exposant les individus à des stimuli spécifiques à leurs peurs, aujourd'hui son utilisation s'étend aux stress post-traumatiques, aux troubles de l'alimentation, aux dysfonctions sexuelles, à certaines dépendances et même aux hallucinations schizophrènes (Glantz et al., 2003). Freeman (2008) cite trois types d'application possibles de la réalité virtuelle à la conception de traitement. Tout d'abord, elle pourrait être utilisée de manière éducative, afin d'éclairer les sujets sur les facteurs agissant de manière positive ou négative sur leurs symptômes. Il est par exemple possible de les aider à ressentir l'impact de leur état émotionnel sur leurs hallucinations, en effectuant des expériences en réalité virtuelle après avoir conditionné les individus en améliorant ou dégradant leur humeur.

La seconde utilisation découle de celle développée pour les troubles de l'anxiété : l'exposition aux facteurs de peurs. Les stimuli restant entièrement contrôlables, les expériences en réalité virtuelle apparaissent plus sûres et plus faciles pour les patients que leurs homologues en situation réelle, et pourraient les précéder. Parsons et al. (2008) ont réalisé une méta-analyse sur la VRET et son efficacité concernant la réduction de l'anxiété et de certaines phobies. D'après les données disponibles, il apparaît qu'elle est « relativement efficace d'un point de vue psychothérapeutique auprès de patients sélectionnés attentivement »²⁷, et révèle un bon potentiel en tant qu'approche de traitement pour différentes phobies spécifiques (vertige, peur de conduire, claustrophobie, peur de l'avion et arachnophobie). Dans d'autres cas, comme celui de la paranoïa, une hiérarchie de tests concernant les croyances de menaces pourrait être effectuée à l'aide de situations virtuelles, en accentuant progressivement certains paramètres de l'environnement.

La troisième utilisation serait celle ayant pour but la réhabilitation des patients en leur enseignant à surmonter les symptômes lorsqu'ils surviennent. Une grande variété de stratégies peuvent être testées, afin de trouver avec le patient une réaction personnalisée, adaptée et reproductible dans le monde réel (Freeman, 2008, Strickland, 1997). Des exercices tels que des jeux de rôles avaient déjà été mis au point afin d'aider les patients schizophrènes à développer leur habileté sociale déficiente, mais restaient limités en raison de la difficulté du formateur à projeter des émotions. En fournissant un monde 3D réaliste et des avatars pouvant fournir des stimuli émotionnels, la réalité virtuelle permet de dépasser cette limite. Ku et al. (2007) rapportent l'efficacité de cette approche dans le cadre d'un essai clinique préliminaire réalisé auprès de 10 patients schizophrènes afin de développer leur capacité de conversation dans des situations spécifiques. Strickland (1997) a quant à lui tenté, à l'aide d'un visiocasque, d'entraîner deux enfants souffrant d'autisme à réaliser certaines tâches dans un monde virtuel.

27. Notre traduction.

Il s'agissait en particulier de reconnaître et suivre visuellement des voitures dans une scène de rue, puis à repérer un panneau-stop, marcher vers lui et s'arrêter à sa hauteur. À la fin des tests, les deux enfants parvenaient à identifier les voitures et les couleurs, à les étiqueter verbalement et à les suivre en tournant leur corps. Ils arrivaient également à trouver dans l'environnement le panneau-stop, mais un seul d'entre eux s'est arrêté une fois l'avoir atteint.

2.6 Objectifs et hypothèses

2.6.1 Conclusion sur l'état de l'art

Au vu de l'état des recherches actuelles, aucun choix de configuration des caméras ni de valeur des paramètres ne se dégage clairement comme le plus efficace, ou offrant les expériences virtuelles les plus confortables, réalistes et immersives. Il apparaît que les différences physiologiques entre les individus et leurs préférences subjectives influencent énormément leurs ressentis et leur capacité de réaliser une tâche en environnement virtuel. La question d'une configuration universelle qui emporterait l'unanimité auprès des observateurs reste donc en suspens, tout comme certaines approches d'ajustement des paramètres et de comparaison des configurations demeurent inexplorées. En particulier, nous n'avons pas eu connaissance de recherches testant l'ajustement continu par interpolation de l'écart inter-caméras en temps réel en fonction de la distance du point de regard de l'observateur, ni d'expériences menées en comparant les réponses oculaires d'individus à l'observation d'une scène réelle et de la même scène reproduite en virtuelle.

2.6.2 Rappel de l'objectif

En se basant sur les observations réalisées au cours de la revue de littérature, l'objectif général déjà cité en section 1.3 a été défini comme suit : concevoir et évaluer sur des critères subjectifs et objectifs une approche ajustant l'affichage stéréoscopique en fonction des mouvements oculaires. Il s'agit en particulier de développer un logiciel permettant de simuler et de comparer plusieurs conditions d'affichage stéréoscopique dans un environnement de réalité virtuelle basées sur la détermination précise du point de regard du participant.

2.6.3 Notre choix de paramètres

Configuration parallèle des caméras

En raison de la configuration architecturale de la voûte de réalité virtuelle, composée de 3 murs et d'un sol perpendiculaires les uns aux autres, nous avons choisi de ne pas faire converger les caméras mais d'utiliser la configuration parallèle. Deux caméras virtuelles

seront positionnées à l'emplacement des deux yeux de l'observateur, ou légèrement décalées en fonction de la configuration, pour chacun de ces écrans (soient huit caméras virtuelles au total), et leurs cônes de projection seront calculés en fonction de la position de l'individu dans la voûte.

Utilisations possibles de la connaissance du point de regard

L'idée qui sous-tend cette recherche est d'utiliser la connaissance du point de regard afin de jouer sur les paramètres de configuration des caméras et d'améliorer à la fois le confort et le réalisme de l'immersion. Comme vu précédemment, cette information peut être utilisée pour :

- Faire varier la distance inter-caméras et donc l'amplitude des disparités stéréoscopiques.
- Générer un flou de profondeur de champ qui simulerait l'accommodation, c'est-à-dire l'effet de la déformation du cristallin sur la vision.
- Faire converger les caméras virtuelles vers ce point, c'est-à-dire les orienter dans l'espace de manière à ce que leurs axes s'intersectent en ce point.
- Enfin, sortant du cadre de cette maîtrise mais s'inscrivant dans le projet à long terme de l'Institut Philippe-Pinel de Montréal, la connaissance du point de regard peut être utilisée à des fins thérapeutiques par les cliniciens de l'Institut, au cours de tests diagnostiques ou de séances de traitements.

Comme présenté dans la revue de littérature, les trois premières utilisations citées ci-dessus influent sur le confort et également sur le réalisme de la perception du monde virtuel (estimation des distances, sentiment de présence, etc.). Étant donné que nous avons écarté la configuration convergente, nous nous sommes concentrés sur la distance inter-caméras et le flou de profondeur de champ.

Distance inter-caméras

Dans l'environnement immersif de l'Institut Philippe-Pinel de Montréal, nous utiliserons un dispositif de suivi qui nous fournira les positions et les vecteurs de direction de vue des deux yeux dans l'espace. Notre système intégrera cette information afin de calculer la position du point de regard de l'observateur, puis de modifier en temps réel la distance entre les caméras virtuelles, en particulier afin de correspondre aux préférences de l'utilisateur, et de situer le flou de profondeur de champ autour de ce point.

Flou de profondeur de champ

Hillaire et al. ont observé que la navigation dans un environnement virtuel monoscopique après ajout d'un flou de profondeur de champ basé sur le suivi de regard était significativement préférée à celles sans flou ou avec un flou calculé sans tenir compte du point de regard (voir la section 2.4.2). Notre implémentation faisant partie de la première catégorie, nous ne devrions profiter que des avantages du flou de profondeur de champ, à savoir un environnement virtuel plus réaliste, un sentiment d'immersion et un plaisir de navigation accru, sans souffrir des inconvénients de celui-ci tels que la gêne ressentie lorsque le flou ne suit pas le regard et affecte l'objet réellement observé.

2.6.4 Objectifs spécifiques

L'objectif général a été décomposé en 7 objectifs spécifiques :

- Mesurer la direction du regard de l'observateur et sa vergence oculaire.
- Déterminer la position 3D du point de regard de l'observateur.
- Concevoir une méthode de modification de l'affichage stéréoscopique en fonction de cette mesure, basée sur deux approches : la distance inter-caméras et le flou lié à la profondeur de champ.
- Évaluer subjectivement notre méthode d'ajustement de l'affichage stéréoscopique, en termes de confort, réalisme du rendu, plaisir de navigation, perception des distances et des profondeurs, sentiment de présence dans le monde virtuel.
- Évaluer objectivement notre méthode d'ajustement de l'affichage stéréoscopique, en comparant les limites de fusion des images stéréoscopiques.
- Évaluer objectivement notre méthode d'ajustement de l'affichage stéréoscopique en termes de perception des distances et des profondeurs, en comparant la vergence oculaire ainsi que l'écart angulaire à la cible entre une scène réelle et une scène virtuelle identique.
- Concevoir des outils efficaces à destination des cliniciens, leur permettant de visualiser les informations liées à l'attention visuelle du patient et d'en extraire des données statistiques.

2.6.5 Hypothèses de travail

H1 : La précision et la stabilité des mesures de positions et de vecteurs de direction de vue fournis par les ETG permettent de déterminer précisément la position du point de regard de l'individu.

Originalité : L'objet d'attention est souvent déterminé dans la littérature grâce à la di-

rection de regard d'un seul œil. Or dans une même direction, le regard peut se porter sur plusieurs objets à différentes profondeurs de champ. L'information apportée par le vecteur de vue du deuxième œil grâce à un système de suivi binoculaire, et qui permet de lever en partie l'ambiguïté sur l'objet d'attention, n'a été que peu étudiée.

Réfutabilité : L'hypothèse est réfutée si les mesures fournies par les ETG ne permettent pas de déterminer la position du point de regard à 10 cm près à 2 m de l'observateur.

H2 : La précision et la stabilité des mesures de positions et de vecteurs de direction de vue fournis par les ETG permettent d'ajuster les paramètres de l'affichage stéréoscopique.

Originalité : L'affichage stéréoscopique dans un environnement virtuel tel qu'une CAVE est généralement calculé sans considération du point de regard de l'individu. Or l'ajustement des paramètres en fonction de la position de ce dernier permet de modifier l'affichage en fonction de l'objet regardé, ce qui, à notre connaissance, n'a pas été testé dans un tel environnement.

Réfutabilité : L'hypothèse est réfutée si les mesures fournies par les ETG ne permettent pas d'ajuster suffisamment précisément ou suffisamment rapidement les paramètres de l'affichage. En particulier, si le rapport signal sur bruit est trop important, il se peut que la modification de la DIC ou l'ajout de flou de profondeur de champ en temps réel se révèle inconfortable. Il faudra déterminer la précision et la stabilité de cette mesure, et évaluer l'impact sur la gêne des participants.

H3 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard influe positivement sur ses impressions subjectives.

H4 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu influe positivement sur ses impressions subjectives.

H5 : Les deux modifications citées précédemment, réalisées simultanément, influent positivement sur les impressions subjectives de l'individu.

Originalité : Dans ce chapitre, nous avons détaillé certaines techniques visant à améliorer l'expérience d'immersion, telle que la réduction des hautes fréquences spatiales par ajout de flou pour accroître le confort, ou la modification de la DIC pour augmenter l'intervalle de fusion. Cependant, à notre connaissance, les modifications de l'affichage que nous proposons n'ont soit jamais été testées (ajustement de la DIC par interpolation en fonction de la distance au point de regard de l'individu) soit jamais été comparées les unes aux autres au cours d'une même expérimentation.

Réfutabilité : chaque hypothèse est réfutée si les notes subjectives attribuées à la configuration correspondante ne sont pas significativement meilleures que celles octroyées à la configuration de contrôle.

H6 : Permettre à l'individu de choisir de manière personnelle la distance inter-caméras influe positivement sur ses impressions subjectives.

H7 : le ratio entre la distance interpupillaire de l'individu et son choix personnel de distance inter-caméras influe sur ses impressions subjectives.

H8 : L'hésitation de l'individu lors de son choix personnel de distance inter-caméras influe sur ses impressions subjectives.

Originalité : Nous n'avons pas connaissance d'expérimentations ayant offert la possibilité à l'individu de choisir sa DIC puis comparé ce choix à d'autres valeurs. *A fortiori*, aucune n'a non plus analysé les résultats en fonction de la différence entre la DIC choisie et la DIP anatomique, ni étudié l'hésitation lors du choix. Pourtant, le ratio DIP/DIC influence directement les disparités perçues par l'observateur et donc sa perception de la profondeur. De même, lorsque l'individu choisit à plusieurs reprises des valeurs similaires pour une distance inter-caméras, on peut supposer que c'est parce que cet écart est clairement plus confortable ou plus réaliste pour lui.

Réfutabilité : **H6** est réfutée si les notes subjectives attribuées à la configuration correspondante ne sont pas significativement meilleures. **H7** et **H8** le sont si aucune relation ne lie respectivement le ratio DIP/DIC ou l'écart-type entre les choix successifs de DIC avec les notes subjectives attribuées à la configuration correspondante.

H9 : La propension à l'immersion de l'individu influe positivement sur ses impressions subjectives.

H10 : La quantité d'inspection visuelle de l'individu, en particulier le nombre et la durée des fixations et des saccades ainsi que le nombre d'objets regardés, influe positivement sur ses impressions subjectives.

Originalité : La propension à l'immersion est évaluée à l'aide d'un questionnaire, et fournit une mesure de la capacité théorique de l'individu à s'immerger dans un environnement virtuel. Il est donc raisonnable de supposer que plus sa propension à l'immersion est importante, plus ses notes sur le critère d'immersion seront élevées. De même, concernant la quantité d'inspection visuelle, la découverte d'un plus grand nombre d'objets dans la scène peut logiquement conduire à des notes de plaisir plus élevées. Nous n'avons pas connaissance d'expérimentations ayant analysé la relation entre ces données et les notes subjectives attribuées par les participants.

Réfutabilité : **H9** et **H10** sont réfutées si aucune relation ne lie positivement respectivement la propension à l'immersion ou la quantité d'inspection visuelle avec les notes subjectives attribuées.

H11 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard augmente l'intervalle proche de profondeur dans lequel il parvient à fusionner les images stéréoscopiques.

H12 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu augmente l'intervalle proche de profondeur dans lequel il parvient à fusionner les images stéréoscopiques.

H13 : Les deux modifications citées précédemment, réalisées simultanément, augmentent l'intervalle proche de profondeur dans lequel l'individu parvient à fusionner les images stéréoscopiques.

Originalité : Bien que nous ayons recensé dans la littérature plusieurs études se penchant sur l'intervalle de fusion stéréoscopique, en jouant sur la DIC et le flou de profondeur de champ, aucune n'utilisait la modification de la DIC que nous proposons. En particulier, l'individu basera son choix de l'une des 3 DIC sur l'observation d'objets proches, la fusion des images stéréoscopiques calculées à partir de ce choix pouvant alors logiquement s'effectuer plus facilement.

Réfutabilité : Chaque hypothèse est réfutée si la limite de fusion, c'est-à-dire la distance minimale entre l'observateur et l'objet pour laquelle il parvient à fusionner les images, correspondant à la configuration n'est pas significativement plus petite que celle de la configuration de contrôle.

H14 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard conduit à des réponses de vergence oculaire plus proches de la réalité.

H15 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu conduit à des réponses de vergence oculaire plus proches de la réalité.

H16 : Les deux modifications citées précédemment, réalisées simultanément, conduisent à des réponses de vergence oculaire plus proches de la réalité.

Originalité : Aucune recherche à notre connaissance n'a testé les réponses de vergence oculaire entre une scène réelle et une scène virtuelle identique. Pourtant, ces réponses fournissent des informations objectives sur la perception de la profondeur par l'individu, celui-ci faisant converger ses yeux à la distance où il perçoit l'objet.

Réfutabilité : Chaque hypothèse est réfutée si la différence entre l'angle de vergence obtenu dans le cas réel et celui obtenu avec la configuration virtuelle correspondante n'est pas significativement plus faible que pour la configuration de contrôle.

H17 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard conduit à un écart angulaire à la cible plus proche de la réalité.

H18 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu conduit à un écart angulaire à la cible plus proche de la réalité.

H19 : Les deux modifications citées précédemment, réalisées simultanément, conduisent à un écart angulaire à la cible plus proche de la réalité.

Originalité : Comme pour les hypothèses **H14**, **H15** et **H16**, il s'agit ici de comparer les écarts angulaires entre la cible et le point de regard dans une scène réelle et une scène virtuelle identique. Cela n'a jamais été testé à notre connaissance. Pourtant, cet écart fournit lui aussi des informations sur le réalisme du rendu des distances et des profondeurs de la scène virtuelle, en révélant par exemple qu'un objet situé à 30 cm à droite du plan médian de la tête de l'observateur sera en réalité perçu à 50 cm.

Réfutabilité : chaque hypothèse est réfutée si la différence entre l'écart angulaire à la cible obtenu dans le cas réel et celui obtenu avec la configuration virtuelle correspondante n'est pas significativement plus faible que pour la configuration de contrôle.

CHAPITRE 3

MÉTHODOLOGIE

3.1 Matériel

3.1.1 Voûte de l'Institut Philippe-Pinel de Montréal

iCUBE-4

La voûte de réalité virtuelle de l'Institut Philippe-Pinel est un modèle iCUBE-4 réalisé par Viz-Tek. Les écrans (gauche, droite, devant et sol) sont en verre acrylique rigide recouvert d'une couche pour la diffusion. Ce matériau, StarGlas 100, prévu pour les projections par l'arrière, a été spécialement conçu pour les applications de mélange aux bords¹ et s'adapte donc très bien aux projections multiples (plusieurs projecteurs par écran). Étant donné que la projection s'effectue également par l'arrière pour le sol, la voûte entière est surélevée sur une plateforme, comme nous pouvons le voir sur la figure 3.1. Les écrans ont une dimension de 2,44 m x 3,14 m² pour l'écran de devant et le sol, et 2,44 m x 2,44 m pour les écrans de gauche et de droite.

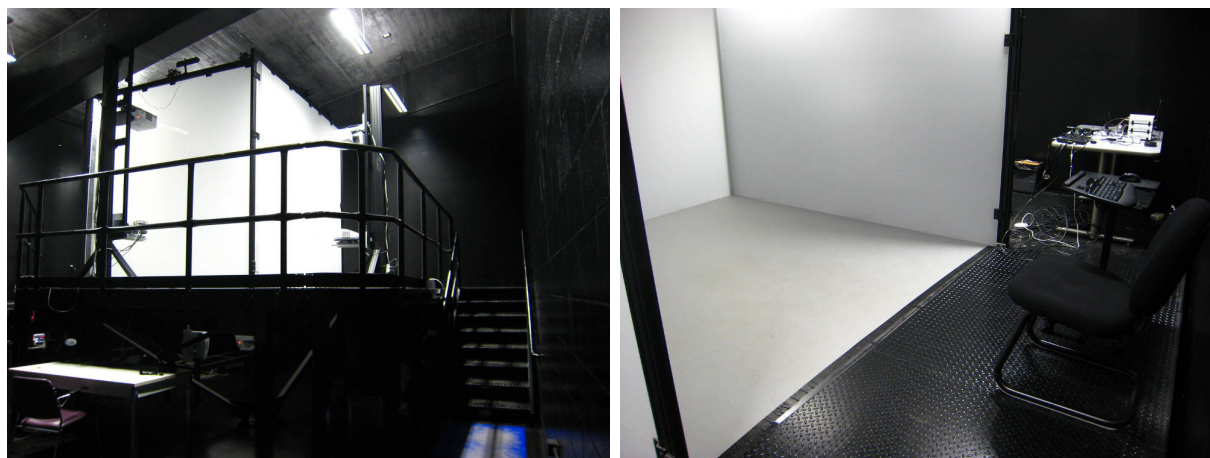


Figure 3.1 Installation de la voûte de réalité virtuelle de l'Institut Philippe-Pinel.

1. En anglais *edge blending*.

2. 8.01 po x 10.30 po.

Carte graphique et projecteurs

La carte graphique utilisée est une Quadro Plex 7000 double (2 blocs de 8 sorties) de la marque Nvidia, qui permet d'atteindre 36 mégapixels au travers de 8 systèmes d'affichage. Elle est associée à 8 projecteurs NEC U310W, utilisés à 120 Hz et 1280 x 720 px, soit une résolution totale de 5120 x 1440 px. La configuration attribuant deux projecteurs à chaque écran, visible en figure 3.2, a été choisie en raison des lieux trop exigus pour éloigner suffisamment les projecteurs autour de la voûte.

La fusion des images est réalisée à l'aide du logiciel AnyBlend (version 2) de VISIO(Visio, 2012), qui peut calculer une sortie jusqu'à 12 projecteurs avec la résolution native. Il réalise automatiquement un mélange doux aux bords³ et la déformation des projections⁴ en quelques minutes à l'aide d'une caméra présentée en figure 3.2, ce qui permet de projeter sur des écrans de forme arbitraire. Il est compatible avec tous les projecteurs et fonctionne sous Windows 7.



Figure 3.2 Deux projecteurs par écran (en haut à gauche). Caméra utilisée par AnyBlend (en bas à gauche) pour réaliser la fusion des images (à droite).

3. En anglais *softedge blending*.

4. En anglais *projector warping*.

Volfoni

Les lunettes stéréoscopiques actives qui furent fournies avec la voûte de l’Institut Philippe-Pinel sont de la marque Volfoni, plus particulièrement du modèle EDGE VR Infrarouge (Volfoni, 2014a), mais le montage des lunettes finales présenté sur la figure 3.6 a été réalisé avec un modèle ActiveEyes IR (Volfoni, 2014b). Le poids de ce dernier est de 29 g, et il offre un champ de vision de 170° horizontalement et 115° verticalement. Ses cristaux liquides permettent d’obtenir des images très lumineuses (17% de lumière résiduelle) et d’éviter les effets d’images fantômes⁵. Un appareil électronique à clipser au col permet de recevoir les signaux IR envoyés par un émetteur situé au-dessus de l’écran de devant pour s’ajuster à la fréquence d’affichage.

Système de repérage

Le système de repérage est un système optique OptiTrack (Optitrack, 2014a) fonctionnant sous Windows 7 avec une précision sous-millimétrique et un temps de latence de 8,3 ms. Il comporte 8 caméras Flex13, visibles en figure 3.3, dont la fréquence d’images est ajustable de 30 à 120 FPS. Elles sont pourvues autour de leur lentille d’une série de DEL IR, et captent les réflexions de la lumière émise par ces dernières sur les marqueurs installés sur les lunettes ou tout objet dont on souhaite connaître la position et l’orientation.



Figure 3.3 Système de repérage optique à 8 caméras.

MotiveTracker (Optitrack, 2014b) est le logiciel développé par Optitrack permettant de récupérer les informations du système de repérage. Il offre la possibilité de définir les entités à suivre, les corps rigides⁶, en enregistrant les positions relatives de leurs marqueurs. Le centre du repère local de celui-ci est positionné au centre de masse de l’ensemble des marqueurs,

5. En anglais *ghosting*.

6. En anglais *rigid body*.

tandis que son orientation est définie au moment de son enregistrement comme étant alignée avec le repère de la voûte, c'est-à-dire avec les vecteurs « avant », « haut » et « droit » respectivement égaux à $(0,0,1)$, $(0,1,0)$ et $(1,0,0)$.

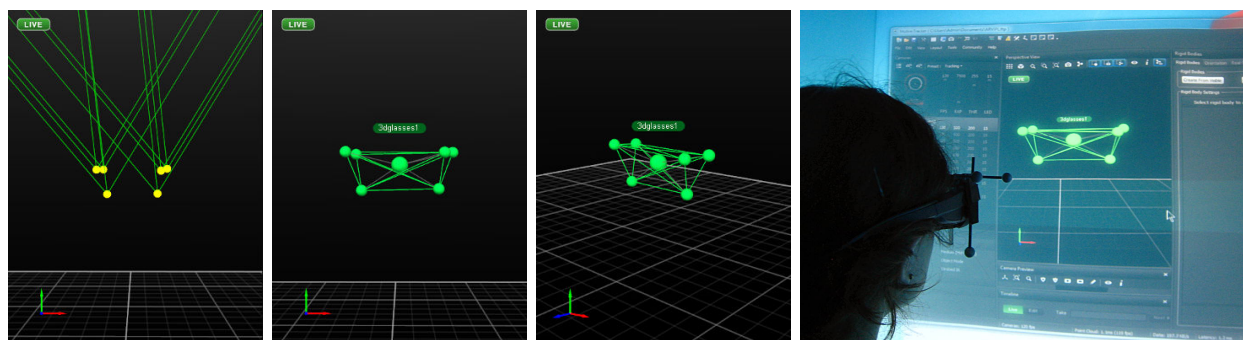


Figure 3.4 Captures d'écran de MotiveTracker : enregistrement du corps rigide des lunettes, représentant la tête de l'observateur.

Unity 3D 4.3.4

Actuellement utilisé pour développer les applications servant aux expérimentations réalisées au sein de l'Institut Philippe-Pinel, Unity est un engin de jeu qui « comprend un engin de rendu puissant complètement intégré et un ensemble complet d'outils permettant de créer des environnements 3D interactifs »⁷ (Unity, 2014). Il intègre 3 langages de programmation : Boo, javascript et C#. Fourni en version professionnelle avec l'iCUBE-4, Unity ne se cantonne plus aujourd'hui aux limites d'un engin de jeu classique, mais constitue un véritable cadre d'application⁸. En outre, la version professionnelle offre la possibilité d'assigner à une caméra un composant **Depth of Field** grâce au paquet **Images Effects**. La communauté Unity est très active et de nombreux modules d'extension⁹ étoffent ses capacités.

MiddleVR

MiddleVR est un « module d'extension générique de réalité virtuelle immersive »¹⁰ (I'm in VR) conçu pour fonctionner entre autres avec Unity 3D (I'm in VR, 2014). Il supporte la liaison avec les systèmes d'interaction 3D, tels que les systèmes de suivi, et permet la synchronisation de l'affichage sur plusieurs écrans. C'est donc lui qui effectue la transition entre le système de repérage et Unity 3D, via un plugin Unity. Ce dernier comporte un

7. Notre traduction.

8. Traduction de l'Office québécois de la langue française de l'anglais *application framework*.

9. En anglais *plugins*.

10. Notre traduction

`GameObject` Unity, `VRManager`, auquel est attaché un certain nombre de scripts. Lorsque l'application est lancée, les traqueurs inscrits dans le fichier de configuration se retrouvent sous le `VRManager` dans le graphe de scène et sont accessibles comme des `GameObject`, ce qui permet d'obtenir leur position et leur orientation en temps réel via leur champ `transform`. Il est possible de définir un gabarit de caméra¹¹ dont les propriétés seront utilisées par les caméras virtuelles servant à calculer les images projetées sur les murs. Dans notre cas, 8 caméras sont associées à la tête : 2 pour chaque mur, représentant l'œil gauche et l'œil droit. Leurs troncs de projection sont automatiquement redéfinis lorsque l'observateur se déplace dans la voûte de manière à systématiquement intersecter les 4 coins des écrans. Un script attaché à notre gabarit de caméra permet de lui assigner en fonction de la configuration courante une profondeur de champ. C'est également via le module d'extension MiddleVR que nous récupérons les entrées clavier et souris et que nous changeons la position des caméras, comme nous le verrons en section 3.3.3.

L'interface graphique de MiddleVR permet entre autres de définir la configuration des dispositifs d'affichage (taille des écrans, dispositions, les caméras stéréoscopiques, etc.), d'indiquer les traqueurs VRPN présents dans la configuration (correspondant aux corps rigides enregistrés avec MotiveTracker) ou encore de lancer une application Unity. Nous pouvons remarquer sur la figure 3.5 la redéfinition des troncs de projection des caméras virtuelles en fonction de la position de la tête dans la voûte.

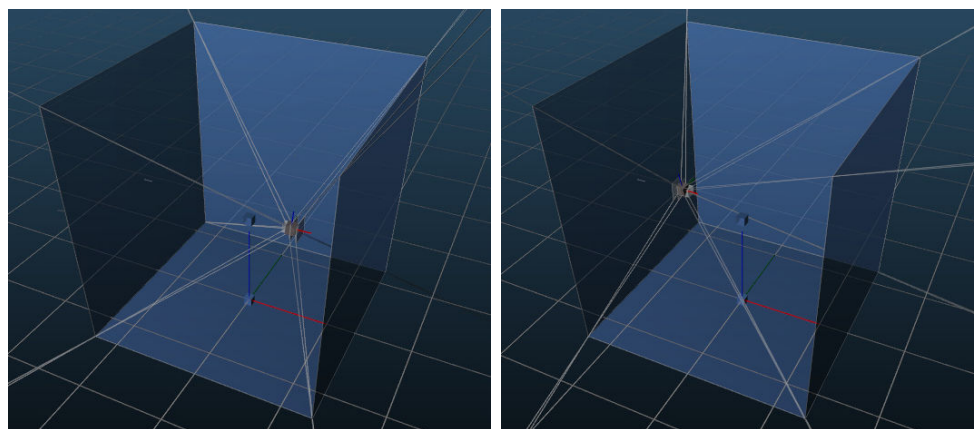


Figure 3.5 Configuration de la voûte dans MiddleVR, et modification des troncs de projection en fonction de la position de l'observateur.

11. En anglais *template camera*.

Eye tracking glasses

Afin de réaliser le suivi oculaire, nous avons utilisé les ETG 2.0 mises au point par SMI. Celles-ci comportent deux caméras fonctionnant au choix à 30 ou 60 Hz, placées dans les montures et dirigées vers les yeux, dont les images permettent de calculer les données relatives au suivi. Filaires, les ETG se connectent en USB et sont fournies avec le logiciel iViewETG chargé du traitement d'images. Il fonctionne en utilisant une modélisation sophistiquée du comportement oculaire développé par l'équipe SMI. La méthode de suivi employée est celle de PCCR vue en 2.2.1, ici consistant à repérer la pupille et les reflets cornéens (jusqu'à 6 par œil) avec compensation automatique de la parallaxe. Un modèle est ensuite appliqué afin de déterminer le point de regard. Ce modèle s'adapte automatiquement aux caractéristiques physiques mesurées sur l'individu, en particulier au cours d'une étape de calibration automatique s'effectuant durant les 15 premières secondes après le lancement du *quick run*, moment durant lequel il doit fixer un point dans le centre de son champ de vision. iViewETG se comporte ensuite comme un serveur transmettant les données de suivi oculaire.

En plus de fournir les vecteurs 3D de position des yeux et de direction de vue, les ETG indiquent la position du point de regard 2D, dont les coordonnées sont données dans le repère de l'image de scène fournie par une troisième caméra située à l'avant de la monture entre les deux yeux. Les spécifications renseignent une précision pour cette position de 0.5° toutes distances confondues dans les limites de profondeur (40 cm - infini) et dans un intervalle de 80° horizontalement et 60° verticalement. Elles indiquent aussi que les ETG fonctionnent avec les lentilles de contact (SMI, 2014), mais j'ai pu remarquer pour les participants en portant que la calibration automatique prenait plus de temps et n'arrivait parfois même pas à s'exécuter jusqu'à la précision maximale (celle-ci étant indiqué par la couleur du point de regard affiché dans la fenêtre de iViewETG, visible en figure 3.6). Le poids total des lunettes est de 68 g. Plus de détails sur les données de suivi oculaire seront fournis en section 3.3.1.

3.1.2 Voûte l'École Polytechnique de Montréal

L'École polytechnique de Montréal a fait l'acquisition d'une voûte de réalité virtuelle en 2000. Celle-ci comporte, tout comme celle de l'Institut Philippe-Pinel, trois écrans et un plancher disposés dans un cube de 3 mètres de côté. L'affichage s'effectue via quatre projecteurs de marque Optoma, qui projettent une image de résolution 1280 x 720 px à la fréquence de 120 Hz. La projection est réalisée par-derrière pour les trois écrans verticaux et par dessus pour le plancher, à l'aide de réflexion sur des miroirs.

Le repérage de la position et de l'orientation de la tête de l'observateur est accompli par un dispositif électromagnétique de repérage géométrique à six degrés de liberté « Flock of

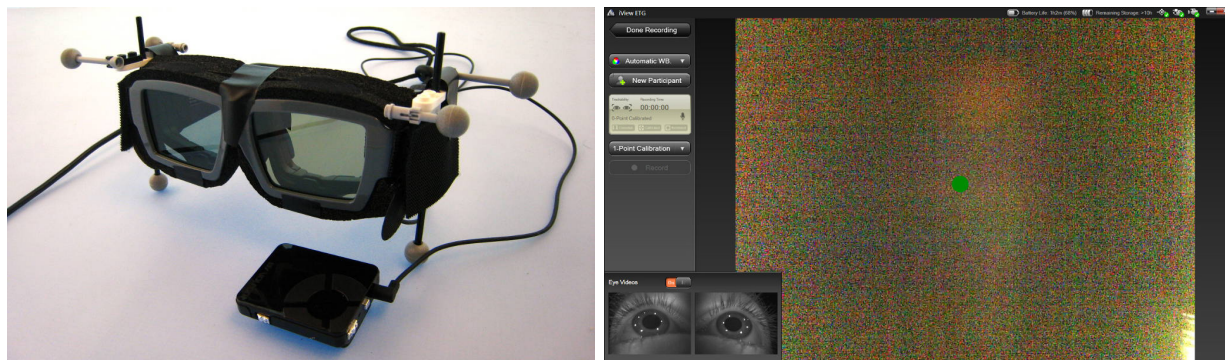


Figure 3.6 Le montage ETG et Volfoni (à gauche) et l'interface utilisateur de iViewETG (à droite).

Birds », fabriqué par Ascension Technologies. Il comprend « une unité qui émet un champ électromagnétique de manière continue (*transmitter*) et un certain nombre de capteurs fixés à l'aide de colliers sur tout élément physique dont on souhaite connaître la position et l'orientation » (École polytechnique de Montréal, 2010-2014). Ces informations sont récupérées via un serveur VRPN par l'ordinateur pilotant la voûte, qui détermine alors la position de l'observateur par rapport à l'unité émettrice et calcule en utilisant VRJuggler la configuration du monde virtuel à afficher. Une grappe de cinq PC est assignée au calcul des images stéréoscopiques, quatre d'entre eux étant affectés à un projecteur. Le système d'exploitation utilisé est un système Unix.

3.2 Tests préliminaires

La revue de littérature nous a permis d'identifier les différentes configurations possibles pour le système de caméras, ainsi que les paramètres intervenant dans ces configurations. Il est nécessaire pour ces derniers de restreindre leur champ de valeurs possibles dans le but de les comparer lors de l'expérimentation avec participants. Par exemple, l'intensité de la profondeur de champ peut prendre des valeurs continues (comprises entre un flou inexistant et un flou très important), tout comme l'intervalle de profondeur qui restera net (celui-ci pouvant être réduit à un plan ou faire plusieurs mètres d'épaisseur par exemple). Aussi, nous avons réalisé des tests préliminaires dans la CAVE de l'École Polytechnique de Montréal, dont la configuration est proche de celle de l'Institut Philippe-Pinel, afin d'explorer le réalisme du rendu et le confort avec différentes configurations du système de caméras. Bien que faisant partie de ces tests préliminaires, la configuration convergente a été abandonnée en raison de la configuration architecturale des environnements immersifs de type CAVE, tel que nous l'expliquerons en section 3.2.1. Ceux réalisés pour déterminer les paramètres de flou

de profondeur de champ se sont déroulés à la voûte de l'Institut Philippe-Pinel de Montréal, avec un participant extérieur.

3.2.1 Tests des configurations du système de caméras

Dispositif

Le dispositif pour les tests préliminaires a été simplifié de manière à se concentrer uniquement sur le rendu des configurations. Ainsi, seul l'écran du fond a été utilisé. La position ainsi que l'orientation de la tête du sujet n'étaient pas suivies, mais considérées comme fixes : au centre de l'écran et à 1,89 m de distance de celui-ci, les trois axes de son orientation coïncidant avec ceux du repère de la CAVE. Son point de regard, utilisé pour certains calculs, en particulier dans le cas de configurations convergentes, n'a pas été suivi avec les lunettes ETG qui seront utilisées à l'Institut Philippe-Pinel, mais présumé correspondre au centre du cube de la scène affichée. Celle-ci a elle aussi été choisie la plus simple possible : un cube translucide, pouvant tourner sur lui-même ou se déplacer dans l'espace, est affiché devant 2 rectangles rouges situés en profondeur au niveau de l'écran. Le cube mesure 3 cm de côté afin de faciliter pour le participant la visée de son centre. La fenêtre d'affichage, mesurant 2,89 m x 1,93 m¹², a été centrée horizontalement dans l'écran et positionnée verticalement en fonction de la taille de l'observateur, de manière à ce que sa ligne de vue intersecte le centre lorsqu'il regarde devant lui.

Le programme de test, basé sur le programme pulsar de Bourke(Bourke, 1999b) a été codé en C++ avec la bibliothèque graphique OpenGL. Les propriétés des deux caméras virtuelles sont définies en utilisant les mesures précises citées ci-dessus. Leurs plans de projection étant fixés au niveau de l'écran, la distance focale qui les sépare de ceux-ci est celle entre la tête de l'observateur, ou plus précisément la marque au sol correspondant à sa position verticale, et l'écran. Les dimensions du tronc de projection au niveau de la distance focale sont alors égales à celles de la fenêtre d'affichage. L'utilisation de ces mesures pour les calculs des angles d'ouverture des caméras et du ratio des dimensions des troncs de projection assure la correspondance entre les distances réelles et virtuelles. Mis à part dans le cas de la simulation des globes oculaires, les caméras sont placées dans la scène virtuelle le long de la droite reliant les deux yeux.

Configurations du système de caméras

Les quatre configurations comparées au cours de ces tests préliminaires sont les suivantes :

- Caméras parallèles, troncs asymétriques

12. 9,48 po x 6,33 po.

- Caméras parallèles, troncs symétriques
- Caméras parallèles, troncs asymétriques et simulation des globes oculaires
- Caméras convergentes

Pour le cas parallèle, le choix de troncs de projection symétriques ou asymétriques correspond aux configurations « parallèle » et « images translatées » décrites en section 2.3.1. Nous avons également eu l'idée de simuler les globes oculaires, en particulier les déplacements des pupilles intervenant lors des mouvements de convergence oculaire, en positionnant les caméras au niveau de celles-ci.

Les propriétés des caméras virtuelles qui diffèrent selon la configuration sont les suivantes : position de la caméra, distance focale, vecteur avant, vecteur haut et limites en X, Y et Z du tronc de projection (celles en X et en Y étant déterminées pour une distance selon l'axe Z donnée). Les prochains paragraphes illustrent la disposition des caméras ainsi que les images stéréoscopiques générées pour chaque configuration, le détail des calculs spécifiques des paramètres pouvant être consulté en annexe A. Le fait que la hauteur en Y des yeux de l'observateur coïncide avec celle de son point de regard et du milieu de la fenêtre simplifie ici les schémas et calculs, nous autorisant à rester dans le plan XZ.

Caméras parallèles, troncs asymétriques et symétriques

La figure 3.7 illustre l'influence de la DIC sur le rendu et les disparités entre les images stéréoscopiques. Deux DIC sont ici représentées, en lignes pleines et en pointillés, les couleurs bleue et orange permettant de distinguer les images destinées à l'œil gauche de celles destinées à l'œil droit. Dans le cas de la seconde configuration, les troncs des caméras sont symétriques et fixes. Les images gauche et droite obtenues sont ensuite projetées sur le même écran, ce qui revient à les translater, augmentant ainsi la disparité naturelle due à l'écartement entre les centres de projection, et conduisant à des disparités beaucoup plus importantes que dans le cas des troncs asymétriques. C'est ce phénomène qui nous a amenés à ne pas retenir cette configuration, malgré le fait que nous l'avions atténué en offrant la possibilité à l'observateur de choisir 3 DIC pour 3 profondeurs du point de regard (c'est-à-dire du centre du cube) différentes.

Caméras parallèles, troncs asymétriques et simulation des globes oculaires

L'idée nous est venue pour cette configuration d'essayer de reproduire le plus fidèlement possible le mouvement des yeux, en positionnant les caméras au niveau des pupilles et en modifiant donc leurs propriétés en fonction de l'orientation des yeux, c'est-à-dire en fonction du point de regard. La figure 3.8 illustre les différences entre les caméras gauche et droite

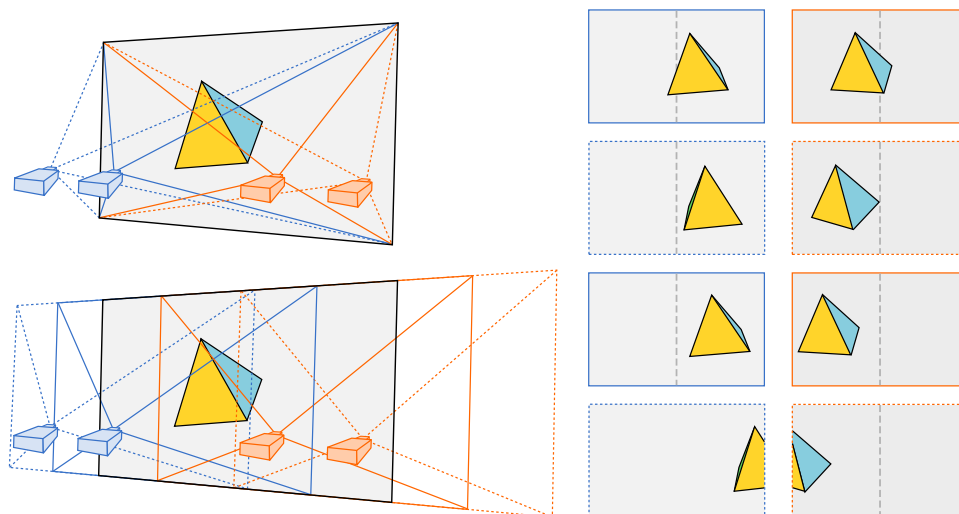


Figure 3.7 Impact de la distance inter-caméras sur les images stéréoscopiques obtenues avec la configuration parallèle, troncs asymétriques (en haut) ou troncs symétriques (en bas).

en raison du fait que le point de regard est excentré par rapport au plan médian. À noter que malgré la modification des troncs nous n'altérons pas l'orientation des caméras, qui sont toujours dirigées perpendiculairement à l'écran. Le rendu obtenu avec cette configuration ne nous ayant pas semblé suffisamment différentiable de celui atteint sans simulation du globe oculaire, nous avons finalement décidé de ne pas la retenir.

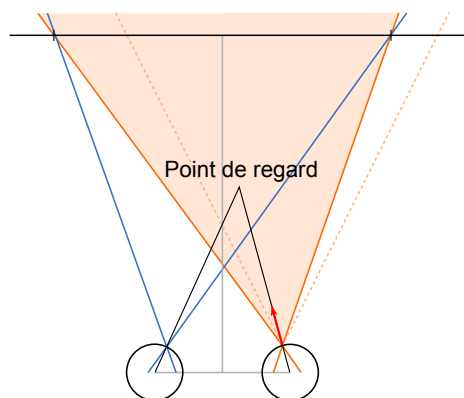


Figure 3.8 Positions et troncs de projection des caméras virtuelles dans le cas de la configuration parallèle, troncs asymétriques et simulation des globes oculaires.

Caméras convergentes

Contrairement aux autres configurations, ici les caméras virtuelles sont orientées vers le centre du petit cube utilisé dans la scène, et non perpendiculairement à l'écran de devant.

Leurs plans de projection ne sont donc plus parallèles à cet écran, mais le fait que le cube soit toujours situé à la hauteur des yeux de l'observateur assure qu'ils restent parallèles à l'axe Y. L'angle dièdre cause des distorsions trapézoïdales qui sont de plus en plus importantes à mesure que l'on s'éloigne de la droite d'intersection des plans le long de ceux-ci, tel qu'expliqué en section 2.3.2. Afin de diminuer les distorsions perçues par l'utilisateur, nous plaçons cette droite d'intersection en alignement avec son point de regard, en faisant se croiser les plans de projection sur la droite partant du centre des yeux et ayant pour direction la moyenne des directions des deux lignes de vues. Nous plaçons également l'intersection dans le plan de l'écran, de manière à ne pas générer de déformation de la taille du cube. En effet, si les plans s'étaient croisés au niveau du cube, alors l'affichage sur l'écran des images stéréoscopiques obtenues aurait conduit à un cube trop petit si celui-ci s'était trouvé devant l'écran, et trop grand s'il s'était trouvé derrière.

Cette configuration n'a pas été retenue car elle ne peut pas s'appliquer au cas d'une voûte de réalité virtuelle composée d'écrans plats positionnés à 90° les uns des autres. En effet, il faut considérer que les images vont devoir être calculées pour chacun des écrans. Or, comme nous pouvons le remarquer sur la figure 3.9, si l'une des caméras se retrouvait par exemple parallèle à l'un des écrans il serait impossible de calculer l'image correspondante à projeter sur celui-ci.

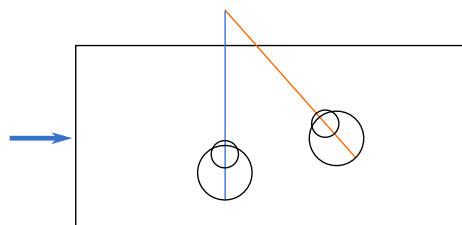


Figure 3.9 Exemple de cas où une image serait indéfinie pour l'un des écrans avec la configuration convergente.

Distance inter-caméras

Comme remarqué au cours de la revue de littérature, la distance inter-caméras est un paramètre influant fortement sur les performances d'une configuration, aussi bien en termes de confort que de perception de la profondeur (voir section 2.3.3). Ce paramètre intervient dans chacune des configurations présentées ci-dessus.

En dehors de la DIP anatomique, qui s'avère une valeur pertinente pour la DIC mais que nous n'avons pas explorée au cours de ces tests préliminaires, nous avons décidé de permettre à l'observateur de choisir lui-même sa DIC, à l'aide du clavier et selon les critères

cités ci-dessus.

Nous avons également envisagé qu’une seule DIC, fixe quelle que soit la distance par rapport à l’objet regardé, pouvait s’avérer trop restrictive. En effet, les disparités horizontales étant plus importantes en nombre de pixels sur l’écran pour des objets proches, un observateur pourrait trouver plus confortable de diminuer la DIC au moment de regarder ces objets, ou au contraire pourrait souhaiter l’augmenter afin de percevoir de manière plus importante leurs volumes. C’est pourquoi deux des configurations testées offraient la possibilité de choisir 3 DIC différentes pour 3 distances du point de regard, c’est-à-dire du cube. Une barre en bois verticale positionnée au niveau du cube, mais légèrement excentrée afin de ne pas occulter son affichage sur l’écran, faisait office d’indice réel de profondeur pour permettre à l’utilisateur de choisir les DIC selon sa perception des distances, en plus de son confort visuel. Une fois les 3 DIC enregistrées, une interpolation en fonction de la distance au cube permettait d’obtenir la DIC adéquate pour un point de regard situé à n’importe quelle profondeur dans l’espace.

Les impressions subjectives recueillies auprès d’un participant nous ont poussés à retenir, pour l’expérimentation finale, les choix d’une ou de trois DIC pour le paramètre de la distance inter-caméras, auxquels vient s’ajouter la DIP anatomique.

3.2.2 Intensité de la profondeur de champ

Le deuxième paramètre mis en exergue lors de la revue de littérature est la profondeur de champ, détaillée section 2.4. Comme expliqué précédemment, celle-ci comporte des propriétés pouvant prendre un ensemble continu de valeurs, en particulier l’intensité et l’intervalle de netteté. Pour ce dernier, après plusieurs essais, nous avons finalement décidé de maintenir la valeur par défaut du **Depth of Field Scatter** de Unity. La valeur de l’intensité a, elle, été choisie avec l’aide des retours d’un participant préliminaire au test final.

Pour conclure, les conditions expérimentales retenues regroupées dans le tableau 3.1 varient selon deux paramètres : la distance inter-caméras et la présence ou non d’une profondeur de champ.

Flou \ DIC	Anatomique	1 choix	3 choix
	Non	DIC1	DIC3
Oui	FLOU	FLOUDIC1	FLOUDIC3

Tableau 3.1 Les 6 configurations stéréoscopiques comparées au cours de l’expérimentation.

3.3 Calcul du point de regard 3D dans la scène virtuelle

Les ajustements de l’affichage stéréoscopique que nous proposons nécessitent de connaître le point de regard de l’observateur dans la scène virtuelle. Pour cela, nous utilisons les systèmes de suivi oculaire et de repérage de la tête présentés en section 3.1. Les données de ce dernier sont directement communiquées à l’application Unity par l’intermédiaire du module d’extension MiddleVR, tandis que nous avons dû réaliser l’intégration des ETG grâce à une bibliothèque dynamique et un module d’extension *ad hoc*. Le calcul du point de regard est enfin effectué grâce à la méthode de *raycasting* à la suite d’une calibration utilisant les multi-quadriques de Hardy.

3.3.1 Suivi oculaire et repérage de la tête

Collecte des données oculaires dans Unity

Afin de collecter les données de suivi oculaire transmises par le serveur iViewETG, nous avons développé une DLL utilisant le SDK de SMI, associée à un module d’extension qui permet de passer des fonctions C++ du SDK au C# des scripts Unity. Les fonctions principales listées en annexe B permettent de gérer la connexion au serveur iViewETG, la souscription au flux de données et la réception des rapports. Chacun intègre les informations de suivi suivantes :

- Les positions 3D des centres des yeux.
- Les vecteurs 3D de direction du regard, c’est-à-dire les vecteurs directeurs des axes optiques.
- Les mesures de confiance sur le repérage des pupilles pour les deux yeux.
- Les mesures d’incertitude sur l’estimation du globe oculaire pour les deux yeux.
- Le point de regard 2D final, ainsi que ceux calculés séparément pour chacun des deux yeux, dans le repère de l’image de scène.
- Le temps en ns.
- Le numéro de l’image de scène à laquelle correspond ce POR.

Parmi celles-ci, certaines ne sont pas stockées mais tout de même imprimées dans le fichier de log, telles que les mesures de confiance et d’incertitude qui donnent des indications sur l’assurance d’iViewETG quant à la qualité du suivi et de la détermination des points de regard. Les autres sont enregistrées dans une **struct** spécifique de notre DLL, laquelle est stockée dans une liste consultée par Unity de manière asynchrone.

Afin d’être le plus réactif possible aux mouvements oculaires du participant, c’est le rapport le plus récent parmi ceux associés au calcul d’une image qui est utilisé pour déterminer le point de regard dans la scène. Plus précisément, le plus récent dont l’ordre de priorité est

le plus important. Cet ordre de priorité est codé par un indice de validité déterminé par la DLL lors du stockage du rapport, qui peut prendre les valeurs suivantes :

- 4 : aucune donnée valide.
- 3 : les positions des yeux sont valides, mais pas les deux vecteurs de direction.
- 2 : un des vecteurs de direction est vers l'arrière.
- 1 : les vecteurs directeurs sont divergents dans le plan XZ.
- 1 : toutes les données sont valides.

Cela permet, dans le cas par exemple où l'observateur a cligné des yeux (conduisant à des valeurs NaN pour les positions et les vecteurs de direction des yeux) au cours du dernier rapport, de récupérer un rapport précédent dont les données étaient valides. Un identificateur unique (pour une connexion) attribué à chaque rapport permet de savoir celui qui a été utilisé lors du calcul du point de regard.

En raison de l'instabilité des données collectées (en particulier le fait que certains vecteurs de direction soient occasionnellement dirigés vers l'arrière), nous avons entrepris de tester deux techniques de lissage, l'une moyennant sur plusieurs images les positions successives du point de regard (5 et 10), l'autre utilisant la méthode **SmoothDamp** de Unity (des précisions sur cette dernière sont fournies en annexe B). Ces deux techniques causant une latence importante dans le suivi du point de regard, très gênant pour les ajustements en temps réel que nous effectuons, nous avons finalement préféré conserver les données bruitées mais réactives.

Couplage au repérage de la tête

Les marqueurs du montage ETG/Volfoni présenté en figure 3.6 ont été positionnés de manière à ce que leur centre de masse, donc l'origine du repère du corps rigide, coïncide avec l'origine du repère local des ETG visible en figure 3.10. L'orientation du corps rigide est, elle, initialisée au moment de son enregistrement dans MotiveTracker (voir section 3.1.1) avec les vecteurs « avant », « haut » et « droit » respectivement égaux à $(0,0,1)$, $(0,1,0)$ et $(1,0,0)$. Afin de s'assurer de la correspondance entre les repères, nous enregistrons le corps rigide associé aux lunettes en les orientant de manière à ce que les axes de leur repère local correspondent aux axes du repère de la voûte. Ainsi, la position des yeux et les vecteurs de direction transmis depuis le serveur iViewETG à Unity sont naturellement exprimés dans le repère de la tête, le passage des coordonnées locales aux coordonnées globales s'effectuant grâce à des fonctions Unity.

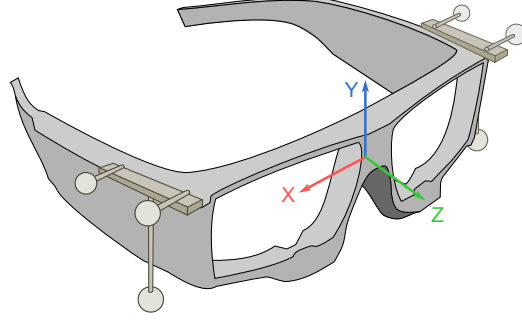


Figure 3.10 Repère local des lunettes ETG.

3.3.2 Calcul du point de regard

Intersection des lignes de vue

Dans un premier temps, nous avons tenté l'approche purement géométrique décrite en section 2.2.2, en positionnant le POR au centre du segment de taille minimale reliant les deux lignes de vue. Notons P_0 et P_1 les positions des yeux, V_0 et V_1 les vecteurs de direction. Soit

un point sur la droite (P_0, V_0) d'équation paramétrique $P_0(t_1) = P_0 + t_0 V_0 = \begin{pmatrix} P_{0x} + t_0 V_{0x} \\ P_{0y} + t_0 V_{0y} \\ P_{0z} + t_0 V_{0z} \end{pmatrix}$

et un point sur la droite (P_1, V_1) : $P_1(t_1) = P_1 + t_1 V_1 = \begin{pmatrix} P_{1x} + t_1 V_{1x} \\ P_{1y} + t_1 V_{1y} \\ P_{1z} + t_1 V_{1z} \end{pmatrix}$

Nous cherchons les paramètres $\hat{t}_0$ et $\hat{t}_1$ qui minimisent la fonction de distance, ou de distance au carré $D(t_0, t_1)^2 = \|P_1(t_1) - P_0(t_0)\|^2$. Le minimum est atteint lorsque les dérivées partielles de cette fonction quadratique à 2 variables s'annulent, ce qui fournit 2 équations linéaires à 2 inconnues. Leurs solutions permettent de calculer la position des deux points $P_0(\hat{t}_0)$ et $P_1(\hat{t}_1)$ dont la moyenne correspond au point dans l'espace le plus proche simultanément des deux droites. À noter que les deux vecteurs de direction ne doivent pas être parallèles, un test permettant de s'en assurer.

Cette méthode de calcul du POR s'est avérée ne pas convenir aux données retournées par les ETG. En particulier, plus l'individu écarte son point de regard du plan médian de sa tête, plus les coordonnées en Y des vecteurs de direction ont tendance à diverger (l'un des vecteurs « montant » ou « descendant » plus que l'autre). Cette différence s'inverse au passage d'une moitié de champ visuel à l'autre lorsque le participant balaye une droite horizontale, les coordonnées en Y étant les plus proches environ au niveau du centre. Nous avons également remarqué que ce phénomène intervient de manière beaucoup plus importante lorsque les lunettes ne sont pas positionnées parfaitement horizontalement par rapport aux yeux de

l'individu. Il a pour effet que le point le plus proche entre les deux droites se retrouve être le centre du segment joignant les deux yeux. Plusieurs solutions ont été tentées, sans succès, pour contrer cela : afficher les droites de direction du regard afin de régler la position des lunettes sur l'observateur pour que les vecteurs aient pratiquement la même coordonnée Y, ou calculer le point de regard avec les vecteurs $(x_0, \frac{y_0 + y_1}{2}, z_0)$ et $(x_1, \frac{y_0 + y_1}{2}, z_1)$. Par ailleurs, pour certains individus, les vecteurs de direction sont régulièrement légèrement divergents dans le plan XZ, cette méthode positionnant donc leur point de regard dans leur dos. Après des tests non concluants et dont les résultats s'avéraient très variables entre les individus, nous avons donc décidé de ne pas retenir cette méthode.

Il apparaît qu'il est très difficile d'estimer la profondeur du point de regard en tenant uniquement compte de la vergence. L'équipe d'SMI avait tenté puis abandonné cette approche, en concluant que la donnée de vergence n'était même pas consistante intra-individu. Dans la revue de littérature effectuée, les expériences réalisées utilisaient une discrétisation de l'espace en couches de profondeurs, ce que nous ne souhaitons pas dans notre cas. Duchowski et al. (2011) avaient bien obtenu une correspondance entre vergence et profondeur du point de regard, mais au moyen de filtre pour lisser les données très bruitées et d'un ajustage de celles-ci.

Raycasting

Nous avons finalement opté pour une méthode dite de *raycasting*, correspondant à celle citée en section 2.2.2, en intersectant les lignes de vues dirigées vers l'avant du participant avec la scène.

La précision obtenue sans aucun traitement n'étant pas satisfaisante, nous avons implémenté une méthode de calibration basée sur les multi-quadriques de Hardy (voir section 2.2.3). Pour réaliser celle-ci, l'observateur est invité à regarder successivement et dans un ordre aléatoire les n points d'une grille 2D affichée devant lui. Sont alors enregistrés pour chaque œil les vecteurs de direction P_i envoyés par les ETG, ainsi que les vecteurs « corrigés » correspondants Q_i : les vecteurs directeurs des droites reliant les centres des yeux aux points. Les couples (P_i, Q_i) conduisent pour chaque coordonnée au système d'équations suivant :

$$\begin{pmatrix} \omega_0(P_0) & \omega_1(P_0) & \dots & \omega_{n-1}(P_0) \\ \vdots & \vdots & \vdots & \vdots \\ \omega_0(P_{n-1}) & \omega_1(P_{n-1}) & \dots & \omega_{n-1}(P_{n-1}) \end{pmatrix} \begin{pmatrix} A_0 \\ \vdots \\ A_{n-1} \end{pmatrix} = \begin{pmatrix} Q_0 \\ \vdots \\ Q_{n-1} \end{pmatrix} \quad (3.1)$$

où $\omega_i(P)$ est pour rappel calculé comme suit :

$$\omega_i(P) = \sqrt{(P - P_i)^2 + R^2} \quad (3.2)$$

La résolution de ce système fournit les coefficients A_i , qui permettent à la réception de chaque nouveau vecteur P mesuré par les ETG de calculer le vecteur corrigé Q correspondant :

$$Q = \sum A_i \omega_i(P) \quad (3.3)$$

Le *raycasting* est ensuite réalisé en utilisant la méthode **Raycast** de Unity (détails en annexe B), qui fournit entre autres les coordonnées du point d'intersection et le nom de l'objet intersecté.

Compte tenu des données disponibles grâce aux ETG, indépendantes pour l'œil gauche et l'œil droit, nous avons eu le choix de réaliser le *raycasting* en utilisant plusieurs rayons, c'est-à-dire plusieurs couples (origine, direction)¹³ :

- PORG : (position de l'œil gauche, vecteur corrigé de direction de l'œil gauche)
- PORD : (position de l'œil droit, vecteur corrigé de direction de l'œil droit)
- POR : (moyenne des positions des deux yeux, moyenne corrigée des vecteurs de direction des deux yeux)

En raison des approximations lors de la résolution des systèmes linéaires (qui sont indépendants pour l'œil gauche et l'œil droit) et du calcul des vecteurs corrigés, il apparaît que le troisième vecteur n'est pas égal à la moyenne des deux premiers (c'est-à-dire que le point de regard correspondant à ce troisième rayon n'est pas situé au milieu des points obtenus avec les deux premiers). Aussi, nous pouvons définir un quatrième rayon : (moyenne des positions des deux yeux, moyenne des vecteurs gauche et droit obtenus après correction), fournissant un quatrième POR distinct : PORM.

Nous avons réalisé des tests de précision (le détail de la procédure est expliqué en section 3.4.2) afin de sélectionner celui qui serait utilisé comme point de regard final et servirait par exemple à déterminer la distance à l'observateur dont la valeur influe sur nos paramètres de configuration. En particulier, nous avons envisagé d'utiliser le rayon gauche lorsque l'observateur regarde plus vers la gauche ($|x_0| < |x_1|$) et le droit lorsqu'il regarde plus vers la droite ($|x_0| > |x_1|$), les ETG étant de moins en moins précises au fur et à mesure que la direction de l'œil s'écarte du centre (c'est-à-dire s'éloigne du vecteur (0,0,1)). Cependant, les histogrammes obtenus et regroupés en figure 3.11 ne confirment pas notre supposition, montrant plutôt une imprécision pour l'œil gauche souvent supérieure à celle de l'œil droit,

13. Notons que c'est aussi grâce au fait que les vecteurs gauche et droit soient considérés indépendamment qu'il n'est pas nécessaire d'utiliser une grille 3D pour la calibration. En effet, si le participant regarde à deux distances distinctes le long de la même ligne de vue gauche par exemple, le vecteur de direction droit sera différent étant donné la variation de vergence oculaire, et le POR correctement calculé.

et une imprécision globale plus basse pour la troisième proposition de rayon, qui fut celle retenue.

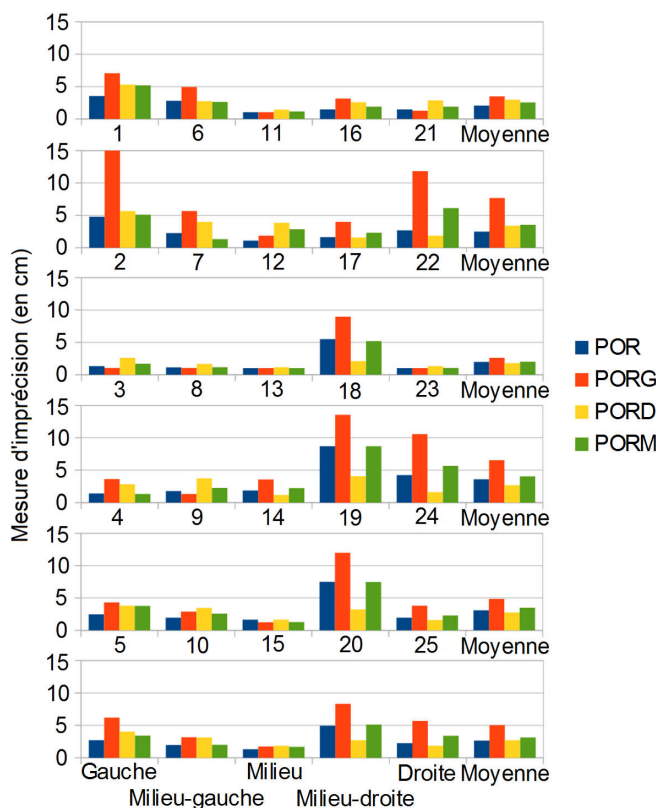


Figure 3.11 Mesures d'imprécision, moyenne et par indice de point de calibration (de 1 à 25 positionnés tel que sur la grille), obtenues avec les différents POR (en cm).

D'autres tests dont les résultats sont présentés en figures 3.12 et 3.13 nous ont permis de comparer plusieurs valeurs pour le paramètre R des multi-quadriques de Hardy et pour le nombre de points de calibration. Bien qu'il apparaisse évident que 9 points plutôt que 5 conduiraient à une meilleure précision, nous souhaitons aussi nous assurer que ce choix n'engendrait pas un temps de latence, du fait du plus grand nombre de calculs à chaque image. Pour l'expérimentation avec participants, nous avons choisi 9 points et un paramètre $R = 0$.

En dehors de la calibration automatique réalisée au lancement d'un *quick run* sur iViewETG, ce programme offre la possibilité d'effectuer manuellement une autre calibration. Il s'agit de demander à l'observateur de fixer 1 ou 3 points précis de son champ visuel, et de cliquer sur ceux-ci dans l'image filmée par la caméra de scène située à l'avant de la monture des lunettes et affichée dans la fenêtre de rendu du programme. Nous avons choisi de ne pas utiliser cette calibration, à la fois parce que les résultats d'imprécision sont légèrement plus

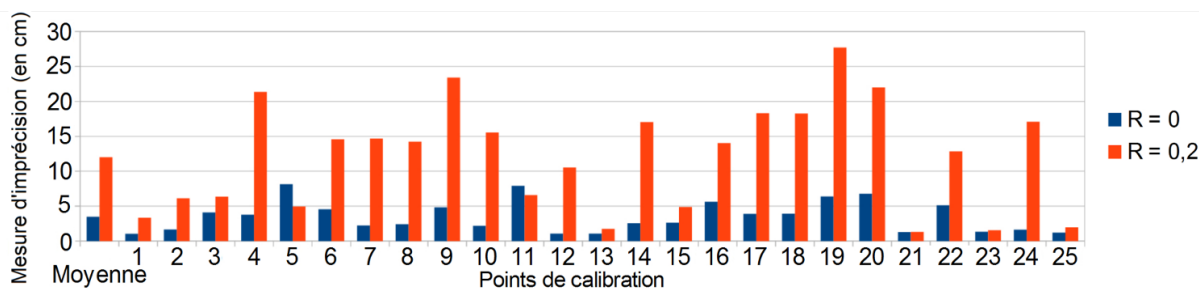


Figure 3.12 Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec différentes valeurs du paramètre R.

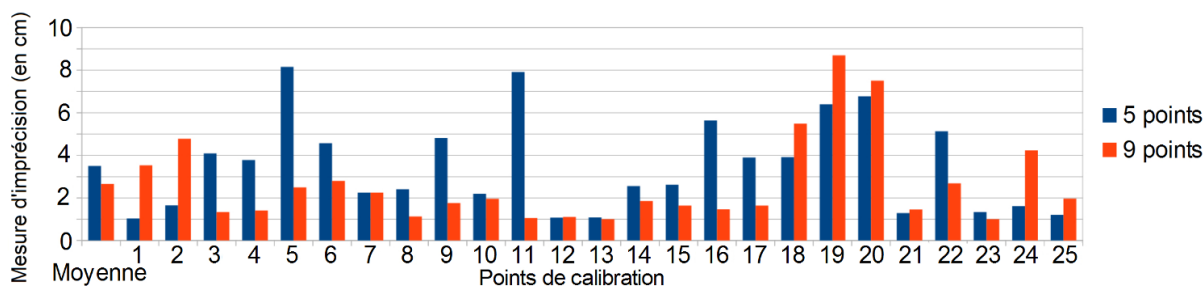


Figure 3.13 Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec 5 ou 9 points de calibration.

élevés avec, comme nous pouvons le voir sur la figure 3.14, mais aussi parce qu'elle nécessite de retirer puis réinstaller les Volfoni qui cachent la caméra de scène dans notre montage de lunettes. Or, si au cours de cette opération les lunettes bougent relativement à la tête du participant, ou si les marqueurs fixés aux ETG sont déplacés, il en résulte une augmentation de l'imprécision comme présentée sur la figure 3.15.

La méthode de *raycasting*, associée à notre procédure de calibration basée sur les multi-quadratiques de Hardy, nous a permis de diminuer l'imprécision de l'estimation du point de regard obtenue avec les ETG. Toutefois, il est important de noter que cette technique nécessite que chaque objet virtuel possède un *mesh collider* afin d'obtenir le point d'intersection entre le rayon et l'environnement, ce qui alourdit la scène.

3.3.3 Calcul de la position des caméras

Par le biais du module d'extension MiddleVR, les caméras virtuelles utilisées pour calculer les images stéréoscopiques sont accessibles et paramétrables dans Unity. Situées sous le *HeadNode* dans le graphe de scène, elles sont au nombre de 8 : une caméra gauche et une caméra droite pour chaque écran, orientées perpendiculairement à celui-ci. Leurs troncs de projection sont mis à jour à chaque image de manière à intersecter les quatre coins de l'écran,

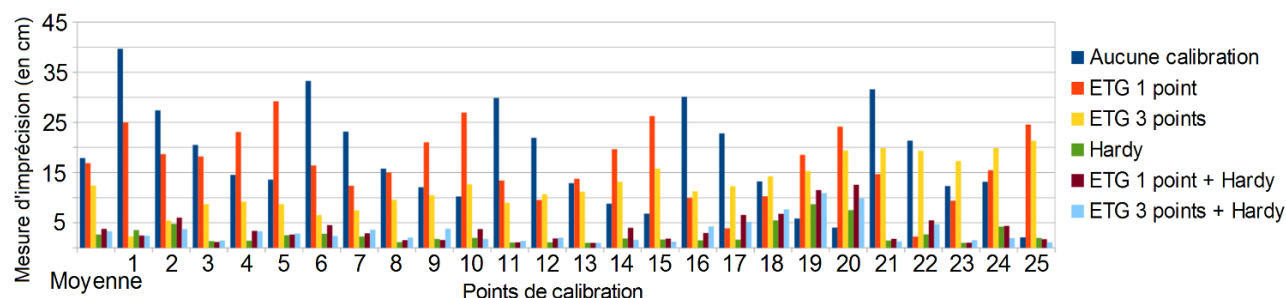


Figure 3.14 Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues avec différentes méthodes de calibration.

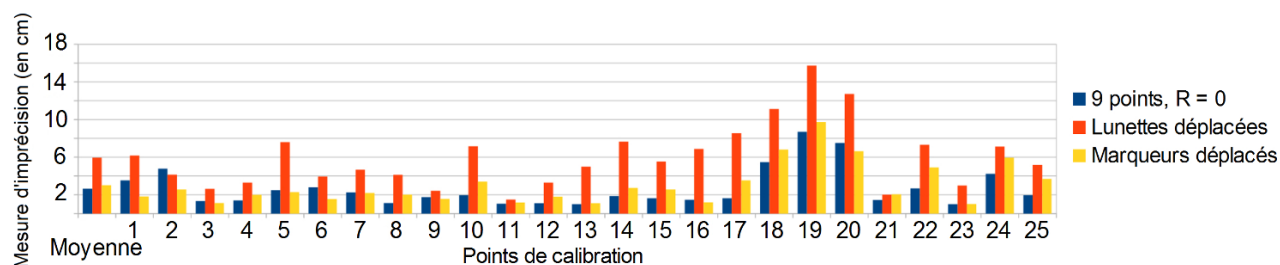


Figure 3.15 Mesures d'imprécision, moyenne et par indice de point de calibration, obtenues après déplacement des lunettes ou des marqueurs.

afin que le plan de projection corresponde toujours à ce dernier.

Le paramétrage de base de MiddleVR place les caméras de part et d'autre du centre du corps rigide représentant la tête, à une distance définie dans le fichier de configuration du système d'affichage. Il est cependant possible de modifier leurs positions, pour les faire coïncider par exemple avec les positions des yeux récupérées du serveur ETG. Étant donné que ces dernières sont exprimées naturellement dans le repère du **HeadNode**, comme nous l'avons vu plus haut, les caméras virtuelles seront alors placées aux positions anatomiques des yeux de l'observateur.

La distance inter-caméras étant un des paramètres de notre expérimentation, nous choisirons cependant parfois de les écarter d'une plus grande distance que la DIP physiologique, ou au contraire de les rapprocher, en fonction des choix des participants. Cette modification n'affectera que la coordonnée en X, les caméras restant placées au niveau des positions des yeux en Y et en Z afin de ne pas augmenter les disparités verticales entre les deux images stéréoscopiques. Mis à part dans les cas **DIC3** et **FLOUDIC3**, où la DIC varie à chaque image en fonction de la distance au POR, les positions locales des caméras ne seront modifiées qu'au moment des changements de configuration.

3.4 Expérimentation avec participants

La revue de littérature et les tests préliminaires nous ont permis d'identifier certains paramètres qui pouvaient améliorer la performance de la configuration du système de caméras. L'ajustement de leur valeur nécessitait en particulier la détermination en temps réel du point de regard de l'observateur, discutée ci-dessus. Nous souhaitons à présent évaluer les configurations stéréoscopiques résultantes, résumées dans le tableau 3.1.

Le protocole mis en place a pour triple objectif de recueillir les impressions subjectives des participants au regard de plusieurs critères de performance, d'évaluer leurs capacités de fusion avec les différentes configurations, et enfin de comparer les variations de vergence entre scène réelle et scènes virtuelles. L'expérimentation se divise donc en plusieurs étapes indépendantes visant chacune à atteindre l'un de ces objectifs, au cours desquelles les participants auront à effectuer différentes tâches. Après les phases de mise en place et de calibration du système, ils devront choisir les valeurs du paramètre de distance inter-caméras pour les configurations n'utilisant pas la distance interpupillaire anatomique. Ils navigueront ensuite au sein d'une scène virtuelle successivement avec les 6 configurations testées, et noteront chacune selon leurs ressentis. L'étape suivante consistera à déplacer deux objets vers l'avant ou vers l'arrière jusqu'à gain ou perte de la fusion stéréoscopique. Finalement, ils devront suivre une cible située au sommet d'un robot se déplaçant le long d'un parcours, à la fois dans le monde réel et dans le monde virtuel. Ce test, qui à notre connaissance n'a jamais été mené, permettra d'observer les divergences de mouvements oculaires engendrés par la vision d'une scène réelle et de la même scène reproduite en virtuel, avec chacune des 6 configurations.

3.4.1 Mise en place

Après avoir lu et signé le formulaire d'informations et de consentement, le participant est invité à remplir deux questionnaires, l'un ayant pour but la collecte d'informations démographiques (âge, genre, port des lentilles correctrices, familiarité avec les jeux vidéos), l'autre l'évaluation de sa propension à l'immersion. Le déroulement et la durée de l'expérimentation lui sont présentés.

À l'intérieur de la voûte, les lunettes sont installées sur le participant selon la procédure détaillée en annexe B et tel que nous pouvons le voir en figure 3.16 (gauche), après quoi il est invité à ne plus y toucher durant le reste de l'expérimentation.

3.4.2 Calibration

Le programme développé pour l'expérimentation comprend une première partie servant à accomplir notre calibration, présentée en figure 3.16 (droite). Une grille virtuelle 2D de 25

points blancs (des sphères de 2 cm de diamètre), d'angles d'ouverture 28° en vertical et 40° en horizontal, s'affiche à 2 mètres devant l'observateur. La position de cette grille et les angles d'ouverture sont ajustables afin de couvrir au maximum le champ de vision de ce dernier, tout en s'assurant que les ETG détectent bien ses yeux alors qu'il observe les quatre coins. La calibration s'effectue sur une sous-grille de 9 points qui s'affichent successivement dans un ordre aléatoire. Au cours de celle-ci, le participant ne tourne pas la tête, mais déplace uniquement ses yeux. Il indique oralement à l'expérimentateur le moment où il observe le point de calibration affiché : les vecteurs, l'un mesuré et l'autre calculé, sont enregistrés et le point suivant apparaît. Finalement, l'affichage de la grille complète et du point de regard obtenu après calibration permet de s'assurer de la qualité de celle-ci avant de continuer, ou éventuellement de la recommencer.

Le participant est ensuite invité, toujours sans bouger la tête, à fixer chacun des 25 points de la grille durant 2 s sans cligner des yeux, afin d'effectuer les mesures d'imprécision et d'instabilité de la calibration réalisée. La mesure d'imprécision correspond à la moyenne sur ces 2 s de la distance entre le POR calculé et la position réelle du point, celle d'instabilité à la moyenne de la distance entre deux POR successifs.

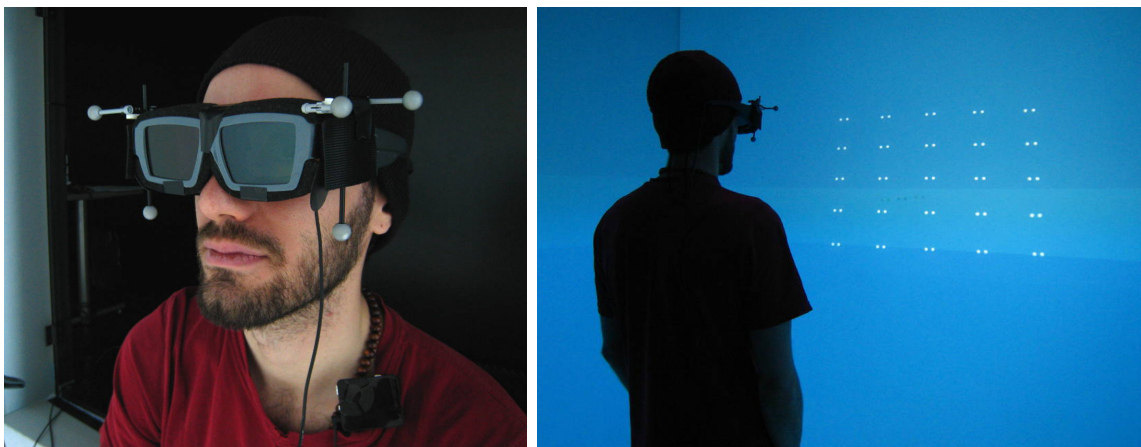


Figure 3.16 Les lunettes installées sur le participant (à gauche). Scène pour la calibration (à droite).

3.4.3 Choix des distances inter-caméras

L'étape suivante consiste à choisir les valeurs des distances inter-caméras pour les configurations **DIC1** et **DIC3** (respectivement **FLOUDIC1** et **FLOUDIC3**). Pour cela, les deux objets visibles sur la figure 3.17 sont placés devant lui à différentes distances : 0,4 m pour DIC3Proche, 1 m pour DIC1 et DIC3Milieu, ou enfin 2,5 m pour DIC3Loin. Il peut

alors à l'aide d'une souris modifier la DIC, c'est-à-dire rapprocher les caméras jusqu'à atteindre la configuration monoscopique ou les écarter sans limites. Nous avons noté que le rapprochement des caméras, en diminuant les disparités horizontales entre les deux images, donne l'impression que les objets grossissent, s'éloignent et s'aplatissent. À l'inverse, éloigner les caméras donne l'impression qu'ils rapetissent, se rapprochent et deviennent plus bombés. Certains participants ont cependant indiqué ne pas distinguer de différence. Il est spécifié au

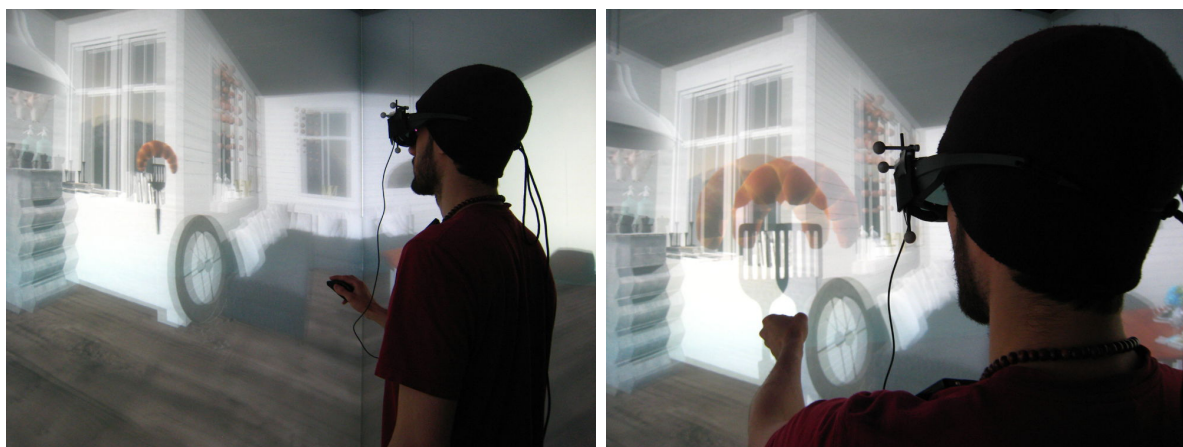


Figure 3.17 Scène pour le choix des DIC.

participant de se baser sur sa perception propre de la profondeur et sur son confort visuel afin d'effectuer ces choix personnels. Concernant la première, il peut s'appuyer à la fois sur ses notions de la taille « normale » d'un croissant et d'une spatule, sur l'indication orale et à l'aide d'un mètre ruban de leur distance, ainsi que sur les indices de profondeur générale donnés par le reste des objets de la scène. Concernant le confort, une disparité moins importante s'avère plus confortable (le cas extrême étant atteint lorsque les caméras se rejoignent et que les deux images sont identiques). Le participant peut donc volontairement réduire l'écartement des caméras afin d'améliorer son confort visuel, au prix de la diminution du réalisme des profondeurs virtuelles.

Chaque choix est réalisé trois fois, la moyenne des trois valeurs enregistrées étant utilisée comme DIC finale. Pour les configurations **DIC3** et **FLOUDIC3**, elle sera interpolée linéairement à chaque image en fonction de la distance entre le centre des yeux et le POR déterminé par le programme. Afin d'éviter, en fonction des pentes des droites d'interpolation, d'atteindre une distance inter-caméras négative (qui inverserait les images gauche et droite) ou trop importante, la DIC est constante lorsque la distance au POR est inférieure à 0,40 m ou supérieure à 2,5 m.

3.4.4 Navigation

Cette partie du test a pour but d'évaluer subjectivement les configurations selon les 5 critères suivants : confort visuel, réalisme du rendu, plaisir de navigation, perception des distances et des profondeurs, sentiment d'immersion.

Scène et chemin utilisés

La scène utilisée pour la navigation, présentée en figure 3.18 et de façon plus détaillée en annexe B.1, est un appartement virtuel dont le modèle a été acheté à Evermotion par l'Institut Philippe-Pinel (EverMotion, 2014) puis modifié. La navigation d'une durée de 90 s, dont le tracé est approximativement représenté en annexe B.1, a été définie de manière à offrir au participant un large intervalle de profondeur pour les objets d'intérêt. Par exemple, au passage des portes ou près des tables, il peut s'attarder sur des éléments très proches, tandis que l'entrée dans le salon ouvre son champ de vision sur des objets plus lointains. Les mouvements de rotation et de translation s'effectuent volontairement plus lentement que ce ne serait le cas avec une marche normale, de manière à diminuer les risques de nausée lors du parcours (voir les risques encourus par les participants en section 3.5.1).

Le choix d'un chemin prédéfini, plutôt que de laisser les sujets parcourir librement l'appartement, est venu de la volonté de diminuer les variabilités inter et intra participants. Par exemple, pour cette première, certains se seraient peut-être limités à l'exploration du salon tandis que d'autres auraient parcouru l'ensemble de l'appartement, les notes de plaisir s'en trouvant alors incomparables. Pour la seconde, un participant aurait pu se lasser de naviguer dans l'appartement et se déplacer de moins en moins au cours des navigations successives, diminuant ainsi certainement son plaisir et son sentiment immersion.



Figure 3.18 Détail de la scène d'appartement utilisée au cours de l'expérimentation.

Notations subjectives

À la fin de chaque navigation, le participant est invité, sans enlever les lunettes, à remplir un questionnaire d'évaluation subjective. Ce temps lui permet également de reposer ses yeux autant qu'il le souhaite.

3.4.5 Limites de fusion

Cette partie de l'expérimentation a été conçue afin de déterminer les limites de fusion pour chacune des configurations, et obtenir ainsi une indication objective de leur efficacité quant à la fusion d'objets virtuels proches. Les deux objets utilisés précédemment sont affichés dans la scène d'appartement selon deux cas de figure :

- Vers l'avant : dans ce sens, ils sont initialement positionnés suffisamment près du participant pour s'assurer qu'il ne puisse pas fusionner leurs représentations stéréoscopiques (7 cm du centre des yeux). Il doit alors les éloigner progressivement jusqu'à gain de fusion.
- Vers l'arrière : dans ce sens, les objets sont placés à une distance garantissant la fusion des images par le participant (40 cm). Il doit cette fois-ci les rapprocher jusqu'à expérimenter de la diplopie.

Comme pour le choix de la DIC, une souris lui permet de déplacer les objets en profondeur. Afin d'obtenir des mesures plus significatives, chaque test est réalisé 3 fois pour chaque configuration et chaque sens, soit 36 itérations pour un participant.

3.4.6 Comparaison entre réel et virtuel

En complément de son évaluation subjective du réalisme du rendu des distances, le comportement oculaire du participant, et en particulier les mouvements de vergence, donne des indications objectives sur sa perception de la 3D. Ces mouvements accompagnent naturellement l'inspection d'une scène réelle lorsque le point de regard passe d'une profondeur à une autre. Dans le cas virtuel, le calcul des images stéréoscopiques ou leur affichage peuvent conduire à une déformation de l'espace, induisant une surévaluation ou une sous-évaluation de la distance qui se répercutera sur l'angle de vergence. Si au contraire les informations de profondeur sont préservées, les mouvements de convergence binoculaires devraient reproduire ceux effectués dans le cas réel, si tant est que le POR suive le même parcours. Ainsi, cette étape de l'expérimentation avait pour but de comparer objectivement les différentes configurations au regard de la configuration réelle, par l'analyse de la réponse oculaire du participant à un stimulus « identique ».

Pour assurer ce dernier point, il est nécessaire de guider le regard du participant. Nous utilisons un robot, prêté par Laurent TREMBLAY et Jérôme COLLIN de l'École Polytechnique de Montréal, capable de suivre un parcours tracé au ruban adhésif noir. Il sera surplombé d'une boîte rectangulaire très simple visible en figure 3.19 (haut). Celle-ci facilite à la fois la modélisation virtuelle et la fixation, pour le participant, d'un point précis : une tête d'épingle bleue placée au sommet. Le parcours, qui sera suivi deux fois, est tracé sur une table blanche de dimensions 80 cm x 120 cm x 76,3 cm¹⁴, reproduite elle aussi virtuellement, au même titre que le parcours et que les murs de la voûte eux-mêmes.

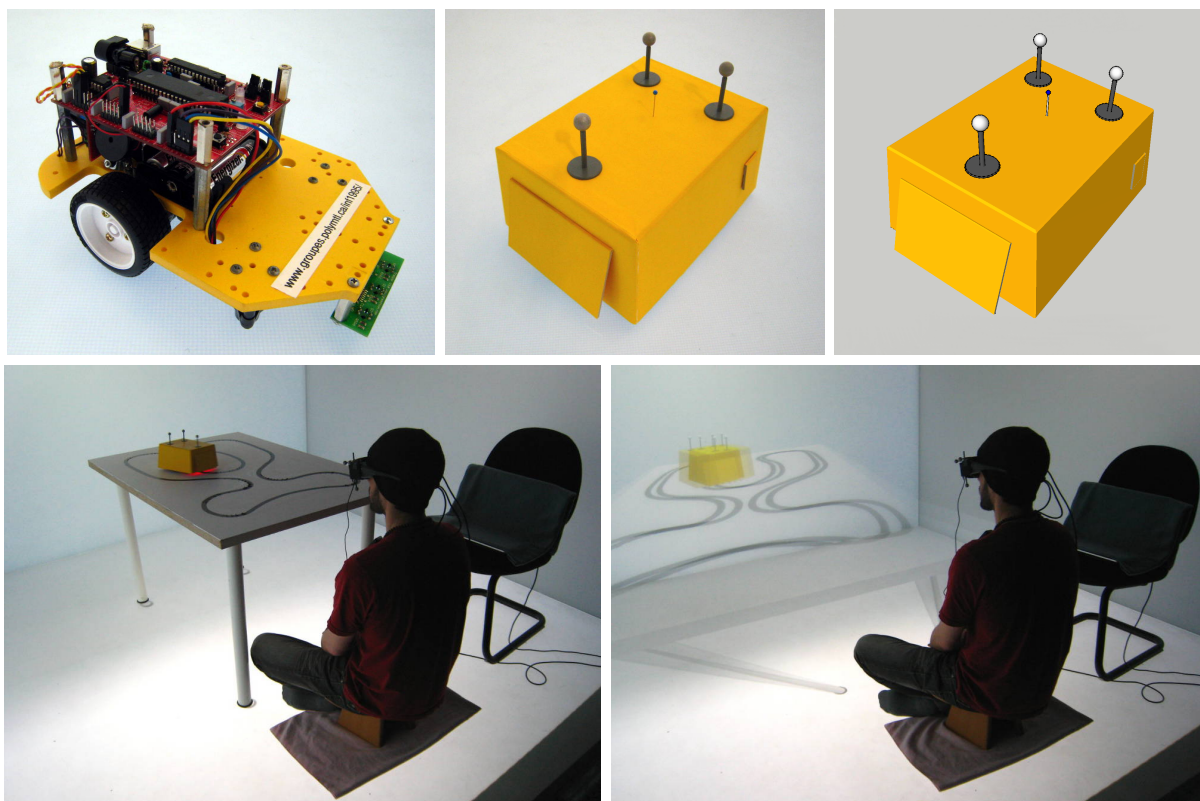


Figure 3.19 Robot réel (en haut à gauche et au milieu) et virtuel (en haut à droite). Installation pour le test réel (en bas à gauche) et virtuel (en bas à droite).

Il est important de noter que les déplacements du robot peuvent varier, du fait de sa recherche tâtonnante en temps réel du tracé à l'aide de détecteurs, mais également de l'usure des piles qui influe sur la vitesse de rotation des moteurs. Afin que les scènes réelle et virtuelle soient les plus similaires possible, nous avons donc décidé d'attribuer au robot un corps rigide, repéré à l'aide des marqueurs placés sur sa boîte. La position et l'orientation de celui-ci sont enregistrées via Unity durant l'étape réelle, puis corrigées et interpolées linéairement avant

14. 31.50 po x 47.24 po x 30.04 po.

d'être appliquées aux 6 robots virtuels successifs. La correction permet d'éliminer d'éventuelles erreurs de mesure, occasionnées par exemple par l'occultation des marqueurs entre eux, causant des retournements verticaux du modèle. L'étape d'interpolation est nécessaire car le temps de calcul d'une image par Unity est légèrement variable, ce qui pourrait provoquer des changements de vitesse du robot. Ainsi, les modèles virtuels ont été réalisés de manière à être les plus ressemblants possible, et leur position dans Unity, qui respecte le système métrique¹⁵, coïncident avec les positions de leurs homologues réels.

Le participant est assis au cours de cette étape, comme nous pouvons le voir en figure 3.19 (bas), afin de réduire la distance minimale entre le centre de ses yeux et le robot en rapprochant ce dernier du plan horizontal contenant les yeux. Ainsi, le robot évoluant approximativement entre 30 cm et 1 m 20, nous obtenons un intervalle plus important de valeurs pour l'angle de vergence.

3.4.7 Considérations sur les ordres

Notre expérimentation fait intervenir plusieurs ordres : celui d'affichage des points lors de la calibration et de la mesure d'imprécision, celui de positionnement des objets pour les choix de DIC, ceux des configurations pour la navigation, la détermination des limites de fusion et le test du robot. Afin de contrebalancer l'accumulation de fatigue visuelle ou la familiarisation à l'environnement au cours des itérations, les participants ont été répartis, pour chaque étape faisant intervenir un ordre, en plusieurs groupes selon le principe des carrés latins. Chaque groupe compte le même nombre de participants, qui partagent le même ordonnancement.

Pour le cas des configurations, deux groupes distincts visualisent des configurations différentes à la même itération, chaque configuration étant visualisée le même nombre de fois aux différentes places possibles. En ce qui concerne l'étape de calibration, 3 séquences de points ont été choisies, parmi les nombreuses possibilités, de manière à maximiser la distance dans l'espace entre deux points successifs.

3.5 Participants

Les 18 volontaires ayant participé aux expérimentations, 3 femmes et 15 hommes, étaient âgés de 22 à 49 ans (moyenne d'âge 28 ans, médiane d'âge 26,5 ans). Ils étaient étudiants de l'École Polytechnique de Montréal, de l'École de technologie supérieure, ou membres du personnel de l'Institut Philippe-Pinel. Parmi eux, 5 portaient des lentilles de contact.

L'ensemble des participants recrutés était familier avec la stéréoscopie, chacun ayant déjà

15. C'est-à-dire qu'un cube dont les arêtes mesurent 1 dans Unity est censé mesurer 1 m s'il est affiché au niveau des écrans de projection.

vu plusieurs films en 3D au cinéma (minimum de 4, maximum de 30, avec une moyenne de 11 films). Bien qu'ils aient tous stipulé voir généralement correctement la 3D à ces occasions, un participant a signalé souffrir parfois de strabisme – ses yeux se désynchronisant en convergence ce qui réduit sa sensation de profondeur, un autre être gêné par les déplacements rapides de l'image pour certains films. Enfin, deux d'entre eux ont remarqué que certaines séquences de quelques films ne leur étaient pas apparues en 3D, l'image semblant plutôt plate, mais ont attribué ce fait à la mauvaise qualité desdits films. Nous n'avons pas testé leur vision correcte de la profondeur en stéréoscopie.

La plupart des participants étaient également coutumiers de la réalité virtuelle de par la pratique de jeux vidéos (seuls 4 ont indiqué ne jamais y jouer, la moyenne pour les autres étant de 8,6 h/semaine). Enfin, 11 d'entre eux avaient déjà été immergés au sein d'une voûte de réalité virtuelle (de quelques minutes à plusieurs années de travail avec ce type d'environnement).

3.5.1 Risques encourus par les participants

La participation de sujets humains imposait de nous pencher sur les risques encourus au cours de l'expérimentation. En effet, l'immersion dans une voûte de réalité virtuelle présente des risques réversibles d'inconfort, voire de fatigue visuelle, pouvant aller de la simple sensation de pression sur les yeux jusqu'à la nausée. Cependant, ce risque n'excède pas celui encouru lors du visionnement d'un film en 3D et d'une visite chez l'optométriste – le test de la détermination des limites de fusion se rapprochant de ceux passés lors d'une telle visite. C'est pourquoi seuls des individus ayant déjà visionné au moins un film en 3D et n'ayant pas ressenti de gêne importante (nausée) à cette occasion furent recrutés. Par ailleurs, certaines mesures ont été prévues de manière à limiter au maximum ces risques : les déplacements lents lors de la navigation réduisent les facteurs de fatigue visuelle (en particulier l'amplitude et la fréquence des variations de profondeur du point de regard), et des pauses jalonnant les étapes permettent au participant de reposer ses yeux. Enfin, nous signalons à celui-ci au début de l'expérimentation qu'il peut y mettre fin à tout moment par simple avis verbal. Notre protocole a reçu l'aval des comités d'éthique de l'Institut Philippe-Pinel, de l'École Polytechnique de Montréal et de l'École de technologie supérieure.

CHAPITRE 4

RÉSULTATS

Les données collectées au cours de l'expérimentation peuvent être regroupées en trois catégories distinctes. Tout d'abord, les participants ont dû effectuer des choix personnels pour certains paramètres, dont les valeurs influencent directement l'efficacité des configurations. Nous présentons et discutons ici ces choix. Les configurations ont ensuite été comparées subjectivement, grâce aux notations octroyées lors de la navigation, puis objectivement en termes de capacité de fusion et de similarité avec la configuration réelle. L'analyse des résultats obtenus nous permettra de tirer les conclusions quant à nos hypothèses dans le chapitre suivant de ce mémoire.

4.1 Remarques préliminaires sur les analyses

Afin d'éclaircir l'interprétation des données, voici quelques détails sur les représentations graphiques employées et les analyses statistiques effectuées. Tout d'abord, nous utilisons des boîtes à moustache pour représenter certains résultats. Les lignes colorées indiquent le minimum et le maximum, la boîte en elle-même étant délimitée par les premier et troisième quartiles. La ligne noire dans la boîte situe enfin la médiane.

Les tests statistiques employés, ANOVA à 1 ou 2 facteurs contrôlés et de Student apparié, s'appuient sur la variabilité des données pour déterminer la probabilité avec laquelle plusieurs échantillons seraient issus de populations différentes. Le test de Student apparié traite les différences entre deux groupes (par exemple le groupe des hommes et celui des femmes), tandis que le test ANOVA peut s'appliquer à un plus grand nombre d'échantillons (par exemple les 6 configurations). Nous avons en particulier utilisé ce dernier pour étudier l'impact de la configuration (1 facteur) ou des deux paramètres : la distance inter-caméras et la présence de flou (2 facteurs). Lorsque deux facteurs sont en jeux, le test ANOVA fournit également une probabilité d'indépendance nous renseignant sur l'additivité de leurs effets (valeur haute de probabilité) ou au contraire sur leur interaction (valeur basse). Nous définissons la valeur significative de probabilité à $p \leq 0,05$ mais parlons également parfois de résultats marginalement significatifs lorsque $p \leq 0,2$. Ainsi, en dessous de ces seuils, les populations (en l'occurrence les configurations ou les valeurs de paramètres) apparaissent significativement ou marginalement significativement différentes au regard du critère de comparaison courant.

Nous avons également étudié le lien entre plusieurs couples de variables, par exemple

la précision de la détermination du point de regard et la moyenne des notes attribuées. Graphiquement, nous représentons l'une des variables par rapport à l'autre et traçons la régression linéaire calculée. Les mesures utilisées pour les analyses sont alors le coefficient de détermination R^2 de cette régression, ainsi que le coefficient de corrélation R entre les deux variables et la probabilité p qu'elles soient indépendantes. Le premier indique le pourcentage de variation dans les données qui est expliqué par le modèle utilisé, ici des fonctions affines. Plus ce coefficient est proche de 1, plus le modèle « explique » les données. En ce qui concerne la corrélation entre les deux variables, les seuils employés pour la probabilité p sont les mêmes que dans le cas des tests ANOVA, $p \leq 0,05$ et $p \leq 0,2$ indiquant respectivement que les deux variables sont significativement ou marginalement significativement corrélées. Pour plus de lisibilité, les valeurs de probabilités significatives et marginalement significatives sont indiquées en gras dans les tableaux.

4.2 Choix des valeurs de paramètres

Cette section présente les choix des participants quant à la distance inter-caméras pour les configurations qui n'utilisent pas la distance interpupillaire anatomique. DIC1 correspond au choix unique et fixe employée par les configurations **DIC1** et **FLOUDIC1**, DIC3 Proche, Milieu et Loin aux 3 choix réalisés pour 3 distances différentes des objets guides qui serviront à interpoler la DIC pour les configurations **DIC3** et **FLOUDIC3**. La figure 4.1 présente les



Figure 4.1 DIP et choix de DIC (en cm).

distances interpupillaires anatomiques de l'ensemble des participants, ainsi que leurs choix de distances inter-caméras. Les premières sont comprises entre 5,51 cm et 7,17 cm, avec une moyenne de 6,28 cm et un écart-type de 0,37 cm. Les DIC choisies sont beaucoup plus espacées, de 2,38 cm à 17,11 cm, avec une moyenne de 8,17 cm et un écart-type de 2,85 cm. Nous remarquons que les valeurs choisies pour DIC1 et DIC3 Milieu sont très proches,

ce à quoi nous nous attendions étant donné que les objets sont placés à la même distance du participant pour ces choix. Nous notons également qu'à 0,4 m la DIC choisie est proche de la DIP anatomique, tandis que pour les autres distances elle est plus élevée de 2,57 cm en moyenne, ayant pour effet d'exagérer l'impression de 3D puisque les disparités sont plus importantes.

Les participants devant effectuer chacun des choix trois fois, nous souhaitions savoir si l'itération du choix avait une influence sur celui-ci, le dernier pouvant par exemple se révéler systématiquement plus petit pour être plus confortable. Cependant, la figure 4.2 atteste que les valeurs ne semblent globalement pas varier au cours des 3 itérations, bien que nous observions une légère décroissance du choix médian de DIC3 Loin.

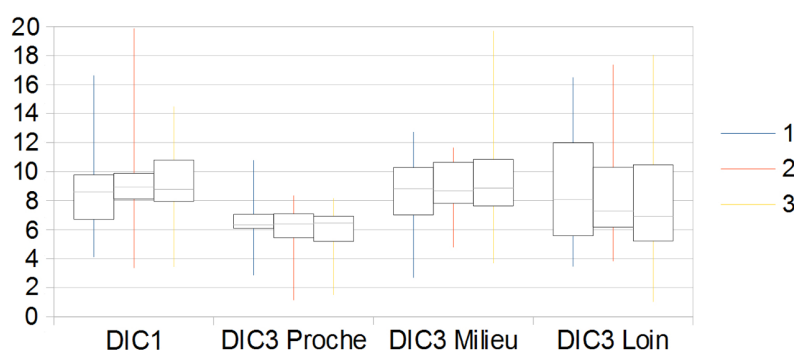


Figure 4.2 Les 3 itérations de choix pour chacune des 4 DIC (en cm).

La figure 4.3 présente enfin les DIC choisies par chaque participant pour la configuration **DIC3**. La DIC croît en moyenne de 2,90 cm quand les objets passent de 0,4 m à 1 m, tandis qu'elle chute ensuite de 0,60 cm. En effet, pour 13 participants, le choix à 2,5 m est plus petit qu'à 1 m. Ce résultat peut s'expliquer par le fait qu'une majorité de participants a signalé avoir des difficultés à effectuer un choix pour la distance 2,5 m, en particulier en raison de l'impossibilité d'utiliser le mètre ruban pour leur donner des indices sur cette distance, et de la petite taille des objets sur l'écran. Ils se sont alors principalement basés sur leur confort à la vue des objets « distants » de la scène, une DIC plus petite étant plus confortable pour les yeux.

4.3 Notations subjectives

L'étape suivante du protocole expérimental consistait en une navigation de 90 s à travers un appartement. Les participants durent alors noter les 6 configurations selon les critères de confort visuel, réalisme du rendu, plaisir de navigation, perception des distances et des

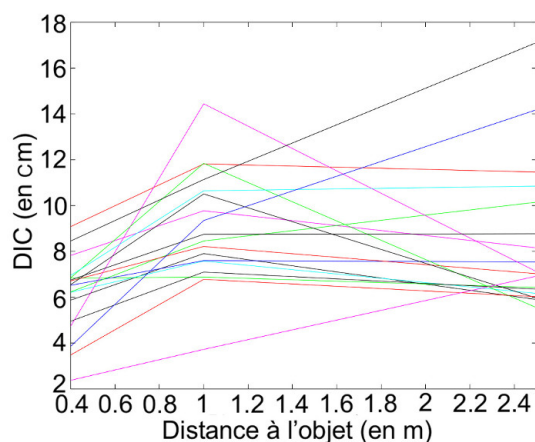


Figure 4.3 Choix de DIC en fonction de la distance de l'objet étalon.

profondeurs, et enfin sentiment de présence, d'immersion dans le monde virtuel. Les indications répertoriées dans le tableau D.1 en annexe les aiguillaient sur les éléments à prendre en considération. L'échelle de notation, continue mais fractionnée en 7 parties, s'étalait des appréciations « très négative » à « très positive ». Les participants étaient libres de placer leur marque à n'importe quelle position sur l'échelle, celle-ci étant par la suite mesurée puis convertie vers une échelle entre 0 et 1. La figure 4.4 réunit les résultats obtenus par chacune

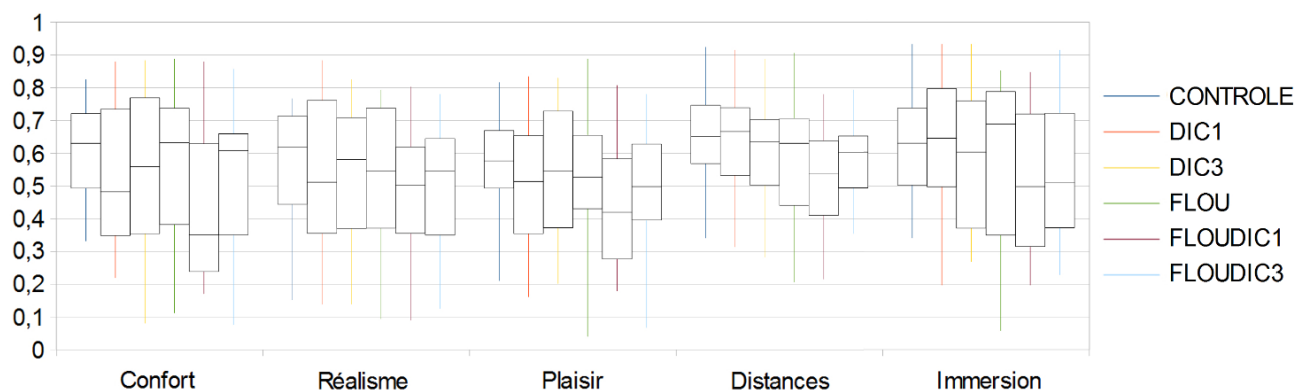


Figure 4.4 Notes obtenues par chacune des configurations dans chacun des critères.

des configurations dans chacun des critères de notation. De prime abord, il apparaît qu'aucune configuration ne se démarque positivement par rapport aux autres dans tous les critères de notation. **FLOU** présente la meilleure médiane pour le confort et l'immersion (mais dans ce dernier critère, c'est également **FLOU** qui a reçu la plus mauvaise note), **CONTROLE** pour le réalisme et le plaisir, et **DIC1** pour ce qui est de la perception des distances. Nous remarquons cependant qu'une configuration se démarque négativement : **FLOUDIC1** a les

	CONTROLE	DIC1	DIC3	FLOU	FLOUDIC1	FLOUDIC3
Note moyenne	0,591	0,572	0,560	0,547	0,481	0,532
Participants	30	26	15	14	7	13

Tableau 4.1 Moyennes et dénombrement des participants ayant octroyé les meilleures notes tous critères confondus.

médianes les plus basses dans chacun des critères. En ce qui concerne le paramètre de DIC, nous observons que l'ordre est le même dans 8 cas sur 10 : $DIP \geq DIC3 \geq DIC1$. Nous pouvons donc conclure que la distance inter-caméras préférée est la distance interpupillaire anatomique. Il n'y a que dans les critères de distances et d'immersion que l'ordre devient $DIC1 \geq DIP \geq DIC3$ pour les configurations sans flou, ce qui peut s'expliquer pour le cas des distances par le fait que les participants ont choisi leur DIC1 en se basant précisément sur leur perception de la profondeur, pour un objet situé à une distance proche de la distance moyenne de tous les objets au cours de la navigation. Concernant le paramètre de flou, nous remarquons que la configuration nette est préférée à son homologue dans 12 cas sur 15, ce qui laisse supposer que l'ajout de flou est défavorable à la configuration. Le tableau 4.1 (complémenté par le tableau D.2 en annexe qui fournit le détail par critère) présente quant à lui les moyennes globales et dénombre pour chaque configuration les participants l'ayant la mieux notée. Il atteste qu'en termes de nombre de participants et de moyenne c'est également la configuration **CONTROLE** qui a le plus de succès, confirmant que la distance inter-caméras préférée est la distance interpupillaire anatomique et que la présence d'un flou de profondeur de champ n'est pas appréciable.

Cependant, les analyses complémentaires révèlent que ces préférences ne sont pas significatives pour la plupart des critères. En effet, le test ANOVA à 1 facteur contrôlé réalisé montre que la configuration n'a pas d'impact significatif sur les notes attribuées à la navigation, comme l'indiquent les valeurs de probabilité regroupées dans le tableau 4.2. Afin de préciser l'étude, nous avons également exploré l'effet de chacun des deux paramètres, choix de la DIC et présence de flou, ainsi que leur interaction. Les résultats du test ANOVA à 2 facteurs contrôlés montrent que le flou a un impact significatif sur les notes dans le critère de perception des distances et des profondeurs, et marginalement significatif sur le sentiment d'immersion. En observant la figure 4.4, nous remarquons que cet impact s'avère négatif, la présence de flou diminuant la médiane des notes dans ces deux critères.

Afin d'affiner les résultats, nous avons étudié l'impact de plusieurs facteurs éventuels, tels que l'accroissement de la fatigue ou de la familiarisation avec l'environnement au cours des navigations successives, l'hésitation lors du choix des DIC ou encore la qualité du calcul du POR. Ces analyses sont discutées dans les sections suivantes.

Critère	Configuration	DIC	Flou	Interaction
Confort	F=0,86 p=0,513	F=1,40 p=0,251	F=1,04 p=0,310	F=0,22 p=0,800
Réalisme	F=0,37 p=0,870	F=0,22 p=0,802	F=1,23 p=0,270	F=0,08 p=0,921
Plaisir	F=0,62 p=0,687	F=0,59 p=0,555	F=1,49 p=0,226	F=0,21 p=0,814
Distances	F=1,27 p=0,283	F=0,22 p=0,802	F=4,91 p= 0,034	F=0,64 p=0,528
Immersion	F=0,72 p=0,613	F=0,07 p=0,932	F=2,59 p= 0,111	F=0,43 p=0,653

Tableau 4.2 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives.

4.3.1 Ordre de navigation

Plusieurs participants ont indiqué que leur plaisir de navigation augmentait au cours des navigations car ils remarquaient de plus en plus d'objets dans la scène. D'autres ont reporté un accroissement de la fatigue visuelle, et par conséquent une possible diminution de leur confort visuel en fonction de l'itération de la navigation. Le test ANOVA à 1 facteur contrôlé effectué, dont les résultats sont présentés dans le tableau 4.3, révèle que l'ordre d'apparition de la configuration dans la suite de navigations n'a pas d'impact significatif sur les notes. Cependant, nous pouvons noter sur la figure 4.5 que la note de plaisir augmente au cours des itérations, tel que souligné par certains participants.

Confort	Réalisme	Plaisir	Distances	Immersion
F=0,33 p=0,893	F=0,41 p=0,841	F=0,85 p=0,521	F=0,30 p=0,913	F=0,13 p=0,985

Tableau 4.3 ANOVA à 1 facteur contrôlé : effet de l'ordre de navigation sur les notations subjectives.

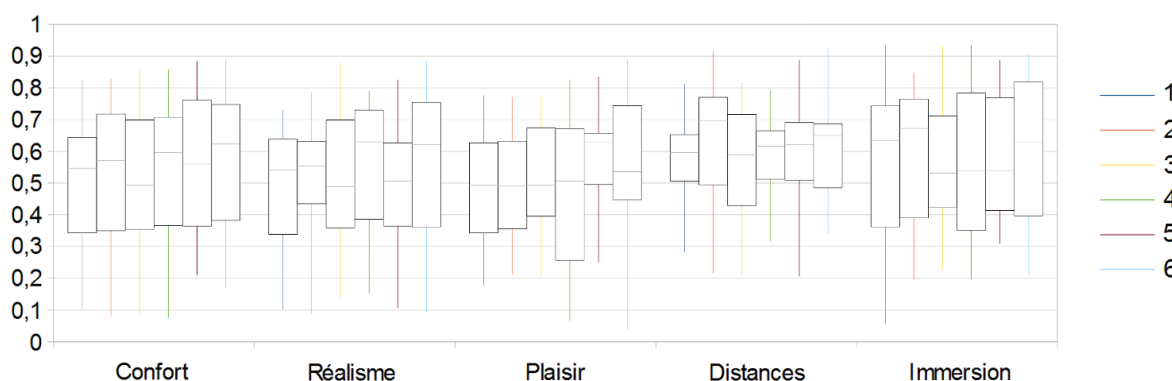


Figure 4.5 Notes obtenues par ordre de navigation dans chacun des critères.

4.3.2 Étude approfondie basée sur la DIC1

Choix de DIC

La DIC ayant un impact direct sur la génération des images stéréoscopiques gauche et droite, et en particulier sur les disparités horizontales présentes entre celles-ci, nous nous sommes intéressés au lien entre le choix de DIC1 et les notes attribuées aux configurations **DIC1** et **FLOUDIC1**. Le tableau 4.4 réunit les probabilités d'indépendance entre ces deux variables dans les différents critères, et indique la croissance ou la décroissance des régressions linéaires associées. Les graphiques illustrant ces dernières peuvent être consultés en annexe (figures D.1, D.2 et D.3). À noter qu'afin de dégager de manière plus significative certaines tendances, nous avons dans un second temps écarté les participants ayant effectué les choix extrêmes (ici 3,64 cm et 17 cm), qui semblaient diverger des autres en termes de notations. Les paragraphes ci-dessous présentent les observations ressortant de ce tableau.

Concernant **DIC1** tout d'abord, nous remarquons dans le cas sans flou qu'une DIC plus importante a un impact significativement négatif pour le critère de perception des distances, et marginalement significatif pour les critères de confort et de plaisir. En écartant les extrêmes, ce résultat devient même significatif pour les trois premiers critères, mais n'est plus que marginalement significatif concernant les distances. La corrélation avec le confort, critère qui peut être fortement lié à celui de plaisir de navigation, s'explique selon nous par le postulat ressortant de la revue de littérature voulant qu'une DIC plus petite soit plus confortable. L'impact sur la perception des distances et le réalisme nous semble cependant plus étonnant étant donné qu'une petite DIC diminue l'impression de 3D. En ce qui concerne le dernier critère, le sentiment d'immersion, aucune tendance ne semble se dégager.

En revanche, la présence de flou semble diminuer l'influence de la DIC car les valeurs de probabilité obtenues sont plus faibles, les tendances restant cependant les mêmes avec des courbes décroissantes en fonction de la DIC. Nous remarquons en particulier que de manière étonnante la DIC ne semble plus avoir d'impact sur la perception des distances et des profondeurs en présence de flou, bien qu'aucune corrélation ne soit ressortie entre ces deux paramètres dans le tableau 4.2. Cependant, ces analyses ne prennent pas en compte la DIP anatomique des participants, dont l'éloignement avec la DIC choisie peut également influencer sur les notations tel que nous le verrons dans la section suivante.

Ratio DIP/DIC

Nous avons vu en figure 4.1 que les DIC1 choisies par les participants étaient la plupart du temps supérieures à leur DIP anatomique, conduisant à des ratios inférieurs à 1 dans 16 cas sur 18. Le tableau 4.5 regroupe les valeurs de probabilité d'indépendance entre ce ratio

	Confort	Réalisme	Plaisir	Distances	Immersion
DIC1	p= 0,184 ↘	p=0,381 ↘	p= 0,178 ↘	p= 0,029 ↘	p=0,370 ↗
DIC1 sans extrêmes	p= 0,007 ↘	p= 0,011 ↘	p= 0,015 ↘	p= 0,055 ↘	p=0,717 ↗
FLOUDIC1	p=0,869 ↗	p=0,658 ↘	p=0,344 ↘	p=0,758 ↗	p=0,372 ↗
FLOUDIC1 sans extrêmes	p= 0,042 ↘	p= 0,105 ↘	p= 0,043 ↘	p=0,458 ↘	p=0,492 ↘
Moyenne DIC1 et FLOUDIC1	p=0,587 ↘	p=0,484 ↘	p=0,210 ↘	p=0,289 ↘	p=0,320 ↗
Moyenne DIC1 et FLOUDIC1 sans extrêmes	p= 0,014 ↘	p= 0,029 ↘	p= 0,016 ↘	p= 0,118 ↘	p=0,838 ↘

Tableau 4.4 Probabilités d'indépendance entre la DIC1 choisie et les notes subjectives attribuées à **DIC1** et **FLOUDIC1**, et variations des régressions linéaires associées.

et les notes attribuées aux configurations **DIC1** et **FLOUDIC1**, et indique les variations des régressions linéaires associées (pour les graphiques détaillés, voir les figures D.4, D.5 et D.6 en annexe). À nouveau, nous écartons dans un second temps un individu en raison de différences de notation par rapport aux autres. Son ratio de 1,51 était le plus important, avec un choix de DIC1 de 3,64 cm contre une DIP de 5,51 cm.

Pour la configuration **DIC1**, nous observons l'émergence de certaines tendances, avec l'attribution de meilleures notes dans les quatre premiers critères lorsque la DIC se rapproche de la DIP anatomique. En écartant la valeur extrême, l'impact du ratio DIP/DIC devient significatif en termes de confort, plaisir et perception des distances, et marginalement significatif pour le critère de réalisme. Ces résultats tendent à indiquer qu'une DIC proche de la DIP physiologique est préférée par les participants : c'est donc bien cette dernière qui est optimale.

Une fois de plus, bien que les tendances restent similaires pour la configuration **FLOUDIC1**, en particulier pour les trois premiers critères, la probabilité d'indépendance s'avère beaucoup plus faible (ainsi que les coefficients de détermination des régressions linéaires dont les valeurs sont indiquées sur les figures D.4, D.5 et D.6 en annexe). Le flou semble donc influencer sur l'effet du ratio DIP/DIC, qui reste cependant significatif en considérant la moyenne des deux configurations dans les critères de confort et de perception des distances.

Hésitation dans le choix de la DIC

Nous avons envisagé que l'hésitation au cours du choix de la DIC1, représentée par l'écart-type entre les 3 choix effectués successivement, pouvait également influencer sur les notes attri-

	Confort	Réalisme	Plaisir	Distances	Immersion
DIC1	p=0,248 ↗	p=0,331 ↗	p=0,422 ↗	p= 0,006 ↗	p= 0,123 ↘
DIC1 sans extrêmes	p= 0,002 ↗	p= 0,087 ↗	p= 0,012 ↗	p= 0,040 ↗	p=0,960 ↗
FLOUDIC1	p=0,926 ↗	p=0,803 ↗	p=0,216 ↗	p=0,695 ↗	p=0,533 ↘
FLOUDIC1 sans extrêmes	p=0,240 ↗	p=0,327 ↗	p= 0,112 ↗	p=0,911 ↘	p=0,904 ↘
Moyenne DIC1 et FLOUDIC1	p=0,541 ↗	p=0,513 ↗	p=0,260 ↗	p= 0,071 ↗	p=0,238 ↘
Moyenne DIC1 et FLOUDIC1 sans extrêmes	p= 0,033 ↗	p= 0,155 ↗	p= 0,027 ↗	p=0,285 ↗	p=0,962 ↘

Tableau 4.5 Probabilités d'indépendance entre le ratio DIP/DIC1 et les notes subjectives attribuées à **DIC1** et **FLOUDIC1**, et variations des régressions linéaires associées.

buées à ces configurations. En effet, si le participant est sûr de son choix, c'est qu'il est sans doute conscient que cette DIC est celle qui est la plus confortable et la plus réaliste pour lui. Les données indiquent que certains individus ont effectué des choix successifs assez éloignés : 5 écarts-types sont supérieurs à 1,5 cm, avec un maximum à 3,61 cm contre une moyenne de 0,96 cm et une médiane de 0,44 cm. Les probabilités d'indépendance entre écart-type et note, ainsi que les variations des régressions linéaires associées, sont regroupées dans le tableau 4.6 (complémenté par les figures D.7, D.8 et D.9 en annexe).

En réduisant le champ aux participants dont l'écart-type est inférieur à 1,5 cm (ce qui correspond aux lignes « sans extrêmes » du tableau 4.6), nous observons pour la configuration **DIC1** une légère tendance dans le sens de notre hypothèse, avec des régressions linéaires décroissantes et des probabilités marginalement significatives pour les critères de réalisme, plaisir et sentiment d'immersion. Les données attestent pourtant que les meilleures notes n'ont pas été données par les participants les plus sûrs de leur choix.

Dans le cas **FLOUDIC1**, nous observons le phénomène inverse lorsque tous les participants sont pris en compte, avec des régressions linéaires croissantes. L'hésitation dans le choix semble en effet avoir un impact positif marginalement significatif dans les trois premiers critères. Cependant, ces tendances disparaissent sans extrêmes, c'est-à-dire en écartant les 5 participants dont l'écart-type est le plus élevé. Nous supposons alors qu'elles pourraient éventuellement être expliquées par un autre facteur que l'hésitation dans le choix de la DIC1, qui regrouperait les 5 participants en question et expliquerait leur tendance à mettre de meilleures notes. Néanmoins, sans tenir compte de ces participants, un écart-type élevé semble une fois de plus conduire de façon marginalement significative à de meilleures notes dans le critère de perception des distances.

En considérant enfin les deux configurations, l'effet marginalement significatif de l'hésitation qui ressort pour le critère de réalisme s'inverse en considérant tous les participants (effet positif) ou en écartant les 5 extrêmes (effet négatif). De manière générale, il apparaît qu'il est difficile de conclure quant à une relation entre l'hésitation dans le choix de la DIC et les notes attribuées aux configurations **DIC1** et **FLOUDIC1**.

	Confort	Réalisme	Plaisir	Distances	Immersion
DIC1	p=0,870 ↗	p=0,322 ↗	p=0,756 ↘	p=0,350 ↘	p=0,610 ↘
DIC1 sans extrêmes	p=0,461 ↘	p= 0,148 ↘	p= 0,178 ↘	p=0,711 ↗	p= 0,163 ↘
FLOUDIC1	p= 0,183 ↗	p= 0,063 ↗	p= 0,093 ↗	p=0,349 ↗	p=0,237 ↘
FLOUDIC1 sans extrêmes	p=0,881 ↘	p=0,214 ↘	p=0,925 ↘	p= 0,092 ↗	p=0,810 ↘
Moyenne DIC1 et FLOUDIC1	p=0,410 ↗	p= 0,145 ↗	p=0,457 ↗	p=0,992 ↘	p=0,701 ↗
Moyenne DIC1 et FLOUDIC1 sans extrêmes	p=0,667 ↘	p= 0,160 ↘	p=0,451 ↘	p=0,252 ↗	p=0,356 ↘

Tableau 4.6 Probabilités d'indépendance entre l'écart-type des 3 DIC1 choisies et les notes subjectives attribuées à **DIC1** et **FLOUDIC1**, et variations des régressions linéaires associées.

4.3.3 Étude approfondie basée sur les DIC3

Hésitation dans les choix de la DIC

De la même façon que pour DIC1, nous nous sommes demandé si l'hésitation dans le choix des DIC3 avait eu un impact négatif sur les notes des configurations **DIC3** et **FLOUDIC3**. Nous avons utilisé la moyenne des écarts-types pour les 3 DIC choisies comme mesure d'hésitation. Les probabilités d'indépendance obtenues ainsi que les variations des régressions linéaires entre moyenne des écarts-types et note sont regroupées dans le tableau 4.7, les graphiques étant consultables en annexe (figures D.10, D.11 et D.12). Aucune conclusion significative ne ressort des résultats dans le cas **DIC3**, hormis le fait qu'un écart-type élevé conduit, en écartant les deux écarts-types les plus importants, de manière marginalement significative à de moins bonnes notes dans le critère d'immersion. En revanche, pour **FLOUDIC3**, nous notons une tendance à des notes plus élevées lorsque l'écart-type moyen pour les 3 DIC est plus important, contrairement à notre hypothèse. Ce résultat s'avère marginalement significatif dans les critères de réalisme et de plaisir.

	Confort	Réalisme	Plaisir	Distances	Immersion
DIC3	p=0,600 ↘	p=0,444 ↘	p=0,479 ↘	p=0,351 ↘	p=0,258 ↘
DIC3 sans extrêmes	p=0,603 ↗	p=0,905 ↗	p=0,742 ↗	p=0,345 ↘	p= 0,071 ↘
FLOUDIC3	p=0,987 ↗	p=0,709 ↗	p=0,666 ↗	p=0,254 ↗	p=0,741 ↗
FLOUDIC3 sans extrêmes	p=0,252 ↗	p= 0,161 ↗	p= 0,159 ↗	p=0,748 ↗	p=0,884 ↘
Moyenne DIC3 et FLOUDIC3	p=0,769 ↘	p=0,801 ↘	p=0,864 ↘	p=0,971 ↗	p=0,663 ↘
Moyenne DIC3 et FLOUDIC3 sans extrêmes	p=0,350 ↗	p=0,423 ↗	p=0,335 ↗	p=0,629 ↘	p=0,295 ↘

Tableau 4.7 Probabilités d'indépendance entre la moyenne des écarts-types des DIC3 et les notes subjectives attribuées à **DIC3** et **FLOUDIC3**, et variations des régressions linéaires associées.

Variabilité entre les 3 DIC

Nous avons également considéré l'impact de l'écart entre les 3 DIC choisies et les notes obtenues pour cette configuration. En effet, la modification de l'écart inter-caméras par interpolation de ces valeurs lors du changement de profondeur du point de regard peut perturber l'observateur, surtout si elle est réalisée avec un temps de latence perceptible. Si celui-ci a choisi des DIC très différentes, cette modification aura sans doute un impact plus important. Le tableau 4.8 résume les valeurs de probabilité d'indépendance obtenues, ainsi que les variations des régressions linéaires (visibles en annexe sur les figures D.13, D.14 et D.15).

Nous remarquons que les notes subjectives attribuées à la configuration **DIC3** dans les trois premiers critères sont inversement liées à l'écart-type entre les 3 DIC, de façon significative pour le confort et le plaisir, et marginalement significative en ce qui concerne le réalisme. Cette tendance diminue cependant après avoir écarté les 4 écarts-types les plus importants (supérieurs à 3 cm).

Ce résultat se retrouve en considérant la configuration **FLOUDIC3** ainsi que les deux configurations en même temps, la présence de flou ne semblant donc pas influencer de manière importante l'effet des valeurs de **DIC3** sur les notations subjectives. Nous notons cependant pour ces deux cas un impact significatif et marginalement significatif de l'écart-type inter-DIC3 sur les notes d'immersion après avoir écarté les extrêmes.

	Confort	Réalisme	Plaisir	Distances	Immersion
DIC3	p= 0,032 ↘	p= 0,121 ↘	p= 0,027 ↘	p=0,465 ↘	p=0,490 ↘
DIC3 sans extrêmes	p=0,461 ↘	p=0,420 ↘	p=0,269 ↘	p=0,864 ↘	p=0,203 ↘
FLOUDIC3	p= 0,034 ↘	p= 0,175 ↘	p= 0,018 ↘	p=0,915 ↗	p=0,274 ↘
FLOUDIC3 sans extrêmes	p=0,506 ↘	p= 0,141 ↘	p= 0,053 ↘	p=0,638 ↘	p= 0,050 ↘
Moyenne DIC3 et FLOUDIC3	p= 0,015 ↘	p= 0,108 ↘	p= 0,010 ↘	p=0,661 ↘	p=0,332 ↘
Moyenne DIC3 et FLOUDIC3 sans extrêmes	p=0,433 ↘	p=0,201 ↘	p= 0,099 ↘	p=0,649 ↘	p= 0,073 ↘

Tableau 4.8 Probabilité d'indépendance entre l'écart-type inter-DIC3 et les notes subjectives attribuées à **DIC3** et **FLOUDIC3**, et variations des régressions linéaires associées.

4.3.4 Questionnaire de propension à l'immersion

Le questionnaire de propension à l'immersion rempli au début de l'expérimentation fournit une mesure de la capacité d'immersion du participant, et nous permet d'évaluer l'impact de celle-ci sur ses notes subjectives. Les réponses ont été considérées dans leur ensemble, ou réunies selon les critères d'implication, de focus, d'émotion et de jeu. Une note importante dans ces critères révèle respectivement une plus grande capacité d'implication dans une tâche, de concentration sur celle-ci, des émotions plus vives ressenties et une propension au jeu. Outre la relation entre le résultat total à ce test et les notes données dans les différents critères, les associations suivantes ont été testées : implication-immersion, focus-immersion, émotions-immersion, émotions-plaisir, jeu-plaisir. Les probabilités d'indépendance résultantes sont présentées dans le tableau 4.9, ainsi que les variations des régressions linéaires (pour les graphiques détaillés, voir la figure D.16 en annexe). Pour le critère de réalisme, une corrélation se dégage significativement entre le résultat total au test de propension à l'immersion et la note moyenne, celle-ci étant plus élevée pour un score plus élevé. Aucune autre relation ne semble lier les scores de QPI et la hauteur des notes mises par les participants.

4.3.5 Qualité du calcul du POR

Le calcul de la DIC pour les configurations **DIC3** et **FLOUDIC3** utilise la distance entre le participant (plus exactement le centre de ses yeux) et son point de regard. Si ce dernier est mal évalué, il en sera de même pour cette distance, et la DIC utilisée par le système pourra se révéler inconfortable pour le participant étant donné sa distance réelle de regard. Le flou de profondeur de champ est également calculé en utilisant la position du POR, une

Total Confort	Total Réalisme	Total Plaisir	Total Distances	Total Immersion
p=0,332 ↗	p= 0,040 ↗	p=0,912 ↗	p=0,828 ↗	p=0,447 ↗
Implication Immersion	Focus Immersion	Émotions Immersion	Émotions Plaisir	Jeu Plaisir
p=0,565 ↗	p=0,646 ↗	p=0,855 ↗	p=0,437 ↗	p=0,590 ↘

Tableau 4.9 Probabilité d'indépendance entre le score au QPI et les notes subjectives, et variations des régressions linéaires associées.

mauvaise estimation pouvant conduire en particulier à flouter l'objet effectivement regardé. La qualité de l'estimation du point de regard peut par ces biais avoir un impact sur les notations subjectives, hormis pour les configurations **CONTROLE** et **DIC1** qui n'utilisent pas cette information pour modifier en temps réel l'affichage. Plusieurs mesures de la qualité de l'estimation ont été employées, telles que l'imprécision, l'instabilité ou encore certaines mesures d'incertitude détaillées plus bas.

Imprécision du point de regard

La valeur d'imprécision utilisée ici est celle obtenue lors de la procédure de mesure effectuée après la calibration. Elle représente la moyenne sur 2 secondes de la distance entre le POR évalué et le POR réel, une valeur élevée indiquant donc que l'estimation du POR est imprécise. Le tableau 4.10 regroupe les probabilités d'indépendance entre cette mesure et les notes subjectives, ainsi que les variations des régressions linéaires associées et présentées sur les figures D.17 et D.19 en annexe. Nous observons qu'aucune corrélation ne semble exister entre l'imprécision et les notes moyennes, d'autant plus après avoir écarté les 3 participants pour lesquels la précision était la plus mauvaise. Le détail par critère confirme ce résultat. Les données brutes indiquent par exemple qu'en ce qui concerne le confort, le participant ayant l'une des 3 plus hautes mesures d'imprécision (16,9 cm) a donné la meilleure note à la configuration **CONTROLE**, qui n'utilise pas le calcul du POR, et la deuxième meilleure note à celle qui l'utilise le plus : **FLOUDIC3**. L'imprécision obtenue par notre système semble donc être suffisamment petite pour ne pas influencer les notes.

Instabilité du point de regard

La mesure d'instabilité utilisée est également celle obtenue lors de la procédure de mesure réalisée après la calibration. Elle correspond à la moyenne sur 2 secondes de la distance entre deux évaluations de POR, une valeur élevée pour cette distance correspondant donc à une estimation instable de celui-ci. Nous remarquons dans le tableau 4.11, qui présente

	Moyenne	Confort	Réalisme	Plaisir	Distances	Immersion
Imprécision	p=0,361 ↘	p=0,544 ↘	p=0,975 ↗	p=0,422 ↘	p=0,230 ↘	p=0,300 ↘
Imprécision sans extrêmes	p=0,960 ↗	p=0,641 ↗	p=0,729 ↘	p=0,861 ↘	p=0,820 ↗	p=0,939 ↗

Tableau 4.10 Probabilité d'indépendance entre la mesure d'imprécision et les notes moyennes, et variations des régressions linéaires associées.

les probabilités d'indépendance entre cette mesure et les notes ainsi que la croissance ou la décroissance des régressions linéaires correspondantes, que l'instabilité et la note moyenne semblent être indépendantes (les graphiques illustrant les régressions linéaires sont consultables en annexe, figures D.18 et D.20). En détaillant par critère, nous notons cependant que de manière marginalement significative, les notes d'immersion sont décroissantes en fonction de la mesure d'instabilité ($p=0,165$) après avoir écarté les deux participants pour lesquels celle-ci est la plus élevée (ce qui correspond à la ligne sans extrêmes). Il apparaît donc que l'instabilité de l'estimation du point de regard influe négativement sur le sentiment de présence des participants.

	Moyenne	Confort	Réalisme	Plaisir	Distances	Immersion
Instabilité	p=0,750 ↘	p=0,977 ↘	p=0,621 ↗	p=0,874 ↘	p=0,429 ↘	p=0,343 ↘
Instabilité sans extrêmes	p=0,449 ↘	p=0,715 ↘	p=0,286 ↘	p=0,713 ↘	p=0,414 ↗	p= 0,165 ↘

Tableau 4.11 Probabilité d'indépendance entre la mesure de stabilité et les notes moyennes, et variations des régressions linéaires associées.

Autres mesures de qualité

L'imprécision et l'instabilité ont été mesurées une unique fois pour chaque participant en début d'expérimentation. *A contrario*, il est ici question de mesures qui ont été enregistrées en temps réel au cours de chaque navigation, puis moyennées. Il s'agit des mesures de confiance des lunettes ETG en leur repérage des pupilles et d'incertitude en leur estimation des globes oculaires (dont certains détails sont fournis en section C en annexe), ainsi que nos propres mesures d'incertitude du calcul du point de regard utilisant les angles entre les différents POR. Les probabilités d'indépendance entre ces mesures et les notes sont réunies dans le tableau 4.12, qui indique également la croissance ou la décroissance des régressions linéaires associées (les graphiques correspondants sont consultables en annexe sur les figures D.21,

D.22, D.23, D.24, D.25 et D.26). Les observations se dégageant des valeurs de probabilité sont discutées dans les paragraphes suivants.

	Moyenne	Confort	Réalisme	Plaisir	Distances	Immersion
Pupilles	p=0,985 ↘	p=0,592 ↘	p=0,244 ↘	p=0,627 ↘	p=0,353 ↗	p= 0,121 ↗
Globes oculaires	p= 0,167 ↘	p= 0,156 ↘	p=0,671 ↗	p=0,213 ↘	p= 0,183 ↘	p= 0,057 ↘
Angle PORG-PORD	p=0,446 ↘	p=0,330 ↘	p=0,940 ↗	p=0,493 ↘	p=0,921 ↘	p=0,555 ↘
Moyenne des 3 angles	p=0,537 ↘	p=0,388 ↘	p=0,753 ↗	p=0,476 ↘	p=0,394 ↘	p=0,353 ↘

Tableau 4.12 Probabilités d'indépendance entre la mesure d'incertitude et la note moyenne, et variations des régressions linéaires associées.

Pupille¹ Aucune corrélation ne se décèle en première ligne du tableau 4.12 entre la moyenne des notes attribuées et la confiance moyenne des ETG en leur repérage des pupilles gauche et droite. Cependant, en détaillant par critère, nous remarquons que de manière marginalement significative les notes d'immersion croissent en même temps que cette mesure de confiance, ce qui laisse supposer un impact positif d'un repérage sûr des pupilles sur le sentiment de présence des participants, via la qualité du point de regard obtenu.

Globe oculaire² Cette fois-ci, la seconde ligne du tableau 4.12 révèle une corrélation marginalement significative entre la moyenne des notes et l'incertitude moyenne de la détection des globes oculaires gauche et droit, la première diminuant avec l'incertitude des ETG quant à cette détection. Ce résultat se retrouve pour les critères de confort, de perception des distances et d'immersion. Il apparaît donc qu'une estimation plus sûre des globes oculaires, en permettant comme dans le cas des pupilles d'utiliser des positions d'yeux et des vecteurs de direction de vue plus fidèles pour le calcul du POR, influe positivement sur les appréciations subjectives des participants.

Angle entre les POR Nous avons enfin considéré comme mesure de qualité les angles entre les différents POR présentés en section 3.3.2, en particulier celui entre PORG et PORD (calculés à partir des rayons obtenus pour les yeux gauche et droit) et la moyenne des 3 angles entre PORG, PORD et le POR effectivement utilisé. En effet, un angle important

1. *N. B.* Plus la valeur envoyée par les ETG est haute, plus la détection de la pupille est sûre.

2. *N. B.* Plus la valeur envoyée par les ETG est basse, plus la détection du globe oculaire est sûre.

entre ces POR indique que les deux directions de vue détectées par les ETG concordent mal, c'est-à-dire ne s'accordent pas sur la position du POR, conduisant à une estimation moins assurée de celui-ci. En considérant les probabilités d'indépendance indiquées dans les deux dernières lignes du tableau 4.12, aucune relation ne semble lier les angles entre les POR aux notations subjectives attribuées par les participants.

Les résultats obtenus en prenant en compte les mesures fournies par les ETG sur leur confiance en leur repérage des pupilles et leur incertitude en leur estimation des globes oculaires font état d'une relation entre la qualité de l'estimation du POR et les notes dans certains critères. Ils entrent donc en contradiction avec les conclusions de l'analyse de la mesure d'imprécision et de celle se basant sur les angles entre POR. Cependant, il est important de noter que la mesure d'imprécision a été réalisée une unique fois en début d'expérimentation et que certains facteurs ont pu par la suite affecter l'imprécision de l'estimation du point de regard, tel qu'un mouvement du participant qui aurait déplacé les lunettes relativement à sa tête. C'est pourquoi les données ETG nous semblent apporter des informations complémentaires intéressantes.

4.3.6 Fixations, saccades et nombre d'objets regardés

L'étude des mouvements oculaires, en particulier fixations et saccades, est source d'informations sur l'inspection visuelle de l'environnement par le participant, et peut expliquer ses préférences subjectives. Nous avons donc extrait les fixations et saccades intervenues durant la navigation, ainsi que le nombre d'objets regardés. Ce dernier, obtenu grâce à l'enregistrement du nom de l'objet intersecté par le rayon utilisé pour calculer le POR, est un indice de la quantité d'exploration visuelle de la scène.

Notre méthode de détection des saccades utilise l'écart angulaire entre deux positions consécutives du point de regard (voir la section 2.2.4). Nous considérons qu'une saccade débute dès que cet écart dépasse un certain seuil, et se termine lorsqu'il repasse en dessous, moment où commence la détection d'une fixation. Avant de nous intéresser à la relation entre ces données et les notations subjectives, nous étudions l'impact de la présence de flou sur celles-ci.

Il apparaît en effet que la présence ou l'absence de flou influence significativement les mouvements oculaires au cours de la navigation, ainsi que le nombre d'objets regardés. Nous observons sur la figure 4.6 que le nombre de fixations³ est significativement plus petit en présence de flou (moyenne de 84,87 contre 104,54), avec un temps moyen plus long (moyenne de 977,38 ms contre 800,39 ms). Les saccades sont également plus longues (moyenne de 130,55

3. La séquence des mouvements oculaires étant fractionnée en saccades et fixations alternées, leurs nombres sont égaux à un ou deux près, donc nous n'utilisons que le nombre de fixations dans le reste de cette section.

ms contre 96,18 ms), la distance angulaire parcourue au cours de celles-ci étant cependant pratiquement la même (moyenne $21,35^\circ$ contre $20,88^\circ$). Le nombre moyen d'objets regardés passe quant à lui de 124,4 à 111,91 après ajout de flou. Les résultats des tests ANOVA à 1 et 2 facteurs contrôlés réalisés sont regroupés dans le tableau 4.13. Nous interprétons ce résultat par le fait que le ralentissement du nombre d'images par secondes dans les configurations avec flou (passant d'une moyenne de 12,96 à 9,38) incite les participants à s'adapter consciemment ou inconsciemment en ralentissant sa vitesse d'inspection, ce qui semble également indiquer que la vitesse d'adaptation du flou est insuffisante. L'influence significative de la présence de flou sur la note de perception des distances et des profondeurs et celle marginalement significative sur le sentiment d'immersion (vues dans le tableau 4.2) peuvent s'expliquer par la durée significativement plus importante des fixations et des saccades.

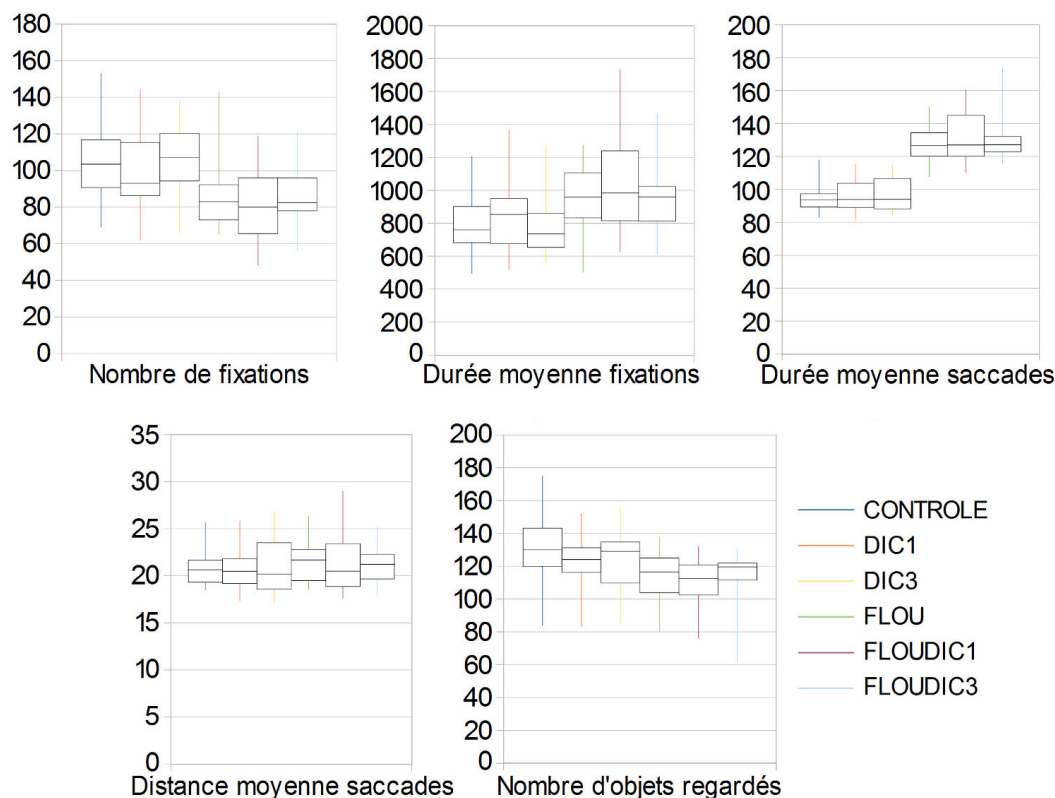


Figure 4.6 Données quantitatives de l'inspection visuelle par configuration.

Nous nous sommes penchés plus en détail sur la relation entre ces mesures de la quantité d'inspection visuelle et les notes subjectives attribuées. Le tableau 4.14 résume les probabilités d'indépendance entre ces variables et indique les variations des régressions linéaires correspondantes. Les paragraphes suivants présentent nos remarques.

Critère	Configuration	DIC	Flou	Interaction
Nombre fixations	F=5,35 p= 0,000	F=0,59 p=0,556	F=25,5 p= 0,000	F=0,02 p=0,980
Durée fixations	F=3,76 p= 0,004	F=0,94 p=0,395	F=16,9 p= 0,000	F=0,03 p=0,971
Durée saccades	F=47,3 p= 0,000	F=0,39 p=0,680	F=236 p= 0,000	F=0,16 p=0,852
Distance saccades	F=0,23 p=0,949	F=0,01 p=0,992	F=1,02 p=0,316	F=0,06 p=0,946
Nombre objets regardés	F=2,83 p= 0,020	F=0,91 p=0,405	F=12,2 p= 0,000	F=0,09 p=0,912

Tableau 4.13 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur la quantité d’inspection visuelle.

	Confort	Réalisme	Plaisir	Distances	Immersion
Nombre fixations	p=0,713 ↗	p=0,672 ↗	p=0,427 ↗	p=0,728 ↗	p=0,584 ↗
Durée fixations	p=0,962 ↘	p=0,949 ↗	p=0,642 ↘	p=0,747 ↘	p=0,734 ↘
Durée saccades	p=0,320 ↘	p=0,917 ↗	p=0,584 ↘	p=0,563 ↗	p=0,729 ↘
Nombre objets regardés	p=0,211 ↗	p=0,547 ↗	p= 0,088 ↗	p=0,232 ↗	p=0,586 ↗

Tableau 4.14 Probabilités d’indépendance entre la mesure de quantité d’inspection visuelle et la note moyenne, et variations des régressions linéaires associées.

Nombre de fixations

Aucune probabilité significative ou marginalement significative de corrélation n’apparaît dans le tableau 4.14 entre le nombre moyen de fixations et les notes subjectives attribuées. Par ailleurs, une remarque annexe ressort des graphiques qui présentent les notes non pas moyennées mais par configuration en ce qui concerne le critère de plaisir. En effet, nous observons pour ce critère sur la figure 4.7 un « triangle vide » en bas à droite, indiquant que lorsque le nombre de fixations est important la note est meilleure (les graphiques pour les autres critères sont consultables en annexe, sur la figure D.27). Il est cependant difficile d’établir le sens de l’implication : le participant a-t-il tendance à déplacer plus souvent ses yeux et à fixer un nombre plus important d’objets lorsqu’il apprécie la navigation, ou au contraire est-ce le fait de parcourir la scène plus minutieusement qui fait que le participant apprécie davantage la navigation ?

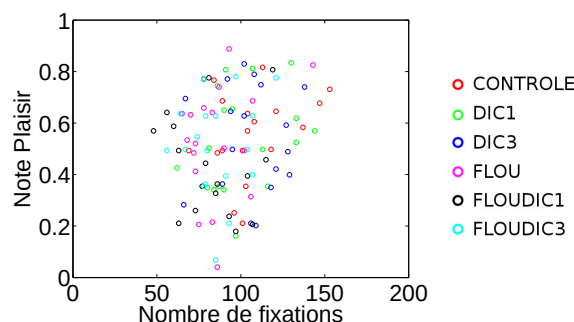


Figure 4.7 Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre de fixations.

Durée moyenne des fixations et saccades

Une fois de plus, les valeurs de probabilités du tableau 4.14 (complémenté par les figures D.28 et D.29 en annexe) ne révèlent pas de corrélation significative liant les durées moyennes des fixations et des saccades aux notes subjectives.

4.3.7 Nombre d'objets regardés

Comme cité en section 4.3.1, plusieurs participants ont déclaré que leur plaisir de navigation augmentait au fur et à mesure des navigations en raison du fait qu'ils remarquaient de plus en plus d'objets dans la scène. Bien que nous n'ayons pas observé d'accroissement significatif inter-navigations, il apparaît au vu du tableau 4.14 que le nombre d'objets regardés au cours d'une même navigation influe positivement et de manière marginalement significative sur les notes de plaisir (les graphiques illustrant les régressions linéaires sont consultables en figure D.30 en annexe). Or, la présence de flou diminuant significativement cette mesure de l'exploration visuelle, nous pouvons conclure qu'il tend également à réduire le plaisir, bien que son impact sur ce critère n'ait pas été révélé de manière évidente dans le tableau 4.2. Les autres critères de notation ne semblent quant à eux pas être impactés par le nombre d'objets regardés.

4.3.8 Cartes de densité, chemin visuel et objets d'intérêt

L'analyse générale des notes subjectives n'a pas conduit à l'émergence d'une configuration significativement préférée aux autres, même en reliant ces notes à d'autres facteurs tels que la précision de l'obtention du point de regard. Cependant, d'autres outils non statistiques permettent de la compléter, tels que les cartes de densité et les chemins de balayage présentés en 2.5.1. Grâce à l'enregistrement continu des informations de position du participant et de

son point de regard dans la scène au cours des expérimentations, nous pouvons générer les cartes et chemins visuels correspondants, et visualiser où le regard du participant s'est concentré au cours des navigations.

Concernant les cartes de densité, chaque position de point de regard est représentée par une sphère, dont la couleur dépend du nombre de sphères dans son voisinage, c'est-à-dire du nombre de sphères en collision avec elle. La couleur va du bleu au rouge (en passant par le vert, le jaune et l'orange), la gamme de couleurs étant répartie en fonction du nombre maximum de collisions entre les sphères pour une navigation donnée.

Les exemples de cartes de densité de la figure 4.8 sont issus des navigations du participant présentant le plus grand écart de notes entre deux configurations, tous critères confondus. Il s'agit de la configuration **DIC3**, qui a obtenu une note moyenne de 0,85, et **FLOUDIC3**, qui n'a obtenu que 0,46. Nous pouvons noter à l'aide de ces images certaines différences au niveau de l'inspection visuelle de la scène entre les deux navigations. Le participant a par exemple regardé le lustre de la chambre avec **DIC3** mais pas avec **FLOUDIC3**, son attention s'étant surtout concentrée sur le lit au cours de cette dernière. Il s'est également davantage concentré sur la table de la cuisine et la table basse du salon avec **FLOUDIC3**, tandis qu'au cours de la navigation **DIC3** son regard semble parcourir plus largement les pièces, en particulier le sol du salon. Pourtant, en termes de nombre d'objets regardés, la configuration **DIC3** n'en compte que 97 contre 110 pour la configuration **FLOUDIC3**.

Comme expliqué précédemment, le chemin suivi par le POR de l'observateur donne quant à lui des indications temporelles, car il permet de distinguer la séquence des fixations. La figure 4.9 présente les chemins visuels pour le même participant.

Nous avons enfin développé un outil permettant d'identifier dans la scène un ou plusieurs objets particuliers, les objets d'intérêt, dont on souhaite connaître l'inspection visuelle par un participant. Nous avons par exemple effectué des analyses sur le lustre présent dans la chambre, qui n'a pas été du tout regardé dans le cas de la configuration **FLOUDIC3**. Dans le cas **DIC3**, l'objet a été observé 3 fois, durant un temps total de 443 ms, soit 147,67 ms en moyenne par passe.

Une étude plus poussée de ces données réalisée sur l'ensemble des participants permettrait éventuellement d'apporter des explications sur les différences de notation, mais requerrait des connaissances en psychologie quant à l'inspection visuelle naturelle d'une scène.

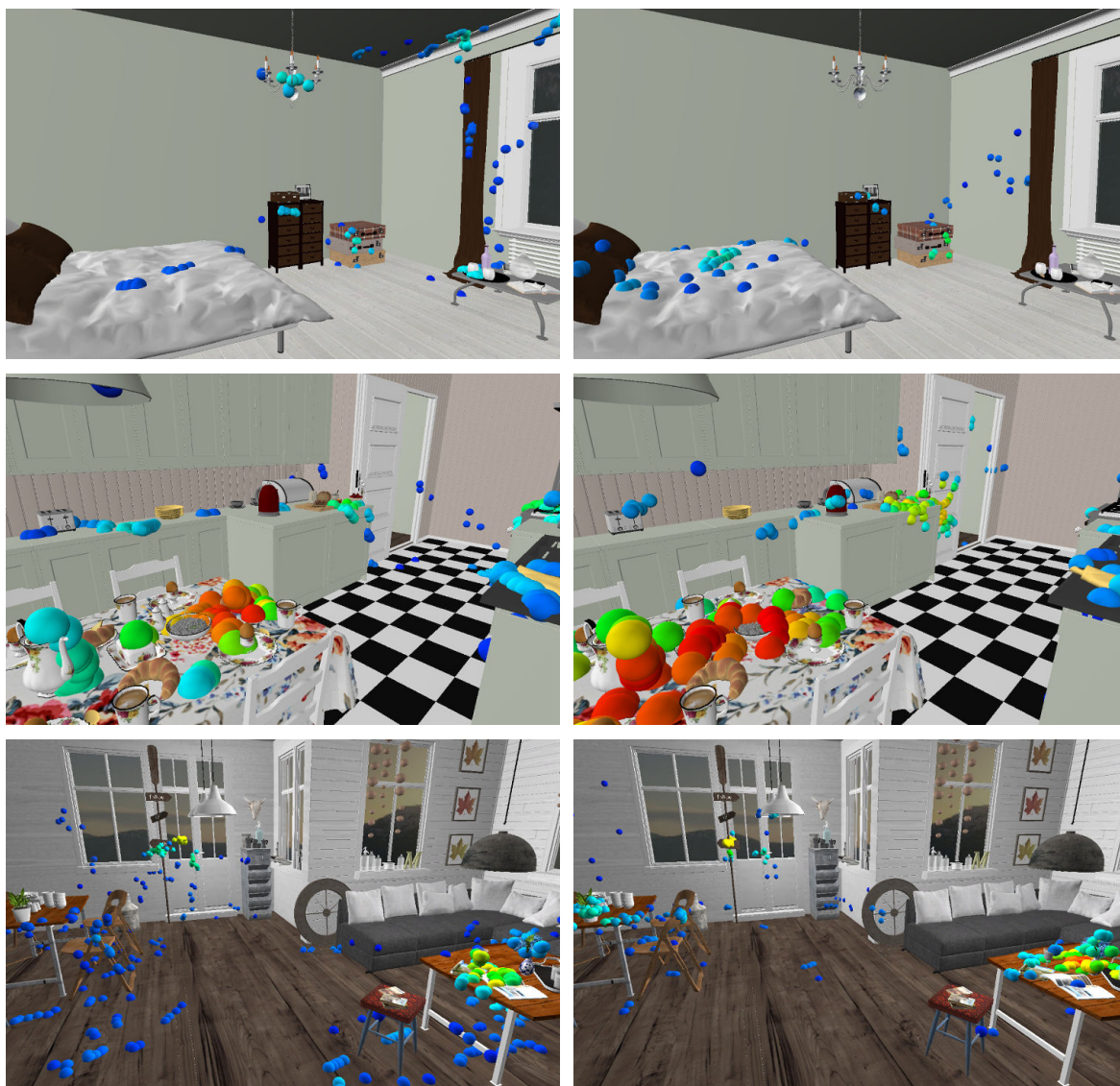


Figure 4.8 Exemples de cartes de densité, pour les configurations **DIC3** (à gauche) et **FLOU-DIC3** (à droite).

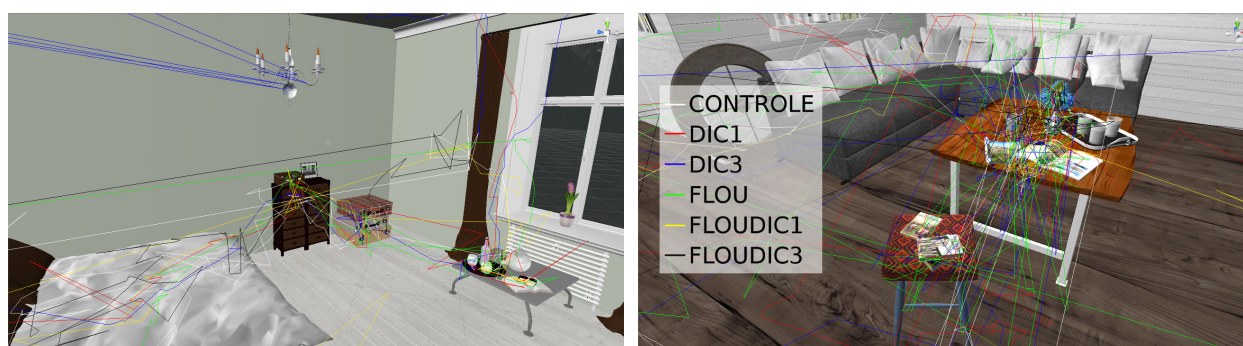


Figure 4.9 Exemples de chemins visuels.

4.4 Limites de fusion

Cette étape de l'expérimentation avait pour but de comparer objectivement l'efficacité de chaque configuration en termes de capacité de fusion. Pour cela, les participants devaient éloigner ou rapprocher deux objets virtuels, dans le premier sens jusqu'à pouvoir fusionner les images stéréoscopiques, dans le second jusqu'à la perte de cette fusion. Afin d'obtenir des données plus significatives, chaque mesure fut réalisée à 3 reprises pour chaque configuration.

4.4.1 Effet de la configuration stéréoscopique

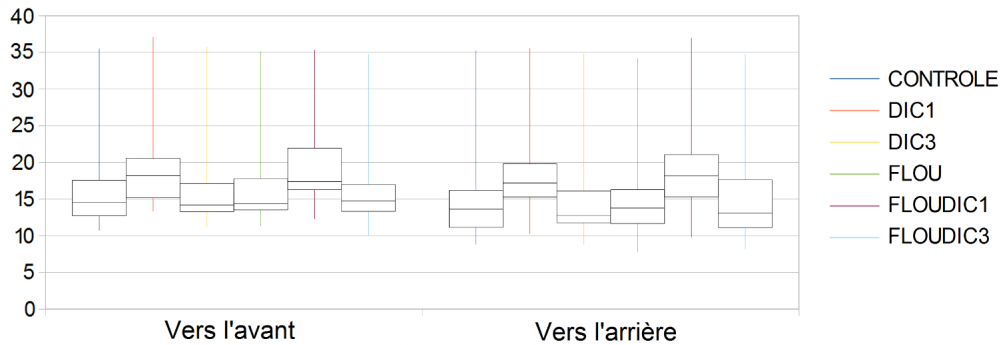


Figure 4.10 Limites de fusion par configuration, pour les cas vers l'avant et vers l'arrière.

La figure 4.10 réunit les limites atteintes avec chacune des configurations, dans les cas vers l'avant et vers l'arrière, la valeur d'une limite (en cm) correspondant à la distance au participant. Nous remarquons tout d'abord que les limites vers l'arrière sont globalement légèrement inférieures à celles obtenues dans le cas vers l'avant, avec une moyenne de 16,24 cm contre 17,11 cm. Ceci est probablement dû au fait que dans ce premier sens les participants s'arrêtaient après perte de fusion, tandis que dans l'autre ils devaient éloigner les objets jusqu'à fusionner à nouveau les images stéréoscopiques, les limites de fusion à proprement parler se situant donc entre les résultats obtenus pour chacun de ces cas.

Les résultats des tests ANOVA à 1 facteur et deux facteurs contrôlés effectués sont regroupés dans le tableau 4.15. Ils indiquent que la configuration stéréoscopique n'influe pas de manière significative sur les limites de fusion. Cependant, en séparant les effets des deux paramètres, nous observons que la DIC a un impact significatif contrairement à la présence ou à l'absence de flou.

Ces résultats se remarquent également dans le tableau 4.16 qui présente les moyennes et dénombre, pour chaque configuration, les participants ayant obtenu la limite la plus basse avec celle-ci. En séparant par DIC et par présence ou absence de flou, on obtient que la DIC la plus efficace est la DIP physiologique (avec 19 participants au total, contre 3 et 14), et que

le flou ne semble pas avoir d'impact significatif (19 participants contre 17 sans). Pourtant, l'étude de la littérature nous avait amenés à penser que la présence d'un flou de profondeur de champ facilitait la fusion des images stéréoscopiques. Nous supposons que l'absence d'impact est ici due au fait que pour les limites de fusion atteintes par la plupart des participants, les objets prenaient une part très importante de l'écran, l'arrière-plan flou se retrouvant donc minoritaire dans l'image.

	Configuration	DIC	Flou	Interaction
Vers l'avant	F=1,57 p= 0,174	F=3,93 p= 0,023	F=0,00 p=0,972	F=0,01 p=0,991
Vers l'arrière	F=1,92 p= 0,098	F=4,74 p= 0,011	F=0,09 p=0,765	F=0,01 p=0,986

Tableau 4.15 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les limites de fusion.

	CONTROLE	DIC1	DIC3	FLOU	FLOUDIC1	FLOUDIC3
Vers l'avant	0,161	0,191	0,161	0,161	0,193	0,160
Vers l'arrière	0,151	0,185	0,146	0,152	0,189	0,151
Moyenne	0,156	0,188	0,153	0,156	0,191	0,155
Vers l'avant	5	0	3	3	1	6
Vers l'arrière	7	1	1	4	1	4
Participants	12	1	4	7	2	10

Tableau 4.16 Moyennes et dénombrement pour chaque configuration des participants ayant atteint la limite la plus basse.

4.4.2 Effet de la DIC

Nous notons sur la figure 4.10 que **DIC1** et **FLOUDIC1**, qui partagent la même DIC, conduisent à des limites plus élevées. En séparant ces configurations des autres, nous obtenons par un test de Student apparié qu'elles sont significativement différentes, en l'occurrence moins efficaces ($p=4,3.10^{-8}$). Ce résultat peut s'expliquer par la valeur de DIC utilisée qui est pour la plupart des participants plus importante que celles employées par les autres configurations, étant donné que **DIC3** et **FLOUDIC3** utilisent pour des distances inférieures à 0,4 m le choix *DIC3 Proche*, plus proche de la DIP anatomique. Or, plus la distance inter-caméras est importante, plus les disparités horizontales entre images gauche et droite seront grandes, pouvant résulter en une difficulté accrue de fusion.

La figure 4.11 (complémentée par E.1 en annexe) présente la moyenne des limites obtenues vers l'avant et vers l'arrière pour les deux configurations (avec et sans flou de profondeur

de champ) utilisant la DIC représentée en abscisse, en écartant le participant aux limites les plus importantes. Nous notons effectivement qu'une DIC élevée entraîne des difficultés significativement plus importantes à fusionner des objets proches. Cependant, il ne faut pas oublier que ce paramètre influe également sur la perception de la profondeur, et bien qu'une DIC nulle (résultant en deux images identiques) permettrait sans doute de fusionner des objets très proches, cela se ferait au détriment de la vision en 3D. Il est donc important de comparer les résultats des limites de fusion aux notes subjectives sur la perception de la profondeur.

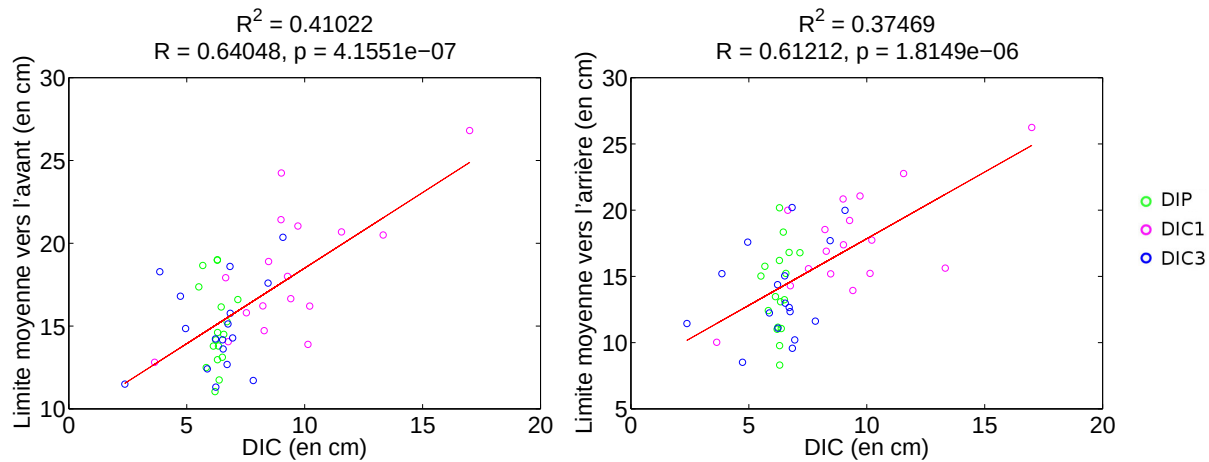


Figure 4.11 Limites obtenues vers l'avant et vers l'arrière en fonction de la DIC, en écartant le participant aux limites les plus importantes.

4.4.3 Lien entre notes subjectives et limites de fusion

Au cours de la navigation, certains objets de la scène se retrouvent très proches de l'observateur, sa difficulté à les fusionner pouvant alors influencer négativement sur ses notes, et ce dans tous les critères. En effet, outre l'inconfort et la mauvaise perception des distances résultant de la diplopie, celle-ci peut également rappeler au participant la présence du système stéréoscopique et diminuer ainsi l'immersion, le réalisme et le plaisir. Cependant, les valeurs de probabilité regroupées dans le tableau 4.17 (complémenté par les figures E.2 et E.3 présentées en annexe) n'établissent pas de lien significatif entre la moyenne des limites de fusion et la moyenne des notes attribuées, avec cependant un résultat marginalement significatif vers l'arrière, après avoir écarté le participant extrême, tendant au contraire à supposer que plus la limite de fusion est haute meilleures sont les notes.

Concernant le critère de perception des distances, le fait que la limite de fusion vers l'arrière influe positivement sur les notes de façon marginalement significative s'explique selon

nous par le phénomène expliqué dans la section précédente : des disparités plus importantes provoquent à la fois une limite de fusion plus haute et une image plus « bombée », une impression de profondeur accrue.

	Note moyenne	Note distances
Vers l'avant	p=0,802 ↗	p=0,870 ↘
Vers l'avant sans extrême	p=0,932 ↘	p=0,576 ↗
Vers l'arrière	p=0,321 ↗	p=0,764 ↗
Vers l'arrière sans extrême	p= 0,197 ↗	p= 0,173 ↗

Tableau 4.17 Probabilités d'indépendance entre les limites et les notes (moyennes ou de distances), et variations des régressions linéaires associées.

4.4.4 Effet de l'imprécision de l'estimation du POR

Le tableau 4.18 (complémenté par la figure E.4 en annexe) n'indique pas d'effet de l'imprécision de l'estimation du point de regard sur la valeur des limites de fusion lorsque tous les participants sont pris en compte. En revanche, en écartant les trois plus hautes imprécisions (lignes sans extrêmes), nous observons une corrélation marginalement significative entre ces deux variables, une imprécision plus haute semblant diminuer la difficulté de fusion (limites plus basses). Ce résultat est étonnant car pour ces distances proches, les objets prenaient une part importante de l'écran de devant, le point de regard étant presque systématiquement estimé à leur niveau mis à part lorsque le participant regardait volontairement à côté. L'imprécision n'avait donc normalement pas d'impact sur l'affichage stéréoscopique.

	Précision
Vers l'avant	p=0,630 ↘
Vers l'avant sans extrêmes	p=0,268 ↘
Vers l'arrière	p=0,586 ↘
Vers l'arrière sans extrêmes	p= 0,182 ↘

Tableau 4.18 Probabilités d'indépendance entre la mesure d'imprécision et les limites, et variations des régressions linéaires associées.

4.5 Comparaison de mouvements oculaires

Pour ce dernier test, le participant devait suivre visuellement une cible, fixée dans un premier temps au dessus d'un robot réel cheminant le long d'un parcours tracé sur une table et dans un second temps sur le même robot reproduit en virtuel. Au cours de la navigation, le

participant a pu donner son avis subjectif sur sa perception des distances et des profondeurs. *A contrario*, ce test fournit une mesure objective de la restitution de la profondeur d'une scène réelle par une configuration virtuelle. En effet, si celle-ci est correcte, en particulier grâce aux paramètres de la configuration du système de caméras, nous nous attendons à ce que les variations de l'angle de vergence et de l'écart angulaire à la cible, témoins du chemin visuel, soient similaires à celles obtenues dans le cas réel. À l'inverse, un rendu peu réaliste peut générer des différences significatives entre les variations de vergence ou d'écart à la cible réels et virtuels, par exemple des angles globalement plus importants ou globalement plus faibles, ou encore des écarts importants de minimums et maximums. Le chemin visuel pourrait également s'avérer plus tâtonnant, le participant effectuant des allées et venues en profondeur dans la scène en raison d'un ressenti ambigu des distances (ou éventuellement en raison d'un inconfort visuel).

4.5.1 Différence de vergence entre cas réel et virtuel

La vergence calculée est l'angle entre les vecteurs de direction des yeux gauche et droit tels qu'enregistrés par les ETG. Nous distinguons également la vergence en X, c'est-à-dire calculée en utilisant la projection normalisée de ces vecteurs sur le plan XZ, ainsi que la vergence en Y, projection sur le plan YZ. L'orientation de l'angle en X a été déterminée de manière à distinguer les cas où les yeux convergent et celui où ils divergent. En effet, les données renvoyées par les lunettes indiquent parfois des angles divergents : dans ce cas, l'angle en X sera étiqueté négatif.

La figure 4.12 représente, pour un participant, les courbes de sa vergence totale et de sa vergence en X divisées par la distance au robot en fonction du temps. En effet, étant donné que la tête du participant n'était pas maintenue à l'aide d'un appareillage pour demeurer fixe, la distance entre le robot réel ou virtuel et le centre des yeux pouvait varier au cours de l'expérimentation – une distance à la cible plus petite entraînant une vergence plus importante. La figure 4.13 illustre quant à elle les vergences totale, en X et en Y en fonction de cette distance.

Les valeurs d'angle de vergence et de distance provenant de fichiers enregistrés indépendamment, elles ont dû être interpolées grâce aux informations de temps inscrites dans chacun de ceux-ci. Elles ont également été lissées en supprimant les valeurs aberrantes de vergence correspondant principalement aux clignements des yeux. Cette opération a été réalisée par Fannie CARRIER, employée spécifiquement à cette fin par l'Institut Philippe-Pinel de Montréal.

En remarque préliminaire, nous observons de manière commune à la plupart des participants que la vergence en Y, et par conséquent la vergence totale, présente plusieurs valeurs

pour une même distance (phénomène particulièrement visible pour les distances inférieures à 0,6 m). Il est pratiquement possible de retracer le parcours du robot sur le graphe. Nous expliquons ce phénomène grâce aux observations décrites en section 3.3.2 sur les différences de coordonnée Y des deux vecteurs de direction lorsque l'individu balaye du regard une droite horizontale.

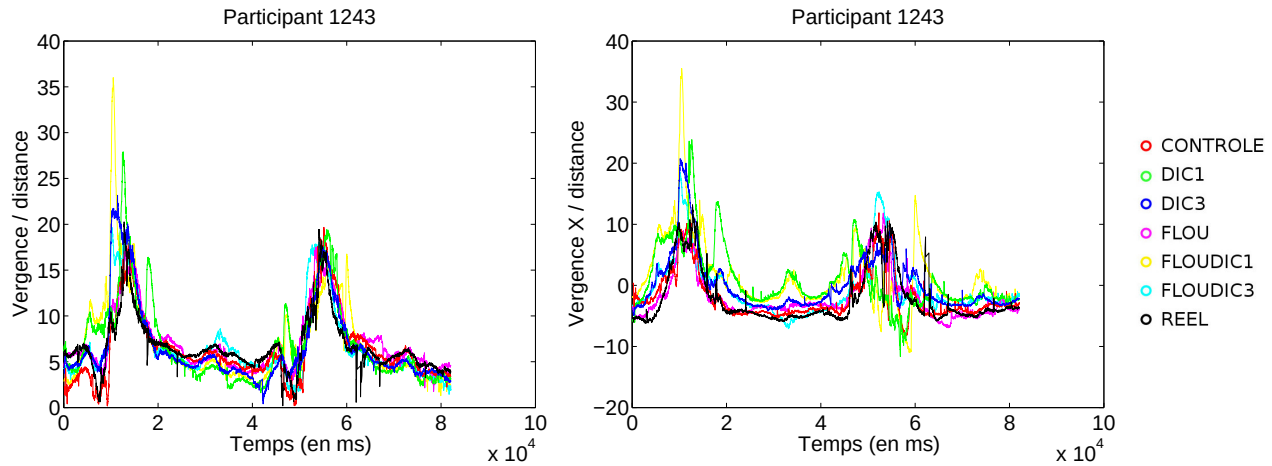


Figure 4.12 Vergence/distance (à gauche) et vergence en X/distance (à droite) en fonction du temps.

	Configuration	DIC	Flou	Interaction
Vergence	F=0,12 p=0,987	F=0,09 p=0,911	F=0,01 p=0,934	F=0,22 p=0,807
Vergence en X	F=0,22 p=0,952	F=0,48 p=0,619	F=0,01 p=0,935	F=0,07 p=0,933

Tableau 4.19 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les différences de vergence/distance et vergence en X/distance entre les cas virtuels et réel.

La figure 4.14 présente les différences de vergence⁴ obtenues par chacune des configurations virtuelles en comparaison de la configuration réelle. De prime abord, aucune ne semble se démarquer positivement ou négativement de manière significative. Ce résultat est confirmé par les tests ANOVA à 1 et 2 facteurs contrôlés dont les résultats sont réunis dans le tableau 4.19. Ces tests ont été réalisés trois fois : en considérant tous les participants ou en écartant certains groupes (comptant respectivement 2 et 6 individus) dont les graphiques s'avèrent bruités ou très bruités même après lissage, sans obtention de résultats plus significatifs. Nous remarquons cependant qu'au regard des deux vergences, la médiane la plus basse est atteinte

4. Les différences de vergence dont il est question dans le reste de la section correspondent en réalité aux différences de vergence divisée par la distance.

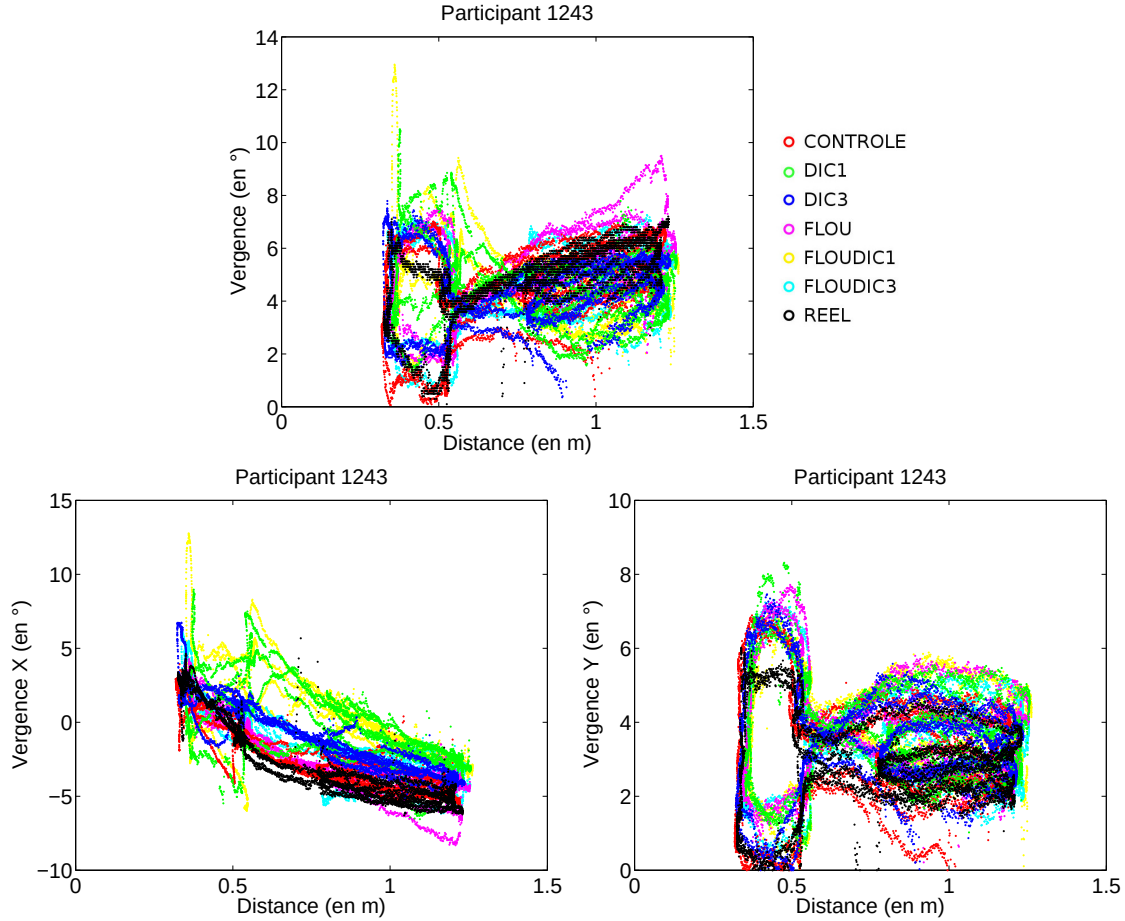


Figure 4.13 Vergence totale (en haut), vergence en X (en bas à gauche) et vergence en Y (en bas à droite) en fonction de la distance au robot.

avec la configuration **CONTROLE**, suivie de **DIC3** en considérant la vergence totale et de **FLOU** en considérant la vergence en X. La présence de flou ne semble pas impacter le réalisme de la représentation des profondeurs, avec des moyennes de médianes de 1,99 et 2,49 contre 1,91 et 2,51 sans flou.

Nous notons également que comme dans le cas des limites de fusion, la DIC commune aux configurations **DIC1** et **FLOUDIC1** conduit à des différences d'angles de vergence plus élevées. Ce phénomène est observable sur les graphiques présentés en annexe F.1 : les angles de vergence en X des configurations **DIC1** et **FLOUDIC1** sont la plupart du temps situés au dessus des autres. Nous obtenons par un test de Student apparié que ces deux configurations sont significativement différentes des autres, en l'occurrence moins performantes ($p = 0.0364$).

Le tableau 4.20 dénombre quant à lui pour chaque configuration les participants ayant obtenu la meilleure approximation de la vergence réelle avec celle-ci, et présente les moyennes.

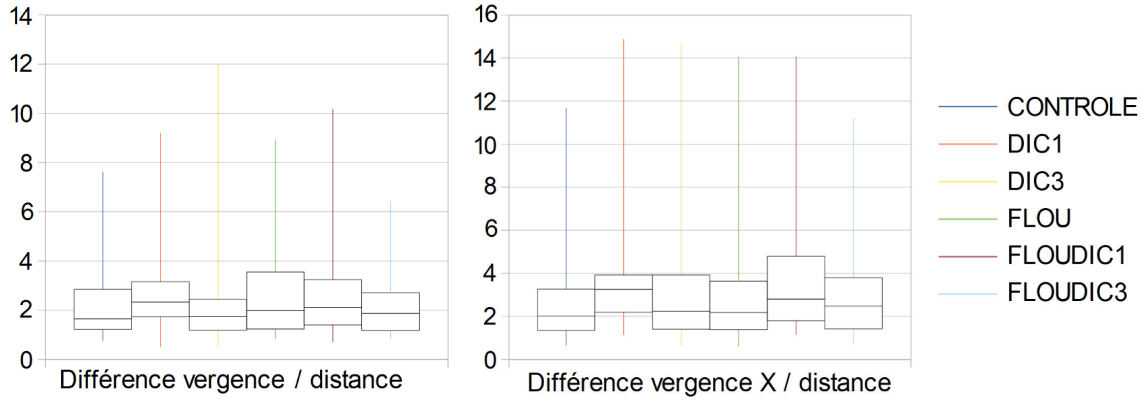


Figure 4.14 Différence de vergence/distance et vergence en X/distance entre les cas virtuels et réel.

En termes de nombre de participants, il s'avère donc que c'est la configuration **DIC3** qui a conduit à des mouvements de vergence plus réalistes, suivie par les configurations **FLOU-DIC3** en considérant la vergence et **FLOU** en considérant la vergence en X. La présence du flou de profondeur de champ ne révèle pas ici non plus une influence nette, en comptant 10 et 10 participants contre 8 et 8 sans flou.

	CONTROLE	DIC1	DIC3	FLOU	FLOUDIC1	FLOUDIC3
Vergence	2,366	2,646	2,622	2,636	2,641	2,258
Vergence en X	3,092	3,989	3,369	3,474	3,902	3,230
Moyenne	2,729	3,317	2,995	3,055	3,271	2,744
Vergence	1	2	5	3	3	4
Vergence en X	1	1	6	6	2	2
Participants	2	3	11	9	5	6

Tableau 4.20 Moyennes et dénombrement pour chaque configuration des participants ayant obtenu les meilleurs résultats en terme de vergence oculaire.

Comparaison avec la perception subjective des distances

La recherche d'un lien entre les notes attribuées pour le critère de perception des distances et les différences de vergence obtenues au cours du test du robot nous semble intéressante et importante, car elle permet de comparer l'impression subjective de réalisme des distances ressentie au cours de la navigation à la réalité objective des mouvements oculaires en réponse aux profondeurs virtuelles. Le tableau 4.21 réunit les valeurs de probabilité obtenues pour l'indépendance entre ces deux variables, les graphiques illustrant les régressions linéaires associées pouvant être consultés en annexe (figure F.2). Aucune relation significative ne ressort

pour le critère de perception des distances, indiquant que les participants n'ont pas systématiquement attribué les meilleures notes dans ce critère à la configuration qui a conduit aux mouvements oculaires les plus réalistes au cours du test du robot. Nous remarquons cependant dans la première colonne du tableau que la différence de vergence en X/distance semble être corrélée de manière marginalement significative à la note moyenne, la croissance de la régression linéaire indiquant étonnamment que plus la différence est importante meilleure est la note. En observant les données brutes, nous remarquons que cela provient principalement du fait que les deux participants ayant eu des différences de vergence en X/distance extrêmes ont attribué en moyenne des bonnes notes aux configurations.

	Note moyenne	Note distances
Vergence	p=0,564 ↗	p=0,529 ↘
Vergence en X	p= 0,185 ↗	p=0,900 ↘

Tableau 4.21 Probabilités d'indépendance entre les différences de vergence/distance et vergence en X/distance et les notes (moyennes ou de distances), et variations des régressions linéaires associées.

Effet de l'imprécision de l'estimation du POR

La qualité de l'estimation du POR a joué un rôle important dans le test du robot, au moment de la mise à jour des paramètres des configurations stéréoscopiques virtuelles. En particulier, la cible sur laquelle les participants devaient fixer leur regard était de petite taille, ce qui a eu pour conséquence qu'à certains moments les rayons de vue ne l'intersectaient pas et s'arrêtaient sur le mur du fond de la voûte, provoquant entre autres le floutage du robot ou un changement brutal de valeur de disparité en fonction de la configuration courante. Cependant, le tableau 4.22 (complémenté par la figure F.3 en annexe) n'indique pas d'effet de l'imprécision de l'estimation du point de regard sur les différences de vergence et vergence en X entre les cas virtuels et réel, et ce même après avoir écarté les plus hautes imprécisions (correspondant aux lignes sans extrêmes).

	Précision
Vergence	p=0,537 ↗
Vergence sans extrêmes	p=0,267 ↘
Vergence en X	p=0,965 ↗
Vergence en X sans extrêmes	p=0,232 ↘

Tableau 4.22 Probabilités d'indépendance entre la mesure d'imprécision et les différences de vergence/distance et vergence en X/distance, et variations des régressions linéaires associées.

Remarque sur les capacités de fusion

En dehors de l'étude générale des résultats que nous venons d'effectuer, une remarque anecdotique se dégage de l'observation indépendante des résultats de chaque participant. En effet, nous remarquons que certains présentent des comportements oculaires très différents en cas réel et virtuel, tel que le présente la figure 4.15, nous amenant à douter en leur capacité de percevoir la profondeur avec un affichage stéréoscopique bien qu'aucun n'ait indiqué de problème à ce sujet dans le questionnaire rempli en début d'expérimentation.

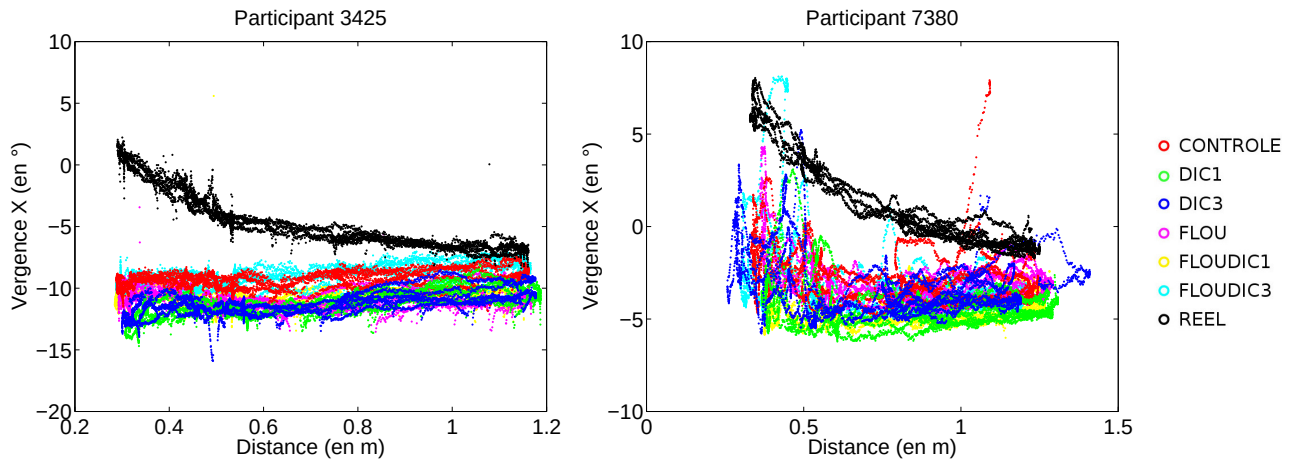


Figure 4.15 vergence en X en fonction de la distance au robot pour deux participants ne semblant pas percevoir correctement la profondeur avec un affichage stéréoscopique.

4.5.2 Écart angulaire à la cible

La seconde mesure objective de comparaison des mouvements oculaires dans les cas virtuels et réel est l'écart angulaire à la cible, c'est-à-dire dans notre cas l'angle entre le vecteur partant du centre des yeux et dirigé vers la cible (le point de repère sur le robot, réel ou virtuel) et celui partant du centre des yeux et dirigé vers le POR. Ce dernier correspond au vecteur de direction moyen du regard obtenu grâce aux ETG. Les données ont été lissées en supprimant les mêmes valeurs que pour la vergence, c'est-à-dire celles correspondant principalement aux clignements des yeux, mais nous remarquons pour certains participants qu'un lissage spécifique pourrait s'avérer nécessaire. À noter que les analyses présentées ci-dessous ne prennent pas en compte l'un des participants dont les mesures d'écart angulaire présentent des erreurs importantes (angle supérieur à 120° plus de 50% du temps avec les configurations **DIC3** et réelle). La figure 4.16 représente, pour un participant, les courbes de ses écarts angulaires à la cible, total et en X, en fonction du temps.

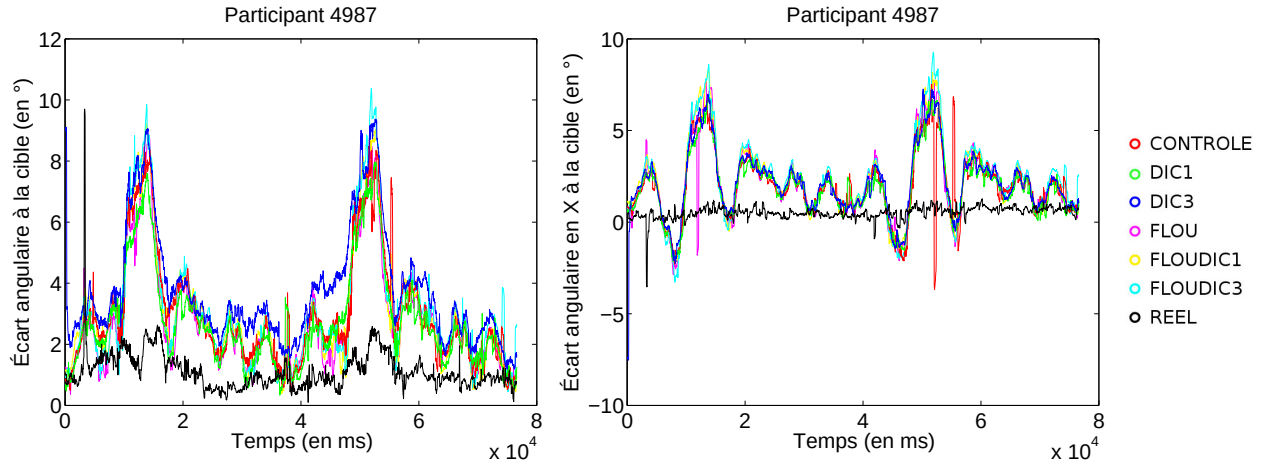


Figure 4.16 Écart angulaire (à gauche) et écart angulaire en X (à droite) en fonction du temps.

Concernant la configuration réelle tout d’abord, nous observons que l’écart en X tourne autour de 0, indiquant que le suivi du robot s’effectue correctement pour cette configuration (d’autres exemples sont présentés en figure F.4 en annexe). Cependant, la figure F.5 en annexe expose le cas de participants pour lesquels l’angle en X dans la configuration réelle reste relativement constant, mais autour d’une valeur différente de 0. Nous pensons que cela peut provenir d’un léger déplacement des lunettes relativement à la tête de l’observateur survenu après l’étape de calibration. Ceci induit un décalage constant entre les vecteurs de direction exprimés dans le repère monde et les directions réelles de regard du participant, mais n’influe pas *a priori* sur les différences de résultats entre les configurations virtuelles. Par ailleurs, nous interprétons les différences entre écart total et écart en X par la présence d’erreurs dans la coordonnée en Y des vecteurs de direction reçus des lunettes ETG, discutée dans la section précédente et en section 3.3.2.

En ce qui concerne les configurations virtuelles, nous observons des écarts angulaires relativement importants, les pics les plus élevés correspondant aux moments où le robot est proche du participant. Nous pensons que ces différences par rapport au cas virtuel, qui semblent communes à toutes les configurations virtuelles, peuvent être dues au dispositif d’affichage stéréoscopique. En effet, si celui-ci présente des erreurs de calibration, il peut ne pas restituer correctement la scène virtuelle, en affichant par exemple le robot plus à droite ou plus à gauche du participant qu’il ne l’est vraiment dans le repère du monde virtuel, et ce quelles que soient les valeurs des paramètres du système de caméras.

La figure 4.17 présente pour chaque configuration la différence entre ses écarts angulaires à la cible, totaux et en X, par rapport à ceux obtenus avec la configuration réelle. Aucune ne semble se démarquer positivement ou négativement, ce qui est confirmé par les tests ANOVA

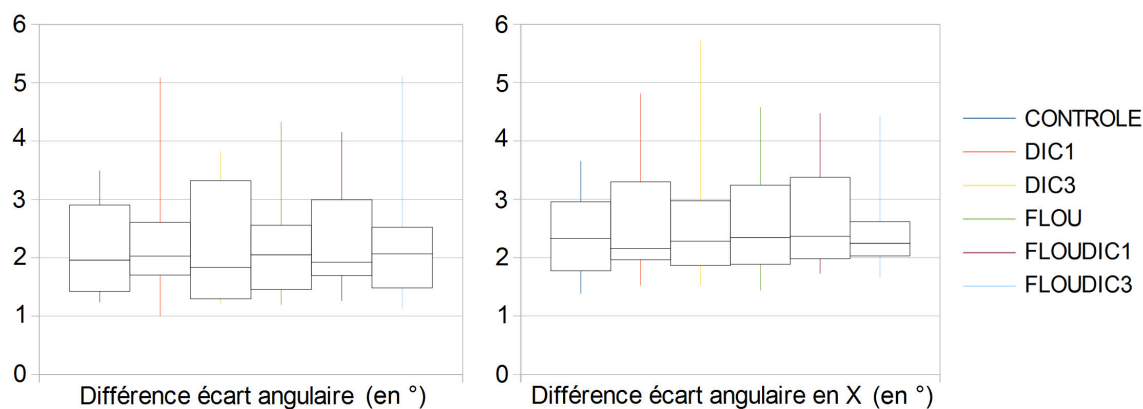


Figure 4.17 Différence d'écart angulaire et écart angulaire en X entre les cas virtuels et réel.

	Configuration	DIC	Flou	Interaction
Écart angulaire	F=0,07 p=0,997	F=0,14 p=0,873	F=0,00 p=0,959	F=0,03 p=0,970
Écart angulaire en X	F=0,31 p=0,908	F=0,55 p=0,590	F=0,08 p=0,773	F=0,17 p=0,840

Tableau 4.23 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les différences d'écart angulaire et écart angulaire en X entre les cas virtuels et réel.

à 1 et 2 facteurs contrôlés dont les résultats sont regroupés dans le tableau 4.23. Nous notons cependant que les configurations obtenant les médianes les plus basses pour l'écart angulaire total sont **DIC3** suivie de **FLOUDIC1**, tandis qu'il s'agit de **DIC1** suivie de **FLOUDIC3** en considérant l'écart angulaire en X.

Le tableau 4.24 dénombre quant à lui pour chaque configuration les participants ayant obtenu la meilleure approximation de l'écart angulaire réel avec celle-ci, et présente les moyennes. En termes de nombre de participants, c'est à nouveau **DIC3** qui a conduit à des mouvements de vergence plus réalistes, suivie par **DIC1** en considérant l'écart angulaire total. Les configurations avec flou comptent quant à elles respectivement 7 et 7 participants en considérant l'écart angulaire total ou l'écart angulaire en X, contre 10 et 10 sans flou.

Comparaison avec la perception subjective des distances

De la même manière que nous l'avons fait pour la vergence, il nous semble intéressant de pouvoir comparer les notations subjectives attribuées aux configurations virtuelles avec les différences d'écarts angulaires à la cible. Le tableau 4.25 (complémenté par la figure F.6 en annexe) ne révèle aucune corrélation significative entre les notes tous critères confondus et les différences d'écart angulaire et écart angulaire en X à la cible. Néanmoins, il semble

	CONTROLE	DIC1	DIC3	FLOU	FLOUDIC1	FLOUDIC3
Écart angulaire	2,192	2,344	2,267	2,195	2,271	2,307
Écart angulaire en X	2,366	2,631	2,555	2,532	2,706	2,468
Moyenne	2,279	2,487	2,411	2,363	2,489	2,387
Écart angulaire	2	4	4	2	2	3
Écart angulaire en X	3	2	5	3	1	3
Participants	5	6	9	5	3	6

Tableau 4.24 Moyennes et dénombrement pour chaque configuration des participants ayant obtenu les meilleurs résultats en terme d'écart angulaire à la cible.

même que l'écart angulaire total ait un impact positif marginalement significatif sur les notes de distances, un individu ayant obtenu des écarts angulaires plus importants au cours du test du robot ayant mieux noté l'ensemble des configurations concernant sa perception des distances et des profondeurs. Nous expliquons ce résultat par le lien reliant ces deux facteurs à la quantité de disparités. D'un côté, des disparités plus importantes peuvent conduire, en combinaison avec des erreurs de calibration de l'affichage, à un écart plus important de la position de la cible perçue dans le monde réel avec sa position dans le monde virtuel. De l'autre, elles donnent une impression de profondeur accrue par le rendu d'une image plus « bombée », pouvant conduire à des notes supérieures dans ce critère.

	Note moyenne	Note distances
Écart angulaire	p=0,949 ↗	p= 0,181 ↗
Écart angulaire en X	p=0,978 ↘	p=0,873 ↗

Tableau 4.25 Probabilités d'indépendance entre les différences d'écarts angulaires et les notes (moyennes ou de distances), et variations des régressions linéaires associées.

Effet de l'imprécision de l'estimation du POR

Comme expliqué pour la vergence, la qualité de l'estimation du POR a joué un rôle important dans le test du robot. Le tableau 4.26 réunit les valeurs de probabilité d'indépendance entre la mesure d'imprécision et les différences d'écart à la cible par rapport à la configuration réelle, les graphiques illustrant les régressions linéaires associées pouvant être consultés en annexe (figure F.7). Nous notons un lien positif significatif entre la mesure d'imprécision et l'écart angulaire en X à la cible, indiquant que ce dernier augmente plus l'estimation du POR est imprécise. Cependant, cette tendance s'inverse en écartant les trois participants pour lesquels l'imprécision était la plus haute, avec une corrélation marginalement significative entre

une bonne précision et un écart angulaire total élevé. Néanmoins, il ne faut pas oublier que la mesure d'imprécision est effectuée immédiatement après la calibration. Or, pour expliquer les écarts angulaires en X non nuls obtenus par certains participants dans le cas de la configuration réelle, nous avons supposé que les lunettes avaient pu être légèrement déplacées au cours du reste de l'expérimentation, impactant ainsi l'imprécision de l'estimation du POR et sa valeur au moment du test du robot.

	Précision
Écart angulaire	p=0,674 ↘
Écart angulaire sans extrêmes	p= 0,113 ↘
Écart angulaire en X	p= 0,031 ↗
Écart angulaire en X sans extrêmes	p=0,285 ↘

Tableau 4.26 Probabilités d'indépendance entre la mesure d'imprécision et les différences de vergence/distance et vergence en X/distance, et variations des régressions linéaires associées.

CHAPITRE 5

DISCUSSION

Les résultats de l'expérimentation détaillés dans le chapitre précédent nous permettent de tirer certaines conclusions préliminaires quant aux hypothèses émises en section 2.6.5. Ces conclusions pourront néanmoins être nuancées dans un second temps, en abordant certaines considérations sur l'influence possible d'aspects démographiques, la possibilité d'analyses complémentaires, et en discutant des limitations théoriques et matérielles rencontrées. Nous reviendrons enfin sur l'accomplissement des objectifs de ce projet.

5.1 Vérification des hypothèses

H1 : La précision et la stabilité des mesures de positions et de vecteurs de direction de vue fournis par les ETG permettent de déterminer précisément la position du point de regard de l'individu.

Vérification : La mesure d'imprécision obtenue suite à la procédure de calibration mise au point, pour des objets situés à 2 m de l'observateur, était de 2,65 cm dans les tests réalisés en amont (voir la figure 3.14) et de 6,47 cm en moyenne sur l'ensemble des participants. Les mesures d'instabilité étaient quant à elles de 0,76 cm et 1,62 cm, ce qui valide notre hypothèse. Cependant, l'imprécision peut augmenter au cours d'une immersion prolongée, par exemple suite à un déplacement des lunettes relativement à la tête.

H2 : La précision et la stabilité des mesures de positions et de vecteurs de direction de vue fournis par les ETG permettent d'ajuster les paramètres de l'affichage stéréoscopique.

Vérification : La méthode implémentée permet donc de déterminer, à partir des données ETG, la position 3D du point de regard dans le monde virtuel avec une bonne précision. Il est ensuite possible à partir de celle-ci de choisir une distance inter-caméras et de générer un flou de profondeur de champ adaptés au point de regard de l'observateur avec la même précision. Les analyses de l'impact de cette dernière sur les résultats des différentes étapes de l'expérimentation n'ont pas révélé d'influence significative, ce qui nous amène à conclure qu'elle était suffisante pour chaque participant afin d'ajuster l'affichage sans ajouter à la gêne ressentie. Cependant, au niveau de la comparaison subjective, les modifications des paramètres proposées se sont révélées peu performantes, les participants préférant dans l'ensemble la configuration **CONTROLE** qui n'utilisait pas du tout l'information du point

de regard. Ce résultat nous amène à penser que c'est la réactivité du système pour mettre à jour l'affichage qui est en jeu, ce dont nous discuterons en 5.4.

H3 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard influe positivement sur ses impressions subjectives.

Vérification : Les notes subjectives octroyées aux configurations **DIC3** et **FLOUDIC3** ne sont pas significativement supérieures à celles obtenues par la configuration **CONTROLE**. De manière générale, un écart important entre les 3 DIC semble impacter négativement les impressions subjectives des participants, mis à part dans le critère de la perception des distances. Diminuer cet écart pourrait alors conduire à une configuration plus performante, mais au détriment des avantages que peuvent représenter la variation de la DIC en fonction de la distance au point de regard sur la restitution des distances et des profondeurs. L'hypothèse semble donc infirmée, bien qu'il pourrait être intéressant de vérifier que cette influence n'est pas liée à un temps de latence trop important, discuté en section 5.4, qui pourrait s'avérer le véritable problème pour les participants ayant moins bien noté ces configurations.

H4 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu influe positivement sur ses impressions subjectives.

Vérification : Les configurations **FLOU**, **FLOUDIC1** et **FLOUDIC3** n'ont pas obtenu de notes significativement supérieures à **CONTROLE**, les configurations sans flou étant préférées à leurs homologues avec flou dans 12 cas sur 15. Les impacts significatif et marginalement significatif de la présence de flou sur les critères de distances et d'immersion, pouvant s'expliquer par le ralentissement de la vitesse d'exploration visuelle induite par cette présence, se révèlent négatifs. Par ailleurs, le flou diminue significativement le nombre d'objets regardés, lié de manière marginalement significative au plaisir ressenti par les participants. L'hypothèse semble donc infirmée, bien que le problème de temps de latence discuté en section 5.4 ait pu biaiser l'effet de l'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu.

H5 : Les deux modifications citées précédemment, réalisées simultanément, influent positivement sur les impressions subjectives de l'individu.

Vérification : La configuration **FLOUDIC3** n'a pas obtenu de notes significativement supérieures à celles de la configuration **CONTROLE**. Elle correspond même à l'avant-dernier choix de configuration en terme de nombre de participants (13 au total sur l'ensemble des critères). Nous remarquons qu'aucune interaction entre les deux paramètres ne ressort des tests ANOVA effectués, leurs effets simultanés ne semblant donc pas synergiques.

H6 : Permettre à l'individu de choisir de manière personnelle la distance inter-caméras influe positivement sur ses impressions subjectives.

Vérification : Aucune des 4 configurations offrant la possibilité à l'individu de choisir sa DIC n'a obtenu de notes significativement meilleures que **CONTROLE** qui utilisait, elle, la DIP anatomique. Au contraire, nous observons même que **DIC1** a obtenu les plus mauvais résultats dans 8 cas sur 10, et que c'est la configuration **CONTROLE** qui décroche la meilleure moyenne générale. L'hypothèse est donc infirmée, les analyses ANOVA effectuées n'ayant par ailleurs pas révélé d'influence significative de la DIC sur les notations.

H7 : le ratio entre la distance interpupillaire de l'individu et son choix personnel de distance inter-caméras influe sur ses impressions subjectives.

Vérification : Le ratio entre la distance interpupillaire de l'individu et son choix de DIC pour les configurations **DIC1** et **FLOUDIC1** a bien un impact sur ses notes dans les critères de confort, réalisme, plaisir et perception des distances, celles-ci étant plus élevées lorsque DIP/DIC se rapproche de 1. En écartant la valeur extrême de ratio, cet impact se révèle marginalement significatif pour le second critère et significatif pour les autres, mais la présence de flou le diminue. L'hypothèse est donc vérifiée et tend à confirmer que la DIP physiologique est préférée par les participants.

H8 : L'hésitation de l'individu lors de son choix personnel de distance inter-caméras influe sur ses impressions subjectives.

Vérification : L'hésitation au cours du choix des DIC1 et DIC3 n'influence pas de manière significative les notations subjectives de l'individu. En effet, bien qu'une certaine tendance se dégage pour les configurations **DIC1** et **DIC3**, amenant à penser que les participants peu sûrs de leur choix l'ont moins bien noté, celle-ci s'inverse avec la présence de flou. L'hypothèse est donc infirmée.

H9 : La propension à l'immersion de l'individu influe positivement sur ses impressions subjectives.

Vérification : La propension à l'immersion du participant, mesurée à l'aide du QPI, influence positivement et de manière significative ses notes de réalisme au cours de la navigation. Il apparaît donc que les individus ayant des facilités à s'immerger dans un monde virtuel le trouvent plus réaliste, plus crédible, bien que leur sentiment d'immersion ne semble pas être affecté. L'hypothèse est donc vérifiée pour ce critère.

H10 : La quantité d'inspection visuelle de l'individu, en particulier le nombre et la durée des fixations et des saccades ainsi que le nombre d'objets regardés, influe positivement sur ses impressions subjectives.

Vérification : Aucun lien significatif ne relie le nombre et la durée des fixations et saccades aux notations subjectives. Cependant, nous observons que le nombre d'objets regardés a bien une influence positive sur les notes de plaisir, mais ce uniquement de façon marginalement significative. Cette hypothèse n'est donc pas vérifiée en utilisant ces mesures pour quantifier l'inspection visuelle.

H11 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard augmente l'intervalle proche de profondeur dans lequel il parvient à fusionner les images stéréoscopiques.

Vérification : Les limites de fusion atteintes dans le cas des configurations **DIC3** et **FLOU-DIC3** ne sont pas significativement plus petites qu'avec **CONTROLE**, **DIC3** ayant tout de même obtenu les deux médianes les plus basses vers l'avant et vers l'arrière, **FLOU-DIC3** arrivant seconde pour ce dernier cas. L'hypothèse semble donc infirmée. Il est tout de même important de noter que le choix de la DIC proche s'effectuait pour un objet situé à 40 cm, toutes les limites de fusion tombant sous cette distance. La DIC pour ces configurations n'était donc pas variable au cours du test des limites, mais constante et égale à ce choix qui s'était révélé proche de la DIP anatomique. À noter également que l'influence significative de la DIC révélée par le test ANOVA est liée à DIC1, les configurations l'utilisant présentant des limites de fusion significativement supérieures.

H12 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu augmente l'intervalle proche de profondeur dans lequel il parvient à fusionner les images stéréoscopiques.

Vérification : Bien que certains articles étudiés durant la revue de littérature soutiennent que la présence de flou facilite la fusion des images stéréoscopiques, celle-ci n'a pas révélé d'impact significatif sur les limites de fusion au cours de notre expérimentation. Le temps de latence ne peut être invoqué ici étant donné que le point de regard était globalement constamment évalué au niveau des objets utilisés pour déterminer les limites de fusion, le flou ne nécessitant pas de mise à jour réactive. L'hypothèse est donc infirmée, sans toutefois que nos résultats indiquent que la présence de flou rende la fusion plus difficile.

H13 : Les deux modifications citées précédemment, réalisées simultanément, augmentent l'intervalle proche de profondeur dans lequel l'individu parvient à fusionner les images stéréoscopiques.

Vérification : Bien que la configuration **FLOUDIC3** compte 10 participants ayant obtenu les meilleures limites, elle est tout de même seconde derrière **CONTROLE** qui en compte 12, ce qui réfute l'hypothèse. Une fois de plus, les deux paramètres ne semblent pas présenter d'interaction et paraissent additifs.

H14 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard conduit à des réponses de vergence oculaire plus proches de la réalité.

Vérification : Les analyses réalisées sur les différences de vergence oculaire entre configurations virtuelles et réelle montrent que **DIC3** obtient la seconde médiane la plus basse en considérant la vergence totale (respectivement troisième pour la vergence en X). De plus, en comparaison des autres valeurs de DIC, la DIC3 compte le nombre maximum de participants dont les mouvements oculaires se sont le plus rapprochés du réel, avec 9 pour la vergence totale et 8 en X. Cependant, les différences de vergence induites ne sont pas significativement plus faibles qu'avec **CONTROLE**, ce résultat pouvant être dû au temps de latence.

H15 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu conduit à des réponses de vergence oculaire plus proches de la réalité.

Vérification : L'ajout d'un flou de profondeur de champ n'a pas d'influence significative sur les réponses de vergence oculaire, ni en négatif ni en positif. L'hypothèse est donc infirmée, mais nous pensons une fois de plus que le temps de latence trop important dans la mise à jour de l'affichage peut être en cause. En effet, malgré un impact éventuellement positif de la présence de flou, une latence notable peut perturber l'observateur et adjoindre un effet négatif qui annule le premier.

H16 : Les deux modifications citées précédemment, réalisées simultanément, conduisent à des réponses de vergence oculaire plus proches de la réalité.

Vérification : Bien que la DIC3 semble conduire à des mouvements de vergence assez réalistes, l'ajout de flou paraît diminuer cette influence. En effet, la configuration **DIC3** obtient 11 participants (en réunissant vergence et vergence en X) contre seulement 6 pour **FLOUDIC3**, l'effet simultané des deux paramètres étant une fois de plus non synergique. L'hypothèse est donc infirmée.

H17 : La variation de la distance inter-caméras pour s'ajuster à la distance entre l'individu et son point de regard conduit à un écart angulaire à la cible plus proche de la réalité.

Vérification : La configuration **DIC3** est la configuration comptant le nombre maximum de participants ayant présenté les écarts angulaires à la cible les plus réalistes. Sa médiane est la plus basse en considérant l'écart angulaire total, celle de **FLOUDIC3** arrivant seconde pour l'écart en X. Cependant, leurs résultats ne sont pas significativement meilleurs que ceux de la configuration **CONTROLE**, ce qui semble infirmer l'hypothèse. Nous pensons cependant que des erreurs au niveau de l'affichage stéréoscopique, qui seront discutées en section 5.4.2, peuvent avoir induit un biais dans les résultats liés à l'écart angulaire à la cible.

H18 : L'ajout d'un flou de profondeur de champ ajusté à la position du point de regard de l'individu conduit à un écart angulaire à la cible plus proche de la réalité.

Vérification : Les tests ANOVA effectués ne révèlent pas d'influence significative de l'ajout d'un flou de profondeur de champ sur le réalisme des écarts angulaires. L'hypothèse est donc infirmée, bien que le temps de latence ait pu une fois de plus biaiser les résultats, en contrebalançant par un effet négatif l'effet éventuellement positif de l'ajout de flou.

H19 : Les deux modifications citées précédemment, réalisées simultanément, conduisent à un écart angulaire à la cible plus proche de la réalité.

Vérification : Les conclusions quant à cette hypothèse sont similaires à celles de **H16** : la distance inter-caméras variable DIC3 semble conduire à des réponses oculaires aussi ou plus réalistes que les autres configurations, mais l'ajout de flou ne paraît pas souhaitable. En effet, **DIC3** obtient 9 participants (en réunissant écart angulaire et écart angulaire en X) contre seulement 6 pour **FLOUDIC3**. L'hypothèse est donc infirmée.

La plupart de nos hypothèses ne semblent donc pas se vérifier expérimentalement. Cependant, certaines remarques présentées ci-dessous peuvent selon nous apporter un éclairage un peu différent sur l'interprétation des analyses et nuancer l'infirmerie des hypothèses.

5.2 Groupement des participants

Les résultats présentés dans le chapitre précédent prenaient la plupart du temps en compte tous les participants, la mise à l'écart de quelques-uns s'effectuant uniquement en raison de différences importantes de leurs données par rapport à la moyenne. Cependant, nous pouvons distinguer certains groupes définis sur des critères démographiques qui s'avèrent présenter des particularités.

5.2.1 Port de lentilles correctrices

Bien que les spécifications des ETG indiquent qu'elles fonctionnent de manière identique en leur présence, nous remarquons que le port ou non de lentilles correctrices a une conséquence significative sur l'imprécision de la détermination du point de regard ($F=13,43$ et $p=0,002$) et son instabilité ($F=7,87$ et $p=0,013$), tel que nous pouvons l'observer en figure 5.1. En particulier, les mesures de qualité de repérage communiquées par les ETG présentent des moyennes différentes pour les participants avec et sans lentilles (voir le tableau 5.1).

	Pupille gauche	Pupille droite	Globe oculaire gauche	Globe oculaire droit
Sans lentilles	8,67	8,57	-3,26	-3,27
Avec lentilles en X	8,23	8,09	-3,12	-2,75

Tableau 5.1 Mesures ETG d'incertitude du repérage des yeux, pour les participants avec et sans lentilles correctrices.

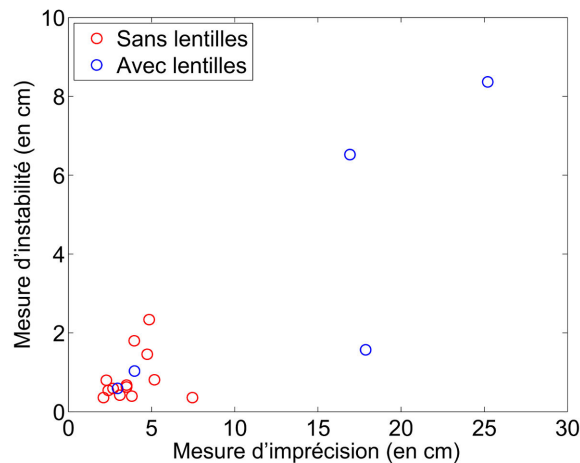


Figure 5.1 Mesures d'imprécision et d'instabilité pour les participants avec et sans lentilles correctrices.

N'ayant pas remarqué de lien clair entre la mesure d'imprécision et les notes subjectives ou les critères objectifs de comparaison des configurations, nous pensons que celle-ci est généralement suffisamment petite et que le port de lentilles correctrices n'influencera pas de manière gênante l'expérience immersive des futurs utilisateurs de la voûte. Néanmoins, il apparaît que les résultats des deux groupes de participants, ceux portant des lentilles et ceux n'en portant pas, sont significativement différents en ce qui concerne les notes subjectives et les limites de fusion, tel que le présente le tableau 5.2 qui regroupe les résultats des tests de Student appariés effectués. Nous avons donc envisagé que nous pourrions obtenir des résultats

plus significatifs quant à l'impact des configurations, de la DIC ou de la présence de flou sur les notes subjectives et sur les limites de fusion en écartant des analyses les 5 participants porteurs de lentilles. Les tests ANOVA à 1 et 2 facteurs contrôlés, dont les résultats sont regroupés dans le tableau 5.3, ne se distinguent cependant pas des précédents, si ce n'est que le flou influence négativement de façon plus importante la perception des distances et le sentiment d'immersion.

	Notes	Limites	Vergence	Écart angulaire
p	0,018	0,007	0,395	0,311

Tableau 5.2 Test de Student apparié comparant le groupe de participants portant des lentilles à celui n'en portant pas.

Critère	Configuration	DIC	Flou	Interaction
Confort	F=0,66 p=0,657	F=0,83 p=0,440	F=1,32 p=0,255	F=0,15 p=0,858
Réalisme	F=0,39 p=0,853	F=0,19 p=0,831	F=1,22 p=0,273	F=0,18 p=0,832
Plaisir	F=0,42 p=0,830	F=0,31 p=0,736	F=1,33 p=0,253	F=0,09 p=0,914
Distances	F=1,52 p= 0,194	F=0,31 p=0,734	F=6,66 p= 0,012	F=0,16 p=0,853
Immersion	F=0,77 p=0,575	F=0,05 p=0,954	F=3,64 p= 0,061	F=0,06 p=0,946
Vers l'avant	F=0,54 p=0,748	F=1,31 p=0,276	F=0,01 p=0,917	F=0,02 p=0,978
Vers l'arrière	F=0,80 p=0,552	F=1,98 p= 0,145	F=0,04 p=0,851	F=0,00 p=0,996

Tableau 5.3 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives et les limites de fusion, en écartant les 5 participants portant des lentilles de contact.

5.2.2 Genre du participant

Le genre du participant peut également avoir une influence sur ses impressions subjectives en raison des différences existantes entre les habilités visio-spatiales des individus féminins et masculins (Baril, 2000). Les 18 volontaires ayant participé à l'expérimentation se divisaient en 3 femmes et 15 hommes. Nous remarquons sur les résultats des tests de Student appariés, regroupés dans le tableau 5.4, que ces deux groupes se révèlent significativement différents, non pas au niveau des notations subjectives comme nous nous y attendions mais pour les 3 autres critères de comparaisons des configurations virtuelles. Cependant, les analyses ANOVA à 1 et 2 facteurs contrôlés réalisés en écartant les individus de genre féminin ne permettent pas de faire ressortir des impacts plus significatifs de la DIC ou de la présence de flou de

profondeur de champ (voir le tableau 5.5). Au contraire, dans le cas des limites de fusion par exemple, l'influence de la DIC passe de significative à marginalement significative, ce qui s'explique par le fait que l'une des participantes était celle dont les limites étaient bien supérieures aux autres.

Cependant, il ne nous semble pas étonnant qu'un aussi petit groupe (3 individus) présente des résultats significativement différents – en particulier lorsqu'il intègre un participant dont les données sont très divergentes de la moyenne – sans que son éviction n'affecte les résultats généraux.

	Notes	Limites	Vergence	Angle
p	0,846	6.099e-08	2.501e-07	1.428e-05

Tableau 5.4 Test de Student comparant le groupe de participants de genre féminin et de genre masculin.

Critère	Configuration	DIC	Flou	Interaction
Vers l'avant	F=5,88 p=0,000	F=14,6 p=0,000	F=0,09 p=0,761	F=0,05 p=0,953
Vers l'arrière	F=5,49 p=0,000	F=13,5 p=0,000	F=0,26 p=0,614	F=0,06 p=0,937
Vergence	F=0,17 p=0,974	F=0,26 p=0,773	F=0,00 p=0,977	F=0,16 p=0,854
Vergence en X	F=0,30 p=0,914	F=0,66 p=0,521	F=0,03 p=0,856	F=0,07 p=0,936
Écart	F=0,05 p=0,998	F=0,07 p=0,933	F=0,05 p=0,831	F=0,03 p=0,967
Écart en X	F=0,26 p=0,935	F=0,36 p=0,699	F=0,48 p=0,489	F=0,04 p=0,957

Tableau 5.5 ANOVA à 1 et 2 facteurs contrôlés : effets de la configuration et de chacun des deux paramètres sur les notations subjectives et les limites de fusion, en écartant les 3 participantes.

5.2.3 Carrés latins

Comme expliqué à la fin du chapitre 3 (voir section 3.4.7), les configurations sont ordonnancées durant les différentes étapes de l'expérimentation de manière à limiter les biais induits par l'augmentation de la fatigue visuelle et de la familiarité avec l'environnement, et les participants répartis en groupes selon le principe des carrés latins. Chaque groupe comptait 3 participants qui visualisaient le même ordonnancement.

Les résultats des tests ANOVA à 2 facteurs contrôlés, réunis dans le tableau 5.6, indiquent que les groupes apparaissent significativement différents pour la plupart des critères de comparaison, mais qu'aucune interaction n'intervient avec les configurations stéréoscopiques. Ainsi, ces groupes divergent dans les notes octroyées ou les limites atteintes, sans pour autant préférer significativement une configuration.

Nous supposons que les différences sont liées à d'autres facteurs que celui de l'ordonnement, tels que le petit effectif des groupes ou leur hétérogénéité (certains comportant des participants portant des lentilles ou de genre féminin, d'autres non). Cependant, nous pouvons observer sur la figure 5.2 que de manière singulière les minimums de limites de fusion atteints par tous les participants des groupes 2 et 3 sont plus élevés que ceux des autres groupes. Une étude plus détaillée des séquences d'ordonnement des configurations pour chaque groupe pourrait apporter des informations sur l'influence de l'ordre de passage des configurations ou de leur alternance quant à leurs performances.

Critère	Groupe	Configuration	Interaction
Confort	F=2,38 p=0,047	F=0,79 p=0,563	F=0,39 p=0,995
Réalisme	F=4,61 p=0,001	F=0,37 p=0,870	F=0,26 p=1,000
Plaisir	F=0,56 p=0,732	F=0,51 p=0,769	F=0,37 p=0,997
Distances	F=7,23 p=0,000	F=1,56 p=0,181	F=0,70 p=0,838
Immersion	F=1,88 p=0,109	F=0,62 p=0,683	F=0,29 p=1,000
Vers l'avant	F=12,2 p=0,000	F=2,12 p=0,073	F=0,18 p=1,000
Vers l'arrière	F=18,6 p=0,000	F=3,17 p=0,012	F=0,15 p=1,000
Vergence	F=3,29 p=0,010	F=0,12 p=0,989	F=0,24 p=1,000
Vergence en X	F=4,82 p=0,001	F=0,22 p=0,954	F=0,14 p=1,000

Tableau 5.6 ANOVA à 1 et 2 facteurs contrôlés : effets de l'appartenance à un groupe et de la configuration sur les notations subjectives, les limites de fusion et les différences de vergence totale et en X.

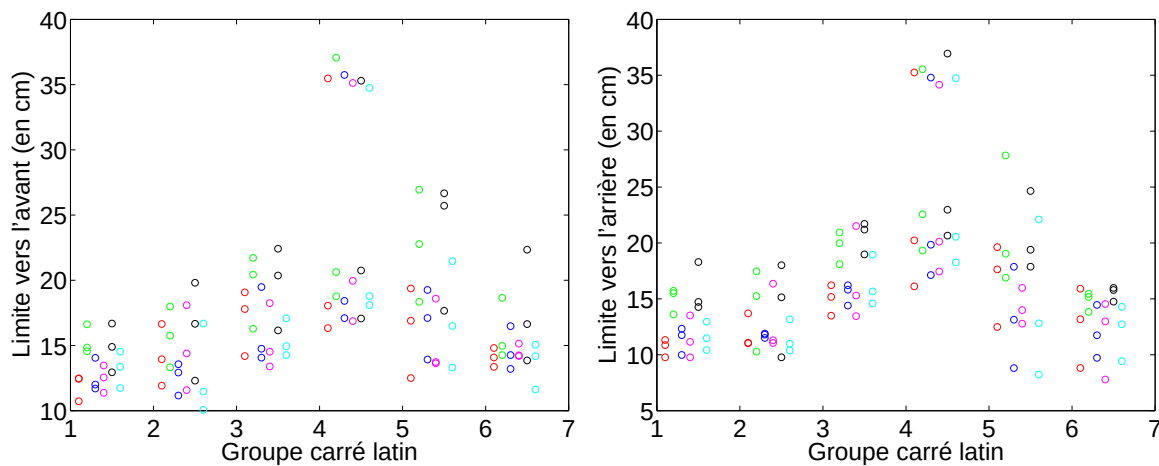


Figure 5.2 Limites de fusion atteintes par les 6 groupes de carrés latins.

5.3 Autres analyses envisageables

L'expérimentation a fourni une quantité très importante de données. À titre d'exemple, les 6 navigations de 90 secondes représentent pour chaque participant 32 400 rapports ETG, et en moyenne 6 026 positions de point de regard calculées dans Unity (une à chaque image). Nos analyses ont majoritairement consisté à extraire et étudier des valeurs statistiques, mais d'autres plus minutieuses sont envisageables, qui pourraient permettre à la fois d'expliquer le peu de résultats significatifs obtenus et éventuellement d'en faire émerger certains en révélant des facteurs dissimulés.

5.3.1 Étude au cas par cas

Comme cité précédemment, les résultats de chaque groupe des carrés latins peuvent être examinés indépendamment, en s'attardant sur l'ordre des configurations les unes par rapport aux autres. Par exemple, Duchowski et al. (2011) ont observé au cours de leur expérimentation que lorsque l'affichage stéréoscopique était suivi du monoscopique, ce dernier était étrangement toujours considéré comme quelque peu stéréoscopique, profitant sans doute d'une répercussion de la perception de la profondeur occasionnée par le premier. Les configurations à disparité plus faibles, comme **CONTROLE**, pourraient alors profiter du même effet secondaire en étant par exemple précédées de **DIC1** ou **FLOUDIC1**.

Les données des autres groupes proposés dans la section précédente pourraient également être considérées plus attentivement, en réalisant et étudiant une à une des boîtes à moustache illustrant les résultats obtenus par chaque configuration. Celles présentées en figure 5.3 révèlent par exemple que les notations sur le critère de perception des distances présentent un schéma particulier pour les individus de genre féminin, avec une préférence aux configurations sans flou (cette tendance pourrait néanmoins s'atténuer avec un nombre plus important de femmes). D'autres classes pourraient être envisagées, regroupant les participants par âge, familiarité avec la stéréoscopie, notes de propension à l'immersion, etc. Nous sommes tentés de croire qu'à partir des résultats étudiés participant par participant, il serait possible de trouver des tendances communes entre certains et de remonter jusqu'au trait démographique pouvant en être à l'origine.

En effet, il n'existe peut-être pas de configuration optimale pour tous les participants, mais une configuration pourrait par exemple convenir aux femmes de plus de 30 ans, une autre aux hommes portant des lentilles de contact. Pour arriver à de telles conclusions, il s'avérerait néanmoins bénéfique de réaliser la même expérimentation sur un nombre plus important de sujets, afin d'obtenir des groupes démographiques de tailles conséquentes.

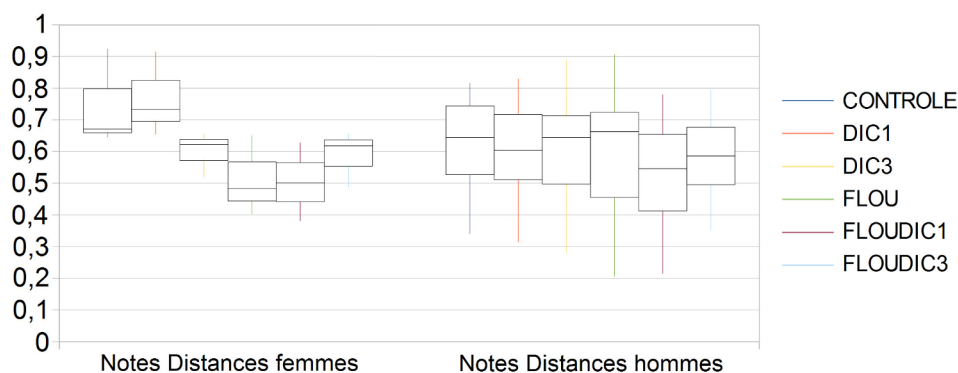


Figure 5.3 Notes données par les femmes et les hommes dans le critère de perception des distances.

5.3.2 Étude d'informations peu exploitées

Certaines informations enregistrées n'ont pas été exploitées à leur plein potentiel, telles que la distance entre l'observateur et son point de regard. Il serait par exemple intéressant d'étudier l'impact de la configuration et de l'ajustement des paramètres sur la distance moyenne à l'objet regardé, ainsi que sur la variation totale de celle-ci. Cette dernière, liée aux allées et venues en profondeur du point de regard de l'observateur, peut avoir un lien avec les notations subjectives attribuées, avec une double implication possible : lorsqu'une configuration est moins confortable en termes d'indices de profondeur, l'observateur aura tendance à tâtonner visuellement et à faire varier la vergence de ses yeux avec une amplitude plus importante que si la configuration est confortable. Inversement, concentrer son attention alternativement sur des objets situés à des profondeurs différentes peut conduire à de l'inconfort et à de moins bonnes notes dans ce critère.

D'autres moments du signal visuel pourraient être examinés. Nous avons par exemple considéré l'angle de vergence et de vergence en X en fonction du temps et en fonction de la distance à la cible sur le robot, mais la vitesse de variation ou une analyse fréquentielle de ces vergences pourraient révéler des schémas distinguant plus assurément les configurations virtuelles, et éventuellement faisant émerger une similitude forte de certaines avec la configuration réelle.

5.3.3 Analyse qualitative de l'attention visuelle

Les mesures d'inspection visuelle employées pour établir des liens avec les appréciations subjectives des participants étaient surtout quantitatives : nombre et durée moyenne des fixations, nombre d'objets regardés, etc. Il pourrait être intéressant de réaliser une analyse qualitative complémentaire, apportant des informations sur la nature des objets regardés. Les

caractéristiques d’une configuration ont pu par exemple pousser les participants à s’attarder sur certains objets, tels que ceux présents dans le champ visuel lointain lorsque la DIC est importante. Inversement, le fait que certaines régions de la scène aient été scrutées plus attentivement a pu orienter les notes attribuées, en particulier dans les critères de réalisme, plaisir et immersion.

Cette analyse se baserait naturellement sur les cartes de densité, les chemins visuels, ainsi que les informations statistiques spécifiques à certains objets ou sous-objets d’intérêt identifiés, mais requerrait des connaissances psychologiques sur l’inspection visuelle naturelle d’une scène. Elle se révélera nécessaire et particulièrement intéressante pour les cliniciens de l’Institut Philippe-Pinel de Montréal lorsque la voûte sera employée à des fins thérapeutiques.

5.4 Limitations de la solution proposée

5.4.1 Limitations théoriques

Au cours de la comparaison expérimentale des configurations, nos propositions d’ajustement des paramètres stéréoscopiques réunies dans la configuration **FLOUDIC3** ne se sont pas révélées plus performantes. Nous émettons plusieurs suppositions qui pourraient expliquer cela sans remettre en cause le potentiel de tels ajustements. Tout d’abord, concernant la variation de la DIC, nous effectuons une interpolation linéaire à partir de 3 valeurs pour 3 profondeurs différentes, ce qui induit une rupture brutale au niveau de la seconde, qui pour la plupart des participants représente le maximum. Une solution pourrait alors consister à réaliser une interpolation quadratique. De plus, nous avons décidé de laisser chaque participant choisir ses DIC, mais les choix subjectifs réalisés ne se sont pas avérés les plus pertinents. Nous avons en particulier remarqué que les configurations **DIC1** et **FLOUDIC1** ont obtenu de mauvais résultats. Nous pensons qu’il pourrait être préférable de comparer un ensemble de DIC fixes et de sélectionner celles obtenant les meilleurs résultats ou de calculer mathématiquement celles qui rendraient la profondeur de la manière la plus réaliste comme le propose Jones et al. (2001).

Ensuite, nous n’avons testé qu’une intensité de flou en raison du nombre déjà important de configurations à comparer, sélectionnée en fonction de nos impressions et de celles d’un participant préliminaire sur la quantité réelle d’un flou de profondeur de champ. Or, ce choix ne convenait peut-être pas à la majorité de nos participants. En effet, le fait que seuls certains sujets de l’expérimentation menée par Hillaire et al. (2008b) aient perçu le flou comme « trop fort » nous laisse supposer que l’intensité optimale dépend de l’individu. Il serait alors intéressant de laisser chaque participant choisir son intensité de flou de profondeur de champ, exagéré ou au contraire diminuée par rapport à un flou « réel », ou encore, afin d’éviter les

déconvenues rencontrées avec DIC1, d'effectuer une expérimentation testant plusieurs valeurs de flou et sélectionner celle préférée par le plus grand nombre d'individus.

Enfin, notre solution calcule le point de regard en utilisant le *raycasting*, méthode que nous avons choisie après avoir testé celle d'intersection des lignes de vue et constaté qu'elle fonctionnait mal avec les données fournies par les ETG. Or, afin d'améliorer la précision, nous avons mis au point une méthode de calibration que nous n'avons pas testée *a posteriori* avec la première méthode, et qui pourrait s'avérer fournir de bons résultats – bien que nous pensons que l'instabilité des vecteurs conduira à une position bien plus instable du point de regard que celle obtenue en intersectant la scène virtuelle.

5.4.2 Matériel et implémentation

L'expérimentation a également révélé des problèmes qui ne sont pas liés à la théorie, mais au matériel utilisé et à l'implémentation que nous avons réalisée de la solution que nous proposons.

Tout d'abord, malgré le fait que la fréquence des lunettes ETG, 60 Hz, soit suffisante par rapport aux caractéristiques de la vision temporelle humaine, nous observons que le nombre d'images calculées par Unity est beaucoup plus faible (12,96 en moyenne durant la navigation pour les configurations sans flou et 9,38 pour celles avec flou). Or, dans notre cas, l'application est à la fois diagnostique – par l'enregistrement pour analyse ultérieure des positions du point de regard, et interactive – par l'ajustement en temps réel des paramètres stéréoscopiques en fonction de ces positions (voir section 2.2.5). Une des préoccupations de ce dernier type d'application concerne la possibilité de suivre le regard et de modifier l'affichage suffisamment rapidement pour que l'observateur ne détecte pas de temps de latence. En effet, si la distance inter-caméras ou le flou de profondeur de champ ne sont pas mis à jour assez vite, ils peuvent influencer le sujet ou perturber son immersion. La présence de flou peut par exemple l'inciter à regarder la partie de l'environnement restée nette après un mouvement de saccade pour lequel l'affichage ne s'est pas encore mis à jour. Nous pensons en particulier que ce phénomène de retard s'avère plus néfaste pour le sentiment de présence que le manque de réalisme de la quantité de disparités ou du flou de profondeur de champ.

Le fait que nous n'ayons pas réalisé de mesure du délai de mise à jour représente une limite au niveau de l'interprétation des résultats. En effet, bien que nous ayons utilisé le rapport ETG le plus récent dans Unity pour le calcul du nouveau POR, et ce afin de limiter ce délai et la distance entre le POR réel et celui utilisé pour l'ajustement des paramètres, nous ne pouvons pas affirmer qu'ils étaient suffisamment petits pour ne pas être perceptibles. Par ailleurs, des discussions informelles avec les participants ont révélé que certains d'entre eux remarquaient la présence de flou par le ralentissement de l'affichage. Bien qu'ils ne l'aient

pas stipulé, nous supposons que cela a pu impacter négativement leur impression de réalisme et leur sentiment d’immersion.

Nous pensons que certains modules de l’implémentation pourraient être optimisés en terme de temps de calcul, ce qui aurait pour conséquences de diminuer la latence de mise à jour et d’augmenter le nombre d’images par seconde pour un rendu plus fluide de l’univers virtuel. Nous avons en particulier utilisé des méthodes Unity pour la génération du flou de profondeur de champ et le calcul du *raycasting*, dont nous ne connaissons ni le fonctionnement ni la complexité en temps. Concernant le premier, un script de génération de flou *ad hoc* basé sur l’une des deux méthodes décrites en section 2.4.1 pourrait s’avérer plus rapide, quitte à perdre en réalisme. La méthode de *raycasting* nécessite quant à elle que chaque objet dispose d’un *mesh collider*, ce qui rend la scène plus lourde, et il est possible qu’elle recherche une intersection avec tous les triangles de l’environnement. Une partition de l’espace tridimensionnel à l’aide d’*octree* ou une hiérarchie de la scène sous forme d’arbre binaire de sphères englobantes pourrait permettre d’accélérer le calcul de l’intersection.

Des problèmes au niveau de l’affichage stéréoscopique dans la voûte se sont également manifestés au cours de l’expérimentation. Tout d’abord, les 8 projecteurs ne parvenaient pas correctement à se synchroniser, certains passant en opposition de phase par rapport aux autres. Ainsi, une partie affichait l’image destinée à l’œil gauche au moment où l’autre affichait l’image destinée à l’œil droit. Relancer le programme à plusieurs reprises a parfois permis de contourner ce problème, mais il est intervenu pour la plupart des participants. Il arrivait même que la synchronisation change à plusieurs reprises au cours de l’expérimentation. Bien que l’inversion droite/gauche ne supprime pas la perception des profondeurs, mais conduise plutôt à une 3D gênante et peu réaliste, ces erreurs de synchronisation s’avéraient particulièrement problématiques lorsqu’elles affectaient les 2 projecteurs d’un même écran, en particulier l’écran de devant sur lequel les participants portaient leur attention la majorité du temps. Une bande horizontale au centre de l’écran, sur laquelle se superposaient les images du haut et du bas, était alors perçue en double. Ce problème a pu grandement jouer sur les notations subjectives, en particulier lorsqu’il n’a affecté qu’une partie des navigations, mais n’a cependant pas eu d’impact majeur au niveau des critères objectifs. En effet, durant le test des limites de fusion les participants se concentraient principalement sur la moitié haute de l’écran, et sur la moitié basse au cours du test du robot.

Les valeurs d’écarts angulaires à la cible enregistrées nous amènent par ailleurs à penser que la calibration de l’affichage ne permet pas de restituer fidèlement la géométrie de la scène virtuelle, en particulier la position des objets. Nous remarquons en effet des écarts beaucoup plus importants pour les configurations virtuelles, indiquant que le robot était af-

fiché de manière décalée par rapport à sa position dans la scène. Ainsi, malgré un calcul des images stéréoscopiques correct ou même optimal, des petites erreurs dans le positionnement des projecteurs ou de correction des projections au cours du mélange¹ peuvent conduire elles aussi à une déformation du repère virtuel par rapport au repère réel. Cette dernière peut grandement augmenter l'imprécision de la détermination du point de regard. En effet, soient deux objets o_0 et o_1 situés dans le monde virtuel aux positions (x_0, y_0, z_0) et (x_1, y_1, z_1) . Si o_0 est affiché sur l'écran à la position (x_1, y_1, z_1) dans le repère réel, c'est alors o_1 qui sera intersecté au moment du *raycasting* et considéré comme l'objet d'attention. La calibration de l'affichage, sur laquelle nous ne nous sommes pas penchés au cours de ce projet, s'avère donc essentielle, bien que la méthode de calibration de la détermination du point de regard utilisée permette en partie de contrebalancer les problèmes d'affichage, en effectuant une conversion du repère réel au repère virtuel lors de la correction des vecteurs de direction de vue.

Ainsi, nous pensons que les principales raisons qui expliquent les résultats obtenus et l'infirmité de la plupart de nos hypothèses ne sont pas liées à la théorie sous-tendant nos propositions, mais aux limitations matérielles. Les problèmes de synchronisation et de rapidité d'ajustement que nous venons d'évoquer peuvent expliquer en particulier qu'aucune configuration ne se soit réellement démarquée des autres en terme de notations subjectives. Les problèmes de calibration de l'affichage peuvent quant à eux être à l'origine de la différence des mouvements oculaires entre configurations virtuelles et réelle, qui est bien plus importante que celle présente entre les différentes configurations virtuelles.

5.5 Synthèse quant aux objectifs

L'objectif de ce projet comportait deux aspects principaux : la conception d'une méthode d'ajustement des paramètres stéréoscopiques et son implémentation logicielle, puis son évaluation par comparaison avec d'autres configurations au cours d'une expérimentation avec participants. Afin d'accomplir cet objectif général, nous l'avons décomposé en plusieurs étapes spécifiques.

Les mesures de la direction du regard et de la vergence oculaire ont été réalisées à l'aide d'un dispositif de suivi binoculaire, les lunettes ETG. Satisfaire cet objectif a nécessité que nous intégrions ces lunettes à la voûte de réalité virtuelle de l'Institut Philippe-Pinel, en particulier en développant une librairie dynamique et un module d'extension Unity. Ces derniers permettent de récupérer depuis Unity les informations de suivi telles que la position des yeux et les directions de regard qui fournissent l'angle de vergence.

1. En anglais *blending*.

Nous avons ensuite développé un script qui détermine le point de regard 3D en ayant recours à une méthode de *raycasting*, et permet également de réaliser une calibration basée sur les multi-quadriques de Hardy pour améliorer la précision de la position.

Après avoir conçu théoriquement notre méthode de modification de l’affichage, **FLOU-DIC3**, qui utiliserait la génération d’un flou de profondeur de champ et l’interpolation de la distance inter-caméra en fonction de la distance au point de regard, nous en avons réalisé l’implémentation.

Afin de l’évaluer, nous avons décidé de la comparer à d’autres configurations, dont les paramètres variables sont la DIC et la présence ou l’absence d’un flou de profondeur de champ, au cours d’une expérimentation faisant intervenir des participants. Leurs appréciations, recueillies à l’aide d’un questionnaire, nous ont permis d’évaluer subjectivement notre méthode sur les critères de confort visuel, réalisme du rendu, plaisir de navigation, perception des distances et des profondeurs et sentiment de présence dans le monde virtuel.

La détermination des limites de fusion atteintes avec chaque configuration nous a permis de comparer sur un critère plus objectif la performance des différentes méthodes. Nous avons également mis au point un test inédit, en recréant virtuellement une scène réelle afin de comparer les mouvements de vergence et les écarts angulaires à la cible provoqués dans chacun des cas.

Enfin, les outils développés pour extraire les éléments statistiques et autres informations pertinentes de l’ensemble des données acquises au cours de l’expérimentation pourront faciliter l’analyse et l’interprétation des études futures menées par les cliniciens de l’Institut Philippe-Pinel. Ils permettent en particulier de calculer la vergence de l’observateur, d’extraire des mesures de ligne de visée en identifiant des objets ou sous-objets spécifiques, ainsi que des données statistiques sur la séquence de fixations et saccades oculaires. Les scripts de génération de cartes de densité et de chemins visuels fournissent les représentations graphiques de l’attention visuelle de l’individu, également vouées à être utilisées à des fins de diagnostic et de traitement.

CHAPITRE 6

CONCLUSION

6.1 Synthèse des travaux

Ce projet est né de la volonté d'améliorer les performances du rendu stéréoscopique dans une voûte de réalité virtuelle, en particulier en réduisant le stress visuel engendré par une exposition prolongée et en optimisant le réalisme pour un sentiment de présence accru. Nous avons décidé tout d'abord d'utiliser la profondeur du point de regard de l'observateur afin d'ajuster la distance inter-caméras, la faisant varier entre différentes valeurs définies pour 3 profondeurs (de 0,4 m à 2,5 m). Le principe est ici de jouer sur la quantité de disparité affichée pour rendre de la manière la plus juste possible la profondeur de la scène en fonction de celle à laquelle l'observateur porte son attention. La position du point de regard a quant à elle permis de générer un flou de profondeur de champ proche de celui qui intervient dans la vision naturelle d'un monde en relief.

6.1.1 Conception de l'expérimentation

Après avoir développé les outils permettant d'intégrer un système de suivi oculaire, les lunettes ETG, à la voûte de réalité virtuelle de l'Institut Philippe-Pinel de Montréal, nous avons défini un protocole expérimental afin d'évaluer et de comparer la configuration stéréoscopique résultant de nos propositions, **FLOUDIC3**, avec 5 autres. Ces 6 configurations se distinguaient par la présence ou l'absence du flou de profondeur de champ et par la valeur de la DIC : DIP, DIC1 ou DIC3. L'expérimentation réalisée auprès de 18 volontaires leur demandait d'effectuer certaines tâches, chacune associée à un objectif de recherche. Une navigation dans un environnement virtuel riche, suivie de la collecte des appréciations du participant, a permis dans un premier temps d'évaluer sur des critères subjectifs la perception de la 3D avec les différentes configurations. Ces critères comprenaient le confort visuel, le réalisme du rendu, le plaisir de navigation, la perception des distances et des profondeurs et enfin le sentiment d'immersion. La détermination des limites de fusion des images stéréoscopiques a quant à elle fourni un indice objectif de performance des configurations, la diplopie générée par des difficultés de fusion étant l'une des sources principales d'inconfort et un intervalle de fusion plus large offrant la possibilité d'utiliser un panel plus varié de scènes virtuelles. La dernière étape du protocole consistait à comparer le comportement oculaire entre une scène réelle et sa reproduction en virtuel. Il s'agissait en particulier de mesurer les différences de

vergence oculaire et d'écart angulaire à la cible occasionnées par le suivi visuel d'une cible mouvante. Ce dernier test, qui à notre connaissance n'a jamais été réalisé, donne des indications objectives sur le rendu des distances et des profondeurs par une configuration. En effet, grâce aux valeurs de vergence, nous pouvons connaître les erreurs de jugement de distances causées par l'affichage stéréoscopique (sous-évaluation ou surévaluation de celles-ci), l'écart angulaire indiquant quant à lui où l'observateur percevait l'objet dans le monde réel et révélant des différences entre les positions dans les repères réel et virtuel.

6.1.2 Analyse des résultats

Les analyses réalisées sur les résultats et présentées dans le chapitre 5 mettent en évidence les influences des paramètres testés. Concernant les avis subjectifs des participants, la variation de la DIC en fonction de la distance au point de regard est très généralement préférée au choix fixe d'une DIC, mais pas à la DIP anatomique. La présence de flou conduit quant à elle à de moins bonnes notes, bien que cela ne soit pas significatif mis à part dans le critère de la perception des distances et des profondeurs, et à un ralentissement de l'inspection visuelle avec des fixations et des saccades plus longues. Du point de vue de l'influence sur les limites de fusion, à nouveau DIC3 est plus appréciée que le choix fixe DIC1 – ce dernier conduisant à des difficultés de fusion significativement plus importantes – mais ne se distingue pas de la DIP. Ceci s'explique par le fait qu'au cours de ce test la DIC n'était pas vraiment variable mais égale au choix réalisé pour un objet situé à 0,4 m, choix s'avérant très proche de la DIP. Le flou de profondeur de champ n'a quant à lui pas influé sur la valeur des limites, bien que l'étude de la littérature nous ait amenés à penser que sa présence facilitait la fusion des images stéréoscopiques. Ce constat s'explique selon nous par le fait que pour les limites de fusion atteintes, les objets utilisés avaient une taille très importante dans l'écran, l'arrière-plan flou se retrouvant donc minoritaire. Le troisième critère de comparaison portait sur les différences de comportement oculaire des observateurs par rapport au réel, en particulier les variations d'angles de vergence et d'écarts angulaires à la cible. Bien qu'en termes de médiane des données aucune configuration ne se démarque significativement des autres, celles employant l'ajustement en temps réel DIC3 comptent le plus grand nombre de participants ayant obtenu les mouvements oculaires les plus réalistes. Le flou n'a quant à lui pas eu d'impact notable sur les résultats de ce test.

Grâce à ces analyses, nous pouvons conclure qu'en prenant en considération l'ensemble de l'expérimentation c'est la configuration **CONTROL**, pour laquelle les caméras virtuelles sont espacées de la DIP anatomique et qui ne présente pas de flou de profondeur de champ, qui a obtenu les meilleurs résultats. Elle reçoit en effet la moyenne de notes subjectives la plus élevée, les moyennes les plus basses en termes de différences de vergence et d'écart

angulaire par rapport à la configuration réelle et la troisième plus basse concernant les limites de fusion. Si ces résultats ne corroborent pas nos hypothèses, nous pensons cependant que les limitations techniques rencontrées au cours de l'expérimentation et détaillées en section 5.4 ont pu contrecarrer les bénéfices de l'ajustement en temps réel des paramètres en fonction du point de regard. Nous proposons en section 6.2 des solutions pour écarter ce facteur.

6.1.3 Applications de la recherche

De par le contexte de ce projet de recherche en collaboration avec l'Institut Philippe-Pinel de Montréal, les applications pratiques à court terme concernent les futures expérimentations menées par les cliniciens de cet institut. Tout d'abord, notre intégration des lunettes ETG et la procédure de calibration mise en place permettent d'obtenir une bonne qualité de l'estimation du point de regard, avec une mesure d'imprécision de 2,65 cm au cours des tests préliminaires et de 6,47 cm en moyenne sur l'ensemble des participants pour des cibles situées à 2 m (ce qui correspond à $0,76^\circ$ et $1,85^\circ$ d'angle visuel). Les cliniciens pourront ainsi connaître précisément et en temps réel l'objet sur lequel le patient porte son regard. De plus, les outils que nous avons développés afin d'analyser les résultats, permettant par exemple de connaître l'écart angulaire aux objets et sous-objets identifiés, le nombre et la durée des fixations sur ces objets, ou encore de visualiser le chemin visuel sous forme de carte de densité, faciliteront leur interprétation des comportements oculaires en termes d'attention. Enfin, en permettant de conclure sur la configuration qui, dans l'installation actuelle de la voûte et parmi les 6 testées, est la plus appréciée subjectivement et conduit en moyenne à un comportement oculaire plus proche de celui exprimé dans le cas réel, ce projet permettra aux cliniciens d'exploiter plus efficacement les capacités de la réalité virtuelle. Bien que le port des lunettes et la réduction du champ visuel induite puissent aussi biaiser le comportement du patient, des mouvements oculaires naturel permettraient de valider l'utilisation de la réalité virtuelle qui offre la possibilité de mettre en scène un panel riche de situations.

Notre étude a également une portée méthodologique, de par la conception d'un test de comparaison objective des configurations stéréoscopiques, basé sur l'enregistrement des mouvements oculaires de l'observateur alors qu'il suit du regard une cible mouvante réelle puis virtuelle. La confrontation avec une scène réelle permet alors de mesurer objectivement le rendu de la géométrie – positions, profondeurs, distances relatives et tailles – par les configurations des systèmes de caméras et d'affichage : plus le comportement oculaire se rapproche de celui exprimé en réel, plus les configurations peuvent être considérées réalistes.

Enfin, à plus long terme, les ajustements en temps réel des paramètres en fonction de la position du point de regard, réunis dans la configuration **FLOUDIC3**, pourraient se révéler plus performants que les valeurs plus classiques pour ces paramètres telles que celles utilisées

par **CONTROL**. Pour s'en assurer, une seconde étude serait nécessaire après avoir réalisé certaines améliorations décrites dans la section suivante au niveau de l'installation de la voûte et de l'implémentation.

6.2 Améliorations futures

Les améliorations futures envisageables découlent en grande partie des limitations exposées dans le chapitre précédent (voir section 5.4). Tout d'abord, il serait nécessaire de réaliser une calibration de l'affichage, dont la précision pourrait par exemple être testée à l'aide d'une scène très simple contenant des cubes et de cubes réels de taille identique positionnés dans la voûte aux mêmes endroits que dans le repère virtuel. Un test d'écart de vergence angulaire à la cible entre réel et virtuel pourrait également être employé. Certaines parties de l'implémentation, concernant la génération de flou et la détermination du point de regard, pourraient être optimisées en termes de temps de calcul afin d'augmenter le nombre d'images par seconde pour une immersion plus fluide. Nous pensons également qu'il serait important de concevoir une méthode de mesure du temps de mise à jour de l'ensemble de l'affichage par rapport au point de regard. Cela permettrait de vérifier si celui-ci a réellement eu un impact sur les résultats obtenus, ou s'il était suffisamment court pour ne pas être distingué par les participants, mais également de mesurer sa diminution lors des modifications d'implémentation. Il s'avère enfin critique de trouver une solution pour s'assurer de la synchronisation de tous les projecteurs tout au long des séances en immersion.

Concernant les ajustements des paramètres que nous proposons, nous pensons qu'il serait intéressant d'évaluer à nouveau leur impact après avoir écarté certains facteurs externes en réalisant les améliorations citées ci-dessus. L'expérimentation pourrait être menée sur un panel plus important de participants, afin d'éviter également les biais démographiques. Il serait de plus intéressant de tester d'autres paramètres ou valeurs pour chaque paramètre, en particulier concernant la variation de la DIC ou l'intensité du flou de profondeur de champ. Un test similaire à celui du robot, c'est-à-dire reproduisant en virtuel une scène réelle, serait utilisable à la fois pour comparer les configurations, mais aussi pour déterminer les valeurs optimales de paramètres. Il s'agirait, après enregistrement des mouvements de vergence du participant avec la configuration réelle, de modifier en temps réel les paramètres tandis que l'individu observe la scène virtuelle, jusqu'à ce que ses mouvements de vergence soient identiques à ceux enregistrés. Cette seconde expérimentation permettrait de valider ou d'infirmer plus sûrement nos hypothèses, et éventuellement de dégager positivement et de manière significative une configuration du système de caméras.

Enfin, nous envisageons également des extensions futures liées aux expérimentations qui

seront menées auprès des patients de l'Institut Philippe-Pinel de Montréal. La détermination de l'objet d'attention intéresse par exemple les cliniciens pour le diagnostic et le traitement de troubles comportementaux. En plus de connaître en temps réel sur quel élément de la scène le patient fixe son attention, les thérapeutes pourraient éventuellement vouloir l'inciter à se concentrer sur un autre objet, en spécifiant par exemple au système de générer un flou de profondeur de champ floutant toute la scène hormis ce dernier. Le flou serait alors utilisé non pas de manière passive afin d'obtenir un rendu plus confortable ou réaliste, mais comme un outil à des fins thérapeutiques. Il influencerait l'observateur pour concentrer son attention sur un objet particulier ou au contraire l'en distraire, comme Hillaire et al. (2008b) le proposent dans la conception de jeux vidéos ou comme c'est déjà le cas dans certains films pour capter l'attention des spectateurs et transmettre des sensations. Dans ce contexte d'application de la réalité virtuelle à la psychiatrie légale, ce projet de recherche et ses futures améliorations prennent alors une dimension dépassant le cadre informatique, en permettant d'exploiter de manière plus optimale le potentiel thérapeutique de la réalité virtuelle.

RÉFÉRENCES

- R. S. Allison, “The camera convergence problem revisited,” dans *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 167–178.
- D. Baril, “Les habiletés cognitives à la lumière de la théorie de l’évolution,” http://www.forum.umontreal.ca/numeros/2000_2001/forum_00_09_18/article06.html, 2000, accès : 2014-09-02.
- A. Bennett et J. Francis, “The eye as an optical system,” *The eye*, vol. 4, pp. 101–115, 1962.
- S. Best, “Perceptual and oculomotor implications of interpupillary distance settings on a head-mounted virtual display,” dans *Aerospace and Electronics Conference, 1996. NAECON 1996., Proceedings of the IEEE 1996 National*, vol. 1. IEEE, 1996, pp. 429–434.
- T. Blum, M. Wiecek, A. Aichert, R. Tibrewal, et N. Navab, “The effect of out-of-focus blur on visual discomfort when using stereo displays,” dans *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, pp. 13–17.
- P. Bourke, “Calculating stereo pairs,” <http://paulbourke.net/stereographics/stereorender/>, 1999, accès : 2012-11-09.
- , “3d stereo rendering using opengl (and glut),” ftp://ftp.sgi.com/sgi/opengl/contrib/kschwarz/GLUT_INTRO/SOURCE/PBOURKE/index.html, 1999, accès : 2013-09-10.
- G. T. Buswell, *How people look at pictures*. University of Chicago Press Chicago, 1935.
- U. Celikkan, G. Cimen, E. B. Kevinc, et T. Capin, “Attention-aware disparity control in interactive environments,” *The Visual Computer*, vol. 29, no. 6-8, pp. 685–694, 2013.
- S. Chartier et P. Renaud, “An online noise filter for eye-tracker data recorded in a virtual environment,” dans *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 2008, pp. 153–156.
- École polytechnique de Montréal, “Laboratoire d’infographie et de réalité virtuelle,” <http://www.polymtl.ca/rv/Materiel/index.php>, 2010-2014, accès : 2014-07-10.
- B. C. Daugherty, A. T. Duchowski, D. H. House, et C. Ramasamy, “Measuring vergence over stereoscopic video with a remote eye tracker,” dans *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 97–100.
- G. C. DeAngelis, “Seeing in three dimensions : the neurophysiology of stereopsis,” *Trends in Cognitive sciences*, vol. 4, no. 3, pp. 80–90, 2000.
- R. L. DeValois et K. K. DeValois, *Spatial vision*. NY : Oxford University Press, 1988.

- A. R. Diefendorf et R. Dodge, "An experimental study of the ocular reactions of the insane from photographic records," *Brain*, vol. 31, no. 3, pp. 451–489, 1908.
- N. A. Dodgson, "Variation and extrema of human interpupillary distance," dans *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 36–46.
- T. J. Doll, S. W. McWhorter, et D. E. Schmieder, "Simulation of human visual search in cluttered backgrounds," dans *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 37, no. 19. SAGE Publications, 1993, pp. 1310–1314.
- A. Duchowski, *Eye tracking methodology : Theory and practice*. Springer, 2007, vol. 373.
- A. T. Duchowski, E. Medlin, N. Cournia, A. Gramopadhye, B. Melloy, et S. Nair, "3d eye movement analysis for vr visual inspection training," dans *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 2002, pp. 103–110.
- A. T. Duchowski, B. Pelfrey, D. H. House, et R. Wang, "Measuring gaze depth with an eye tracker during stereoscopic display," dans *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*. ACM, 2011, pp. 15–22.
- K. Essig, M. Pomplun, et H. Ritter, "A neural network for 3d gaze recording with binocular eye trackers," *The International Journal of Parallel, Emergent and Distributed Systems*, vol. 21, no. 2, pp. 79–95, 2006.
- EverMotion, "Archinteriors vol. 30," http://www.evermotion.org/modelshop/show_product/archinteriors-vol-30/10469/0/0/, 2014, accès : 2014-08-10.
- D. Freeman, "Studying and treating schizophrenia using virtual reality : a new paradigm," *Schizophrenia bulletin*, vol. 34, no. 4, pp. 605–610, 2008.
- D. Freeman, K. Pugh, A. Antley, M. Slater, P. Bebbington, M. Gittins, G. Dunn, E. Kuipers, D. Fowler, et P. Garety, "A virtual reality study of paranoid thinking in the general population," *The British Journal of Psychiatry*, vol. 192, no. 4, pp. 258–263, 2008.
- T. Fukushima, M. Torii, K. Ukai, J. S. Wolffsohn, et B. Gilmartin, "The relationship between ca/c ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images," *Journal of Vision*, vol. 9, no. 13, p. 21, 2009.
- W. Ganong, *Physiologie médicale*. De Boeck Supérieur, 2005.
- J. Garding, J. Porrill, J. Mayhew, et J. Frisby, "Stereopsis, vertical disparity and relief transformations," *Vision research*, vol. 35, no. 5, pp. 703–722, 1995.
- K. Glantz, A. Rizzo, et K. Graap, "Virtual reality for psychotherapy : Current reality and future possibilities," *PSYCHOTHERAPY-RIVER EDGE-*, vol. 40, no. 1/2, pp. 55–67, 2003.
- R. L. Gregory, *Eye and brain : The psychology of seeing*. Princeton university press, 1997.

- J. M. Henderson, "Object identification in context : the visual processing of natural scenes." *Canadian Journal of Psychology*, vol. 46, no. 3, pp. 319–341, 1992.
- J. M. Henderson et A. Hollingworth, "Eye movements during scene viewing : An overview," *Eye guidance in reading and scene perception*, vol. 11, pp. 269–294, 1998.
- S. Hillaire, A. Lécuyer, R. Cozot, et G. Casiez, "Depth-of-field blur effects for first-person navigation in virtual environments," *Computer Graphics and Applications, IEEE*, vol. 28, no. 6, pp. 47–55, 2008.
- , "Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments," dans *Virtual Reality Conference, 2008. VR'08. IEEE*. IEEE, 2008, pp. 47–50.
- K. Holmqvist, M. Nyström, et F. Mulvey, "Eye tracker data quality : what it is and how to measure it," dans *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012, pp. 45–52.
- Q. Huynh-Thu, C. Vienne, et L. Blondé, "Visual storytelling in 2d and stereoscopic 3d video : effect of blur on visual attention," dans *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 865 112–865 112.
- A. Iijima, M. Haida, N. Ishikawa, H. Minamitani, et Y. Shinohara, "Head mounted goggle system with liquid crystal display for evaluation of eye tracking functions on neurological disease patients," dans *Proceedings of the IEEE conference on Engineering in Medicine and Biology Society*, vol. 4. IEEE, 2003, pp. 3225–3228.
- W. A. IJsselsteijn, H. de Ridder, et J. Vliegen, "Subjective evaluation of stereoscopic images : effects of camera parameters and display duration," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, no. 2, pp. 225–233, 2000.
- I'm in VR, "Chapter 6. middlevr for unity3d," <http://www.imin-vr.com/doc/current/ch06.html>, 2014, accès : 2014-06-20.
- , "Middlevr," <http://www.imin-vr.com/middlevr/>, accès : 2014-08-07.
- G. R. Jones, D. Lee, N. S. Holliman, et D. Ezra, "Controlling perceived depth in stereoscopic images," dans *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics, 2001, pp. 42–53.
- J. Ku, K. Han, H. R. Lee, H. J. Jang, K. U. Kim, S. H. Park, J. J. Kim, C. H. Kim, I. Y. Kim, et S. I. Kim, "Vr-based conversation training program for patients with schizophrenia : a preliminary clinical trial," *CyberPsychology & Behavior*, vol. 10, no. 4, pp. 567–574, 2007.
- Y.-M. Kwon, K.-W. Jeon, J. Ki, Q. M. Shahab, S. Jo, et S.-K. Kim, "3d gaze estimation and interaction to stereo display," *IJVR*, vol. 5, no. 3, pp. 41–45, 2006.

- M. Lambooi, M. Fortuin, I. Heynderickx, et W. IJsselsteijn, “Visual discomfort and visual fatigue of stereoscopic displays : a review,” *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 30 201–1, 2009.
- M. Lambooi, M. Fortuin, W. IJsselsteijn, et I. Heynderickx, “Measuring visual discomfort associated with 3d displays,” dans *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 370K–72 370K.
- M. T. Lambooi, W. A. IJsselsteijn, I. Heynderickx *et al.*, “Visual discomfort in stereoscopic displays : a review,” *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, no. 1, 2007.
- L. Leroy, P. Fuchs, et G. Moreau, “Real-time adaptive blur for reducing eye strain in stereoscopic displays,” *ACM Transactions on Applied Perception (TAP)*, vol. 9, no. 2, p. 9, 2012.
- L. Lipton, “Stereographics, developers handbook,” *StereoGraphics Corporation*, vol. 3, 1997.
- J. Lorenceau, “Cursive writing with smooth pursuit eye movements,” *Current Biology*, vol. 22, no. 16, pp. 1506 – 1509, 2012. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0960982212006719>
- B. H. McCormick, D. A. Batte, et A. T. Duchowski, “A virtual environment : Exploring the brain forest,” *Computación y sistemas*, vol. 1, no. 1, pp. 5–13, 1997.
- P. Milgram et M. Krüger, “Adaptation effects in stereo due to on-line changes in camera configuration,” dans *SPIE/IS&T 1992 Symposium on Electronic Imaging : Science and Technology*. International Society for Optics and Photonics, 1992, pp. 122–134.
- S. Nagata, “The binocular fusion of human vision on stereoscopic displays—field of view and environment effects,” *Ergonomics*, vol. 39, no. 11, pp. 1273–1284, 1996.
- Office québécois de la langue française, “Attention,” http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8363842, 2002, accès : 2013-03-25.
- , “Écartement des yeux,” http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=17499584, 2002, accès : 2013-07-26.
- Optitrack, “Motive :tracker, motion capture & 6 dof object tracking,” <https://www.naturalpoint.com/optitrack/products/motive/tracker/>, 2014, accès : 2014-08-01.
- , <https://www.naturalpoint.com/optitrack/>, 2014, accès : 2014-08-01.
- T. D. Parsons, A. A. Rizzo *et al.*, “Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias : a meta-analysis,” *Journal of behavior therapy and experimental psychiatry*, vol. 39, no. 3, pp. 250–261, 2008.

- L. Perez-Bayas, "Human factors involved in perception and action in a natural stereoscopic world : an up-to-date review with guidelines for stereoscopic displays and stereoscopic virtual reality (vr)," dans *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics, 2001, pp. 251–267.
- T. Pfeiffer, "Measuring and visualizing attention in space with 3d attention volumes," dans *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012, pp. 29–36.
- K. Rayner et A. Pollatsek, "Eye movements and scene perception," *Canadian Journal of Psychology/Revue canadienne de psychologie*, vol. 46, no. 3, pp. 342–376, 1992.
- E. M. Reingold, L. C. Loschky, G. W. McConkie, et D. M. Stampe, "Gaze-contingent multi-resolutional displays : An integrative review," *Human Factors : The Journal of the Human Factors and Ergonomics Society*, vol. 45, no. 2, pp. 307–328, 2003.
- P. Renaud, S. Bouchard, et R. Proulx, "Behavioral avoidance dynamics in the presence of a virtual spider," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 6, no. 3, pp. 235–243, 2002.
- P. Renaud, J. L. Rouleau, L. Granger, I. Barsetti, et S. Bouchard, "Measuring sexual preferences in virtual reality : A pilot study," *CyberPsychology & Behavior*, vol. 5, no. 1, pp. 1–9, 2002.
- G. Riguer, N. Tatarchuk, et J. Isidoro, "Real-time depth of field simulation," *ShaderX2 : Shader Programming Tips and Tricks with DirectX*, vol. 9, pp. 529–556, 2003.
- S. K. Rushton et P. M. Riddell, "Developing visual systems and exposure to virtual reality and stereo displays : some concerns and speculations about the demands on accommodation and vergence," *Applied Ergonomics*, vol. 30, no. 1, pp. 69–78, 1999.
- W. L. Shebilske et D. F. Fisher, "Understanding extended discourse through the eyes : How and why," dans *Eye Movements and Psychological Functions : International Views*, R. Groner, C. Menz, D. F. Fisher, et R. A. Monty, édés. Hillsdale, NJ : Lawrence Erlbaum Associates, 1983, pp. 303–314.
- SMI, "Eye tracking glasses 2.0," http://www.eyetracking-glasses.com/fileadmin/user_upload/documents/smi_etg20_flyer.pdf, 2014, accès : 2014-06-17.
- L. B. Stelmach, W. J. Tam, F. Speranza, R. Renaud, et T. Martin, "Improving the visual comfort of stereoscopic images," dans *Electronic Imaging 2003*. International Society for Optics and Photonics, 2003, pp. 269–282.
- D. Strickland, "Virtual reality for the treatment of autism," *Studies in Health Technology and Informatics*, pp. 81–86, 1997.

T. Takeda, K. Hashimoto, N. Hiruma, et Y. Fukui, “Characteristics of accommodation toward apparent depth,” *Vision research*, vol. 39, no. 12, pp. 2087–2097, 1999.

K. Talmi et J. Liu, “Eye and gaze tracking for visually controlled interactive stereoscopic displays,” *Signal Processing : Image Communication*, vol. 14, no. 10, pp. 799–810, 1999.

Unity, “Create the games you love with unity,” <http://unity3d.com/unity>, 2014, accès : 2014-08-07.

Visio, “Vioso anyblend : Edge blended projection integrated in windows 7,” http://www.vioso.com/products_anyblend.php, 2012, accès : 2014-06-20.

Volfoni, “Activeyes,” <http://www.volfoni.com/en/virtual-reality/17>, 2014, accès : 2014-08-07.

—, “Edge vr,” <http://www.volfoni.com/en/virtual-reality/15>, 2014, accès : 2014-08-07.

J. P. Wann, S. Rushton, et M. Mon-Williams, “Natural problems for stereoscopic depth perception in virtual environments,” *Vision research*, vol. 35, no. 19, pp. 2731–2736, 1995.

A. J. Woods, T. Docherty, et R. Koch, “Image distortions in stereoscopic video systems,” dans *IS&T/SPIE’s Symposium on Electronic Imaging : Science and Technology*. International Society for Optics and Photonics, 1993, pp. 36–48.

A. L. Yarbus, *Eye movements and vision*. Plenum press, 1967.

Y.-Y. Yeh et L. D. Silverstein, “Limits of fusion and depth judgment in stereoscopic color displays,” *Human Factors : The Journal of the Human Factors and Ergonomics Society*, vol. 32, no. 1, pp. 45–60, 1990.

G. Zachmann *et al.*, *Virtual Reality in Assembly Simulation : Collision Detection, Simulation Algorithms, and Interaction Techniques*. Fraunhofer-IRB-Verlag, 2000.

ANNEXE A

Notations et calculs des paramètres du système de caméras

D	Distance focale	α_v	Angle d'ouverture verticale
d	Limite proche en Z du tronc de projection		$= 2 \times \arctan(\frac{H}{2 \times D})$
L	Largeur du tronc de projection à la distance en Z D	$H_{1/2d}$	Demi-largeur du tronc de projection à la distance en Z d
H	Hauteur du tronc de projection à la distance en Z D		$= d \times \tan(\alpha_{1/2v})$
dic	Distance inter-caméras	$L_{1/2d}$	Demi-hauteur du tronc de projection à la distance en Z d
p	Point de regard		$= H_{1/2d} \times \frac{L}{H}$

Tableau A.1 Paramètres des caméras virtuelles dans le cas général.

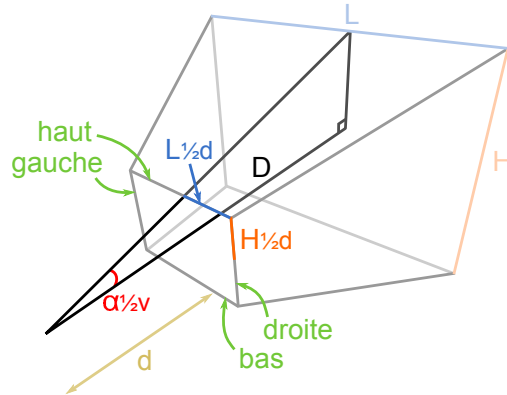


Figure A.1 Tronc de projection.

Δl	Décalage avec les limites du tronc symétrique
	$= dic_{1/2} \times \frac{d}{D}$

Tableau A.2 Paramètre spécifique à la configuration parallèle avec troncs asymétriques.

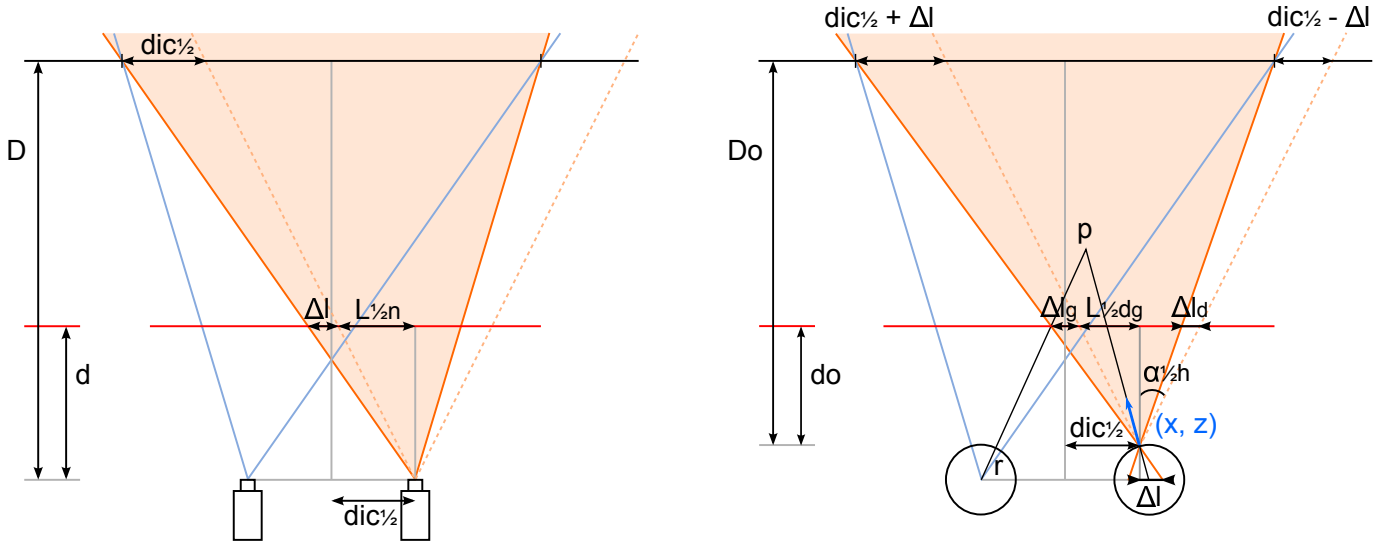


Figure A.2 Configuration parallèle avec troncs asymétriques (à gauche) et avec simulation des globes oculaires (à droite). Les lignes en pointillés correspondent au tronc symétrique.

r	Rayon de l'œil	Δl	Décalage avec les limites du tronc symétrique $= z \times r \times \tan(\alpha_{1/2h})$
D_o	$D - z \times r$	Δl_g	Décalage à gauche avec les limites du tronc symétrique $= (dic_{1/2} + \Delta l) \times \frac{d_o}{D_o}$
d_o	$d - z \times r$	Δl_d	Décalage à droite avec les limites du tronc symétrique $= (dic_{1/2} - \Delta l) \times \frac{d_o}{D_o}$
$dic_{1/2}$	Distance entre la pupille droite et le centre des yeux $= \frac{dic}{2} + x \times r$		
$\alpha_{1/2h}$	Demi-angle d'ouverture horizontale $= \arctan(\frac{L_{1/2d}}{d_o})$		

Tableau A.3 Paramètres spécifiques à la configuration parallèle avec troncs asymétriques et simulation des globes oculaires.

Les calculs sont ici détaillés pour la caméra de droite, et sont applicables symétriquement pour la caméra de gauche.

(x_o, z_o)	Coordonnées de la caméra droite $= (dic_{1/2}, 0)$	L_2	$= l_2 \times \tan(\alpha)$
(x_d, z_d)	Coordonnées de l'intersection entre l'axe de vue et l'écran $= (x_o + \frac{D}{z} \times x, D)$	D_o	$= L_1 + L_2$
(x_m, z_m)	Coordonnées de l'intersection entre l'axe de vue moyen et l'écran	Δl	Décalage entre le milieu horizon- tal du tronc de projection et la ca- méra $= x_m - l_2$
l_1	$= x_o - x_m$	Δl_d	Limite à droite du tronc de pro- jection à la distance en Z d $= \frac{L_{1/2} - \Delta l}{D_o} \times d$
L_1	$= \frac{l_1}{\cos(\alpha)}$	Δl_g	Limite à gauche du tronc de pro- jection à la distance en Z d $= -\frac{L_{1/2} + \Delta l}{D_o} \times d$
l_2	$= \cos(\pi - \alpha) \times (x_m - x_d)$		

Tableau A.4 Paramètres spécifiques à la configuration convergente.

Les notations l_1 , l_2 , L_1 et L_2 correspondent à des valeurs algébriques qui peuvent être négatives, ce qui est nécessaire pour que le calcul de la distance focale D s'effectue correctement quel que soit la position du point de regard.

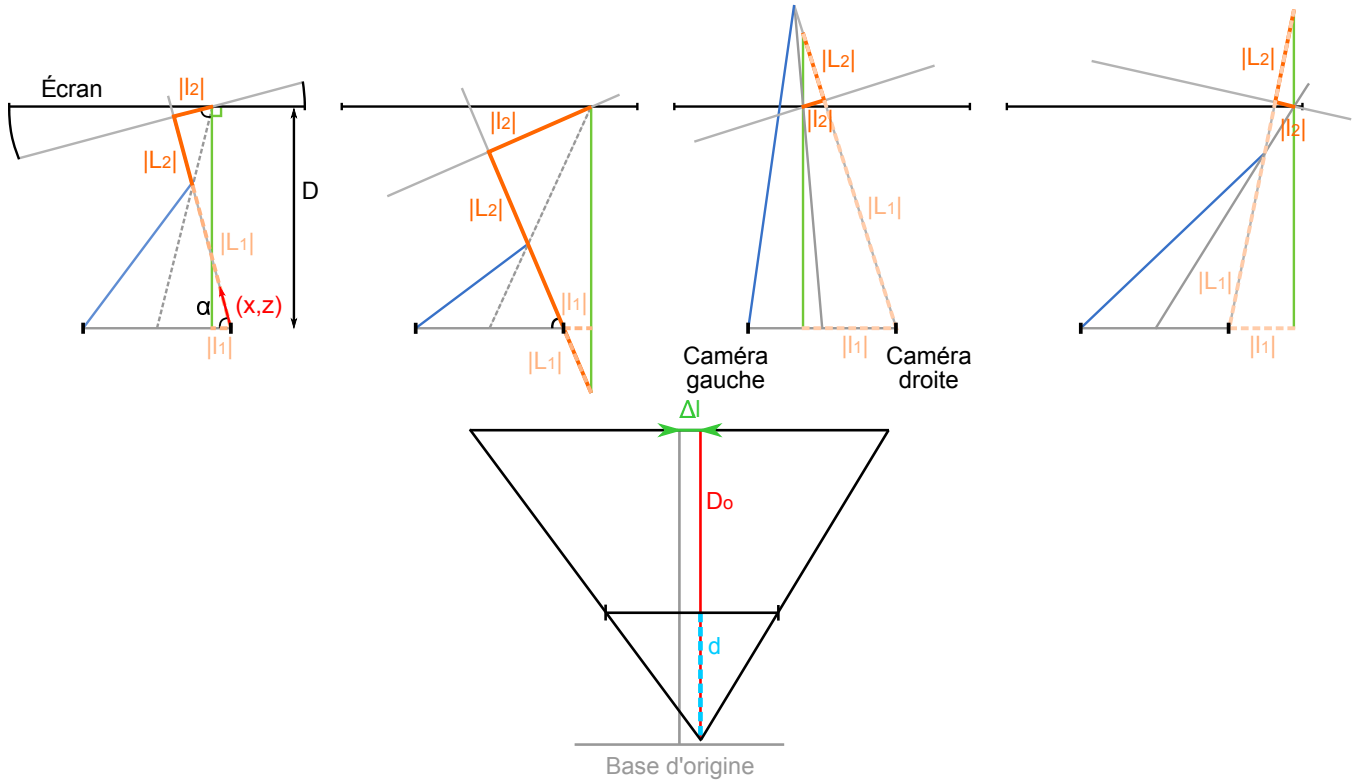


Figure A.3 Paramètres de la configuration convergente dans différents cas de figure dépendant de la position du point de regard (en haut). Exemple du tronc de projection de la caméra droite (en bas).

ANNEXE B

Informations supplémentaires sur l'implémentation et l'expérimentation

Détails sur la DLL développée

Les fonctions principales de la DLL développée comprennent :

- int EXPORT_API ConnectServer (char* serverIP, char* fileLog, char* fileLogAddReport, double scaleUnityETG);
- int EXPORT_API CloseConnectionServer ();
- int EXPORT_API SuscribeDataStream ();
- int EXPORT_API UnsuscribeDataStream ();
- void EXPORT_API GetReports (gaze_report ** cpy, int& cnt);
- void EXPORT_API CallbackSaveReports (iViewTicket * const ticket);

La dernière fonction est appelée à la réception de chaque rapport sur le flux de données de type `IVIEWDATASTREAM_GAZE_INFORMATION`. Les informations contenues dans celui-ci qui sont destinées à être transmises à Unity sont enregistrées dans une **struct** spécifique de notre DLL : `_gaze_report`, laquelle est stockée dans la liste de notre classe **StreamDataWrapper**. Cette liste est consultée par Unity de manière asynchrone avec la réception des rapports et sert en quelque sorte de tampon, s'assurant qu'aucun rapport ne sera perdu si le calcul des images dans Unity s'effectue à une fréquence inférieure à 60 Hz (ce qui fut le cas pour notre expérimentation). Un **mutex** permet de s'assurer que les opérations d'ajout – par **CallbackSaveReports** – et de suppression – lorsque les rapports sont récupérés dans Unity – sur la liste ne rentrent pas en conflit. À noter que les positions 3D des centres des yeux sont données en mm, et dans le repère local aux ETG dont l'axe X est initialement opposé à celui de Unity. Les coordonnées en X pour les quatre vecteurs 3D sont donc inversées, et les positions divisées par un facteur 1000 afin de correspondre à l'échelle métrique de Unity et à son repère.

Méthode **SmoothDamp** de Unity

L'une des techniques de lissage testées utilisait la méthode **SmoothDamp** de la classe **Vector3** de Unity. Celle-ci prend en argument un vecteur de vitesse, qu'elle met à jour automatiquement et de manière intelligente en se basant sur les précédents appels. Ses autres arguments comprennent deux valeurs permettant d'ajuster le lissage, **smoothTime** et **maxSpeed**. Le premier indique le temps approximatif pour atteindre la cible, le deuxième, optionnel,

permet de borner la vitesse maximale : une valeur basse stabilisera le point de regard lors d'une fixation mais diminuera la réactivité lors d'une saccade, une valeur haute préservera la réactivité au détriment du lissage.

Méthode Raycast de Unity

Le *raycasting* est réalisé en utilisant la méthode `Raycast` de la classe `Physics` de Unity, qui prend en argument un objet `Ray` composé d'une origine et d'une direction, un pointeur vers un objet `RaycastHit` et une longueur pour le rayon (50 m dans notre cas). C'est dans le `RaycastHit` que seront stockées les informations résultantes du *raycasting*, telles que les coordonnées du point d'intersection ou le nom de l'objet intersecté.

Changement de repère

Au repère global d'une scène Unity, communément appelé repère monde, s'ajoute un repère local pour chaque `GameObject`. Le passage des coordonnées d'un repère à l'autre s'effectue grâce aux fonctions `localPosition` et `position`, ainsi que `localRotation` et `rotation`.

En ce qui concerne les caméras virtuelles utilisées pour calculer les images stéréoscopiques, il s'agit d'objets `MiddleVR` qui ne sont accessibles que via le `VRDisplayMgr` de la manière suivante :

```
dimgr = MiddleVR.VRDisplayMgr;
dimgr.GetNode("CameraStereo_Floor.Left");
```

Elles sont situées sous le `HeadNode` dans le graphe de scène, c'est-à-dire que leurs coordonnées locales sont exprimées dans le repère de ce dernier. Étant donné que les positions des yeux récupérées du serveur ETG sont également exprimées dans celui-ci, les caméras peuvent être positionnées directement à leur niveau (ou de manière décalée en X lorsque la DIC utilisée ne correspond pas à la DIP anatomique) en utilisant les fonctions `SetXLocal`, `SetYLocal` et `SetZLocal`.

Mise en place des lunettes sur le participant

À l'intérieur de la voûte, les lunettes sont installées sur le participant et maintenues grâce à une dragonne ajustable à l'arrière de la tête. Deux tailles de ponts nasaux permettent de régler la hauteur de ses yeux par rapport aux caméras des lunettes, à l'aide d'`iViewETG`, qui indique les ajustements verticaux à effectuer afin que le suivi oculaire soit optimal. Le corps rigide correspondant à la tête est enregistré via `Motive :Tracker`, de manière à ce qu'il soit correctement orienté dans la voûte, tel qu'expliqué en 3.3.1. Après le lancement du serveur

ETG, le participant est invité à fixer durant une quinzaine de secondes un point situé au centre de son champ visuel tandis que la calibration automatique s'effectue. Les altérations des positions relatives des lunettes ou des marqueurs pouvant conduire à une augmentation de l'imprécision de l'estimation du POR, il est indiqué au participant de ne plus toucher aux lunettes une fois ces procédures terminées. Les pauses destinées au repos de ses yeux peuvent s'effectuer par exemple en les fermant, ou en interrompant leur occultation alternée par les lunettes actives Volfoni en éteignant l'appareil clipsé au col.

Détails de la scène utilisée pour l'expérimentation

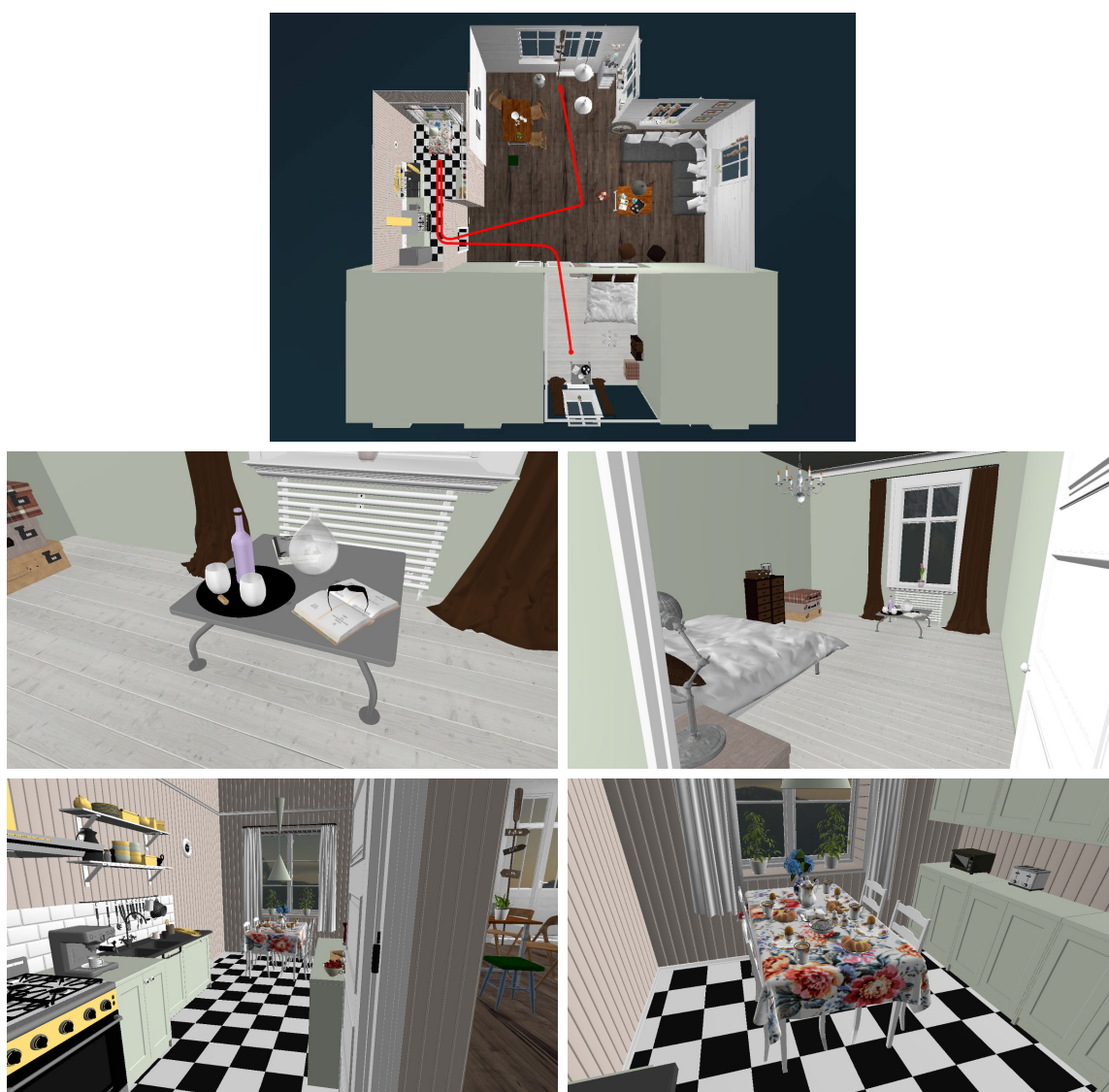


Figure B.1 Chemin de navigation (en haut) et détails de la scène d'appartement.

ANNEXE C

Mesures de la qualité suivi par les ETG

Comme présenté en section 3.3.1, les lunettes ETG fournissent des mesures d'incertitude sur leur détection de la pupille et leur estimation du globe oculaire pour chacun des deux yeux. De manière annexe au projet, nous nous sommes intéressés aux variations de ces mesures au cours du temps, en supposant que certains facteurs pouvaient augmenter les incertitudes et d'autres les diminuer. Par exemple, si les lunettes se déplacent relativement à la tête, l'angle de vision des caméras par rapport aux yeux est affecté et le modèle utilisé pour repérer ces derniers dans les images filmées n'est plus optimal, ce qui pourrait augmenter l'incertitude. *A contrario*, ce modèle peut se mettre à jour et s'améliorer au fil du temps, conduisant à une diminution de l'incertitude. Les différences de luminosité entre les étapes peuvent également influencer sur la taille de la pupille et donc la détection de celle-ci. Cependant, aucune remarque générale ne se dégage de l'observation des variations des mesures d'incertitude au cours du temps de chaque participant. La figure C.1 présente les exemples de deux participants pour lesquels les variations sont contraires : la détection des pupilles et des globes oculaires s'améliore pour le premier, tandis qu'elle empire pour le second.

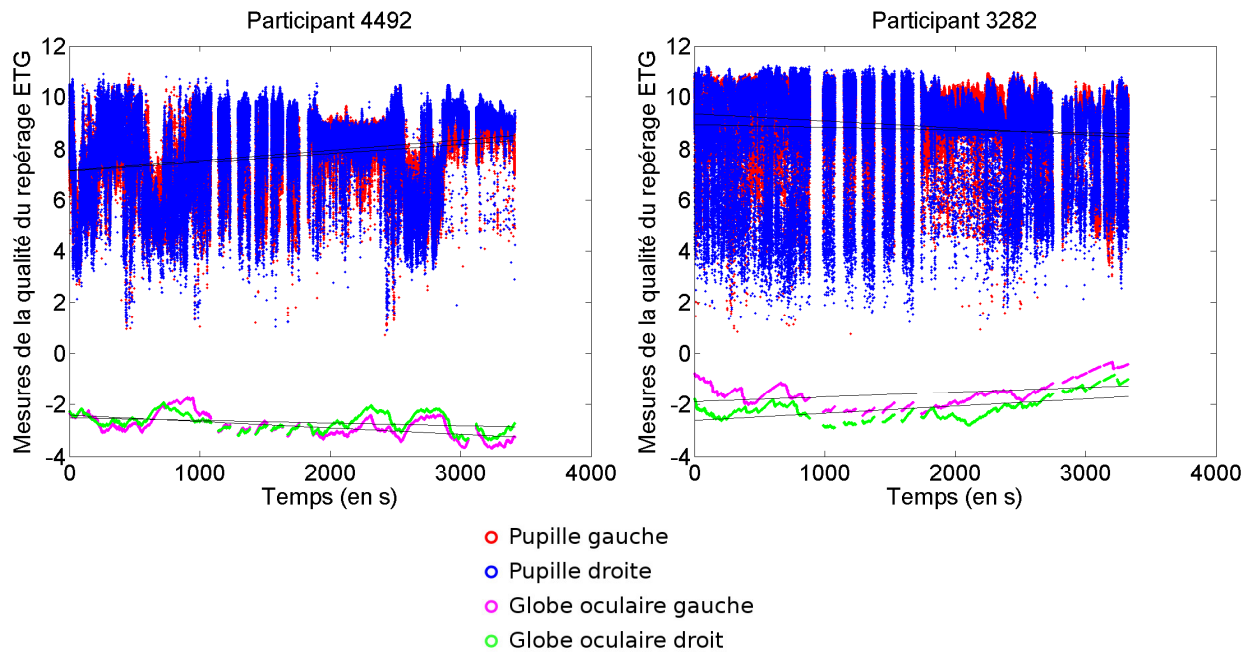


Figure C.1 Incertitudes de la détection des pupilles et des globes oculaires au cours du temps.

ANNEXE D

Notations subjectives

Critère	Se concentrer sur ...
Confort	... La gêne et la fatigue visuelles ressenties durant de la navigation, les efforts nécessaires pour réaliser la fusion des images stéréoscopiques
Réalisme	... Le réalisme, la crédibilité du monde virtuel et des objets environnants. La scène et les objets qui la composent paraissent-ils réels ?
Plaisir	... Le plaisir, l'amusement ressenti durant la navigation. Par exemple, celui-ci peut être diminué par la gêne ressentie ou le manque de réalisme de la scène
Distances	... Les distances, les tailles des objets et leurs proportions : sont-elles bien rendues ? Réalistes ? Par exemple, le rendu des distances peut être affecté par des difficultés de fusion
Immersion	... Le sentiment d'immersion ressenti durant la navigation. Croyez-vous à ce que vous voyez ? Vous sentez-vous immergé dans la scène virtuelle ? Ressentez-vous la présence des objets et des personnages autour de vous, même indépendamment de leur côté réaliste ?

Tableau D.1 Indications sur les critères de notation subjective.

	CONTROLE	DIC1	DIC3	FLOU	FLOUDIC1	FLOUDIC3
Confort	0,600	0,531	0,534	0,558	0,452	0,524
Réalisme	0,547	0,536	0,532	0,523	0,472	0,483
Plaisir	0,555	0,530	0,536	0,524	0,449	0,509
Distances	0,640	0,634	0,605	0,571	0,526	0,582
Immersion	0,612	0,628	0,592	0,561	0,507	0,562
Confort	7	5	4	1	1	2
Réalisme	6	4	2	3	2	2
Plaisir	7	7	3	2	1	2
Distances	5	3	2	4	3	4
Immersion	5	7	4	4	0	3

Tableau D.2 Moyennes et dénombrement des participants ayant octroyé les meilleures notes, détaillés par critère.

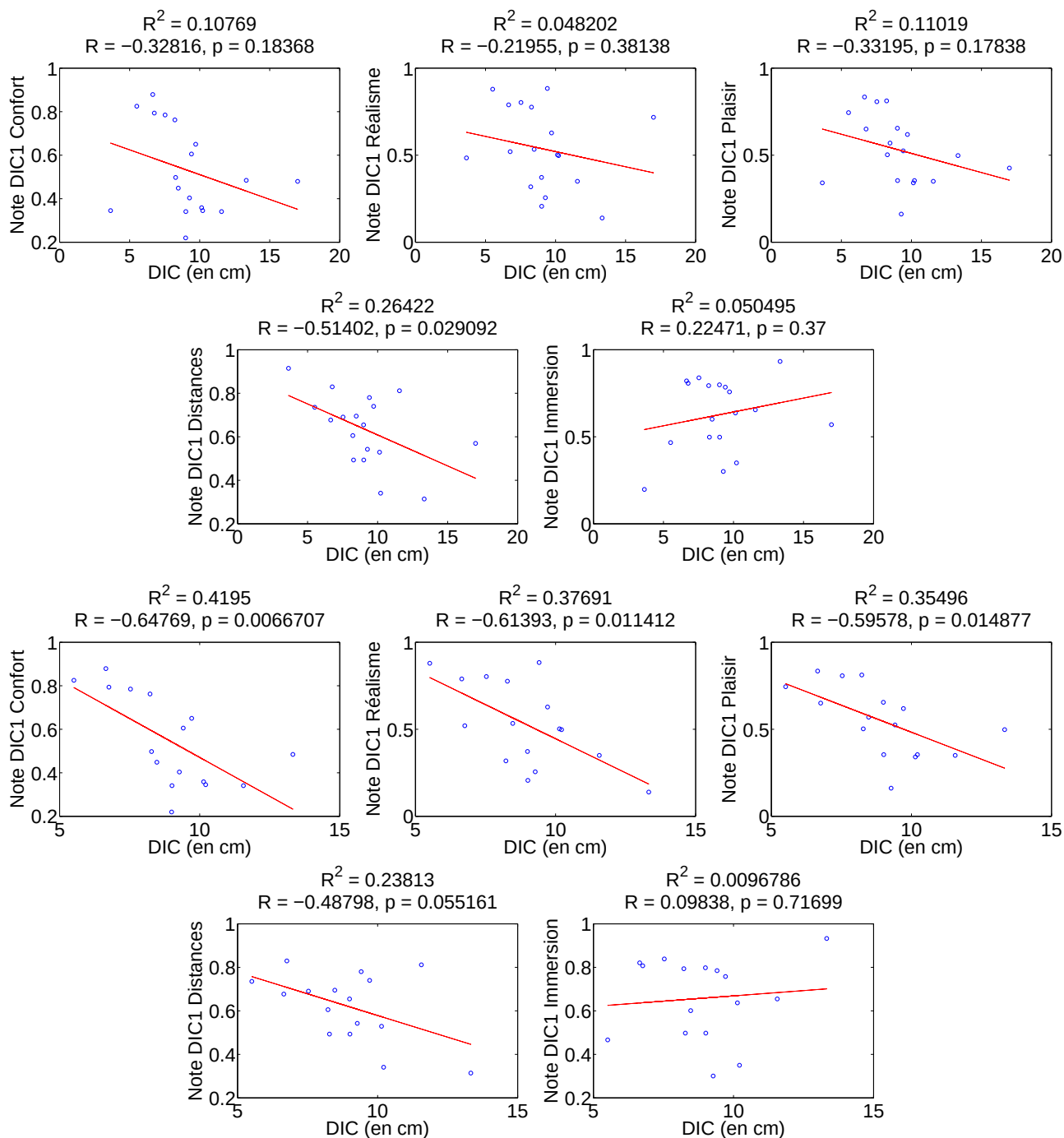


Figure D.1 Notes obtenues par **DIC1** en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).

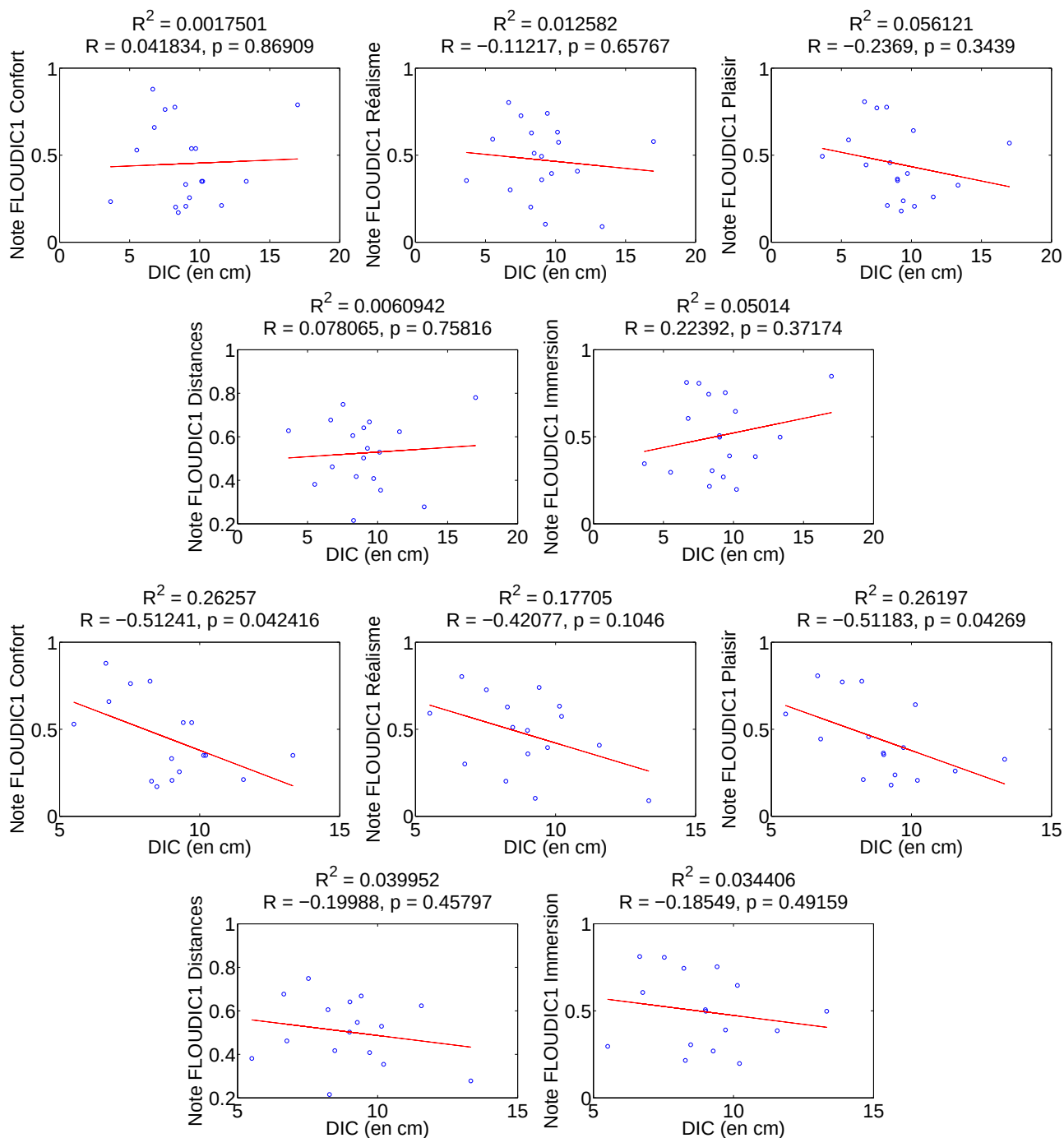


Figure D.2 Notes obtenues par **FLOUDIC1** en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).

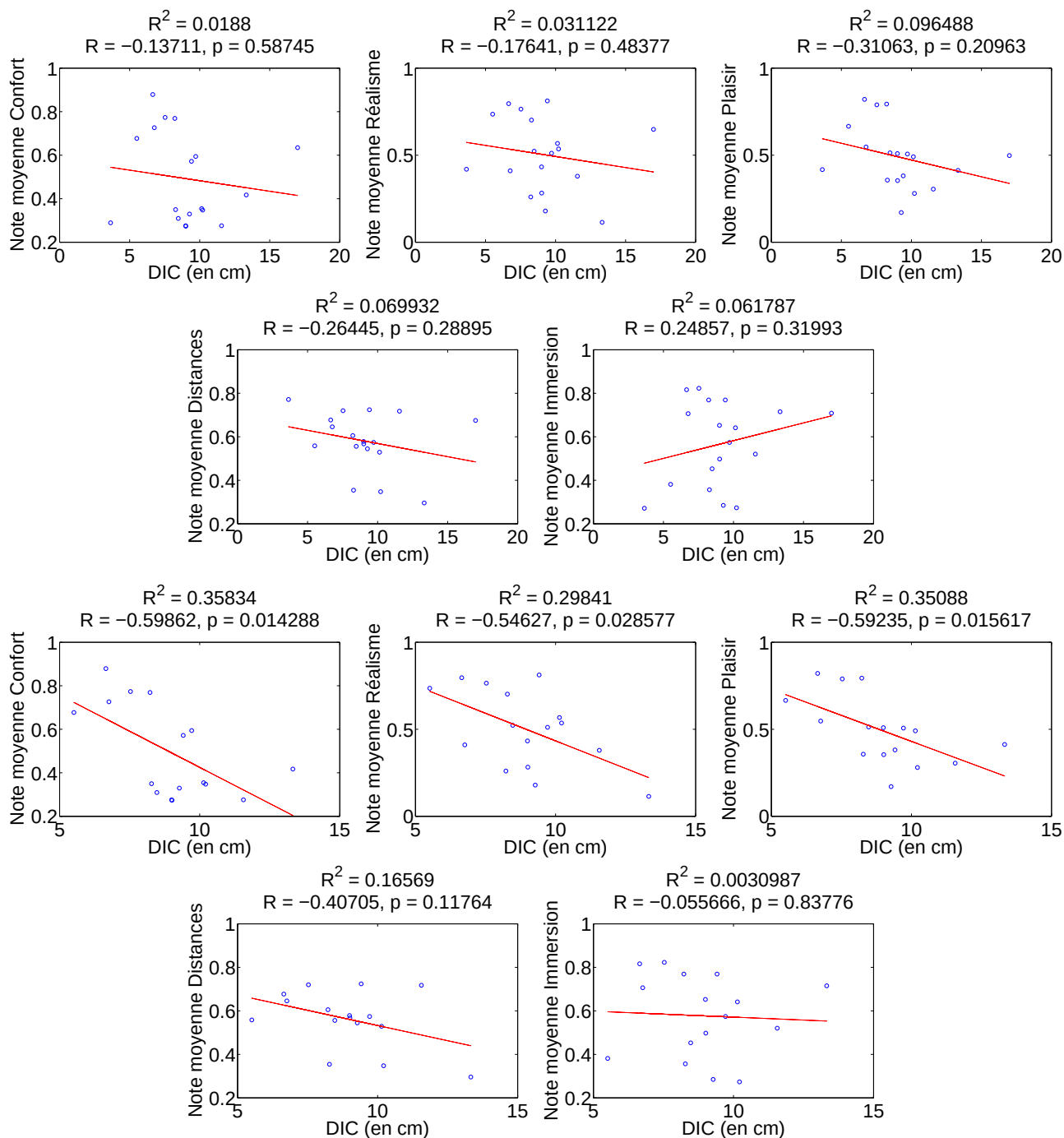


Figure D.3 Notes moyennes **DIC1** et **FLODIC1** en fonction du choix de la DIC1 (5 graphiques du haut), puis en écartant les 2 DIC extrêmes (5 graphiques du bas).

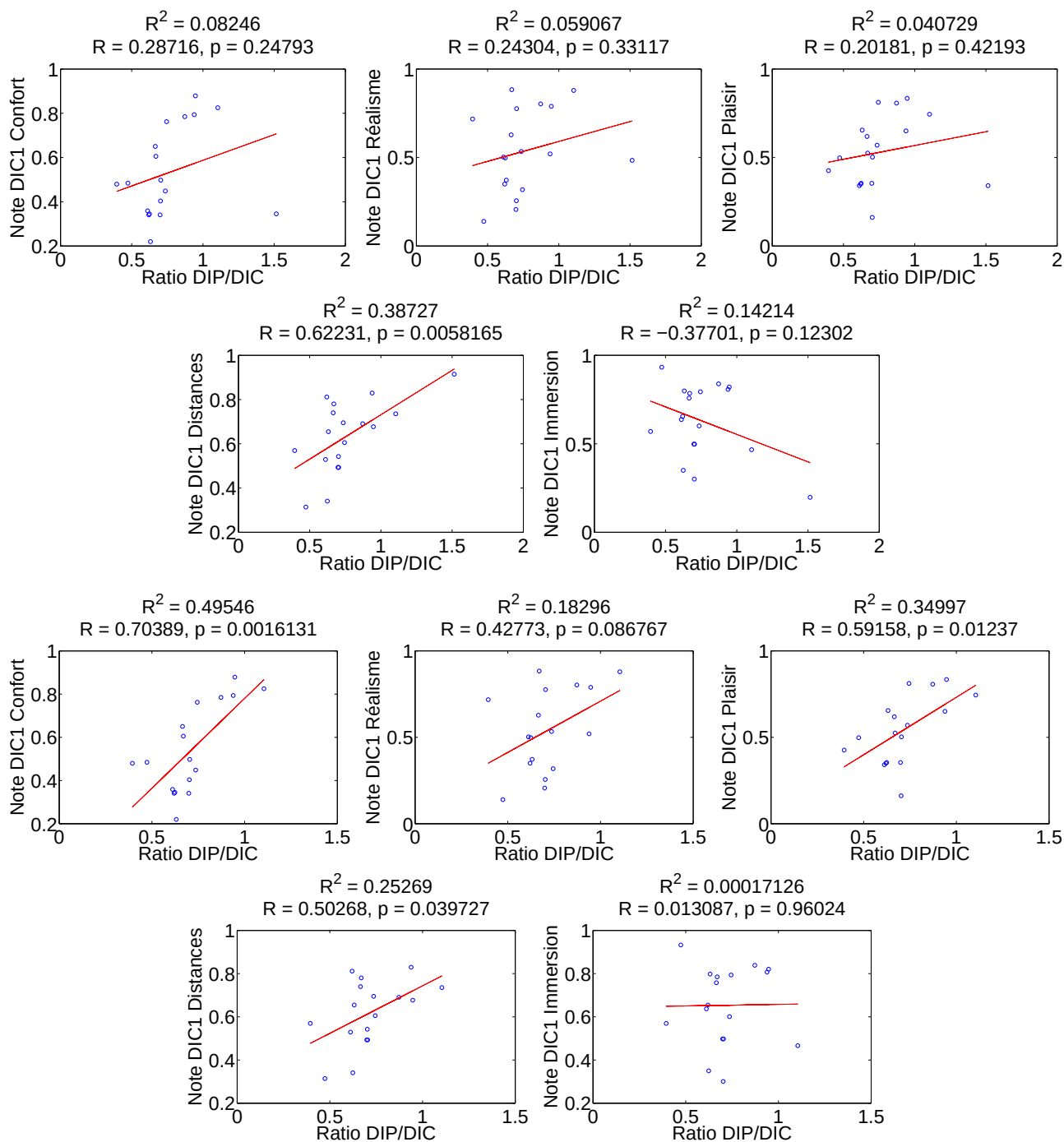


Figure D.4 Notes obtenues par **DIC1** en fonction du ratio DIP/DIC1 (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).

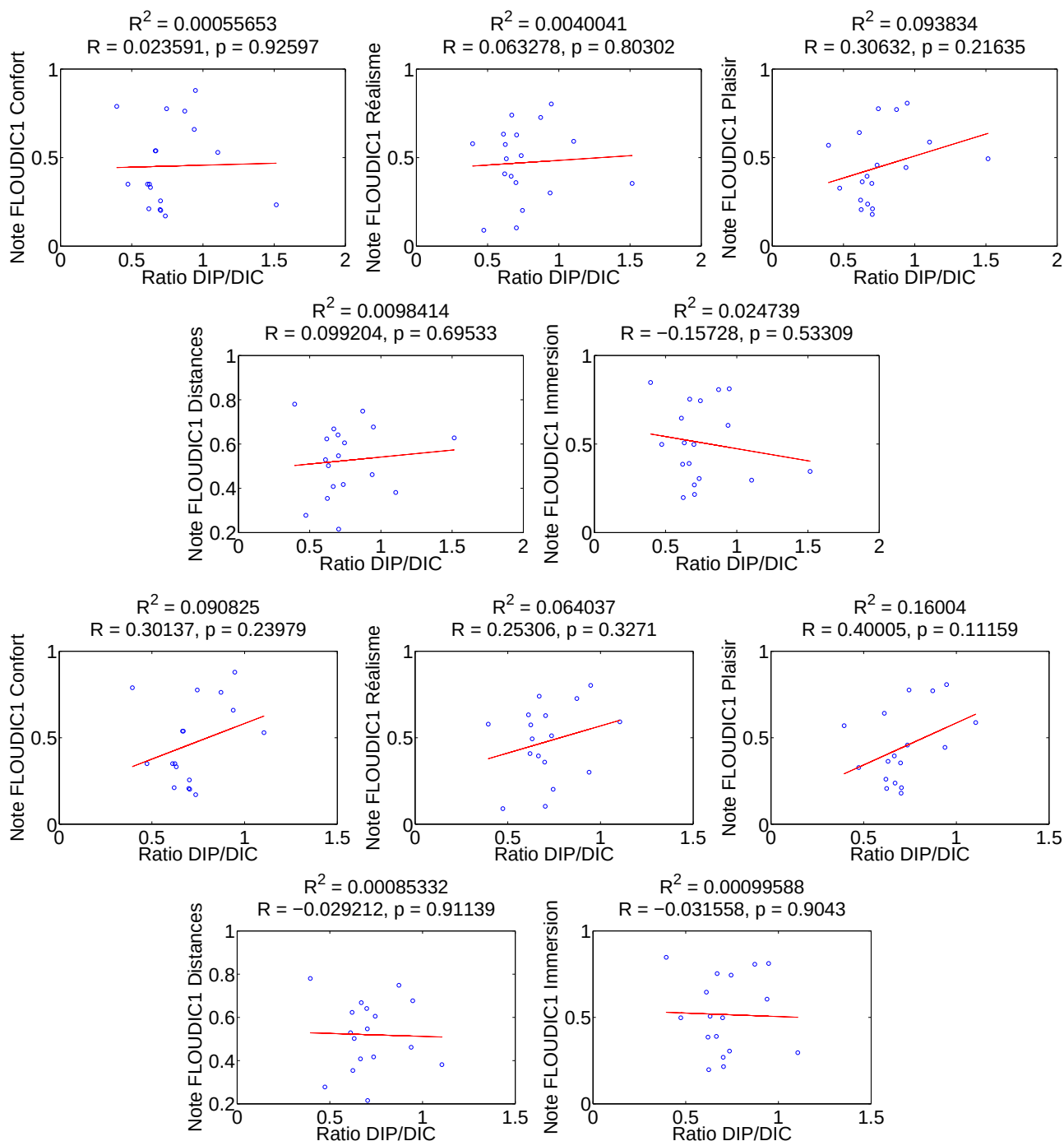


Figure D.5 Notes obtenues par **FLOUDIC1** en fonction du ratio DIP/DIC (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).

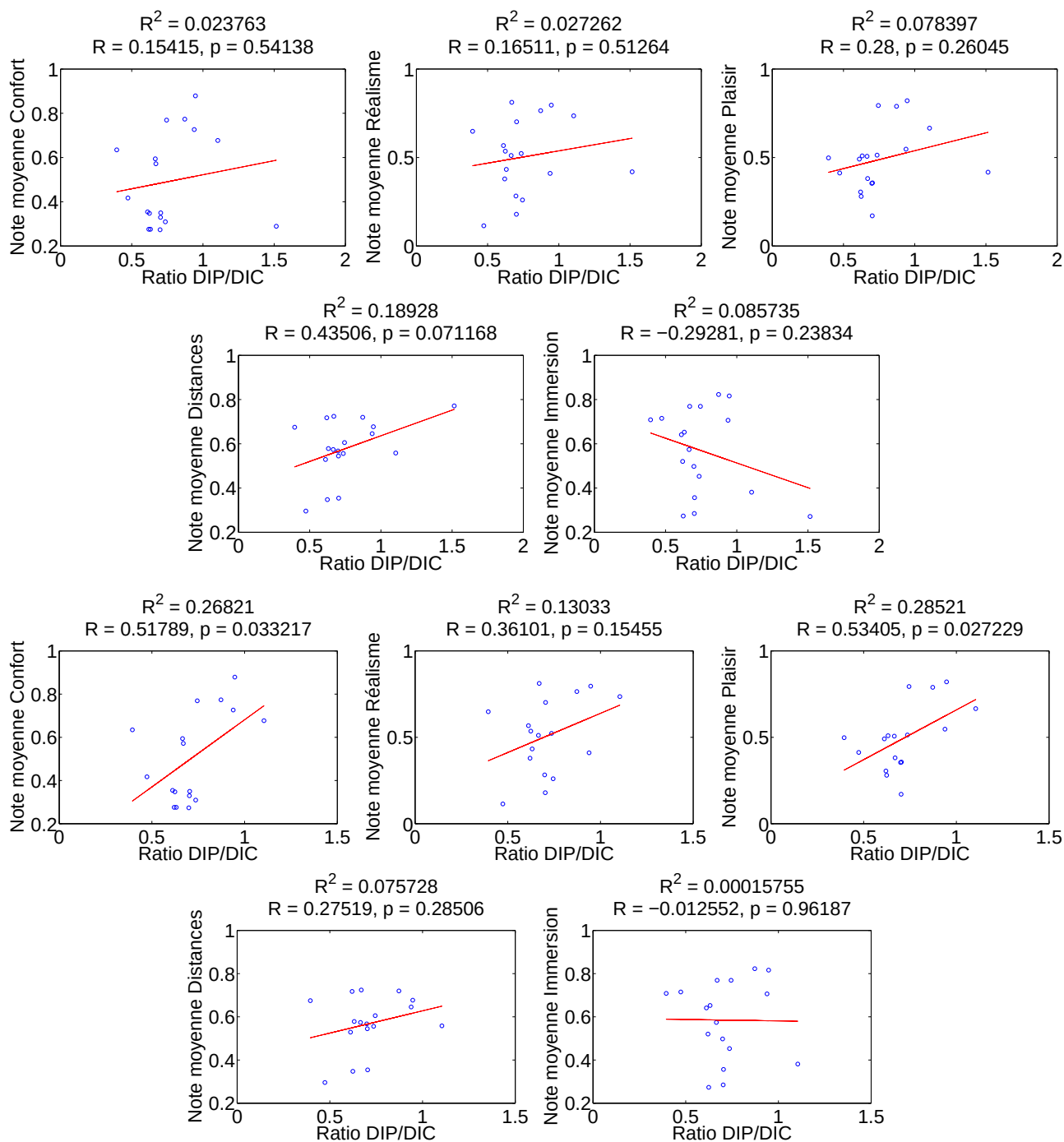


Figure D.6 Notes moyennes **DIC1** et **FLOUDIC1** en fonction du ratio DIP/DIC (5 graphiques du haut), puis en écartant la valeur extrême (5 graphiques du bas).

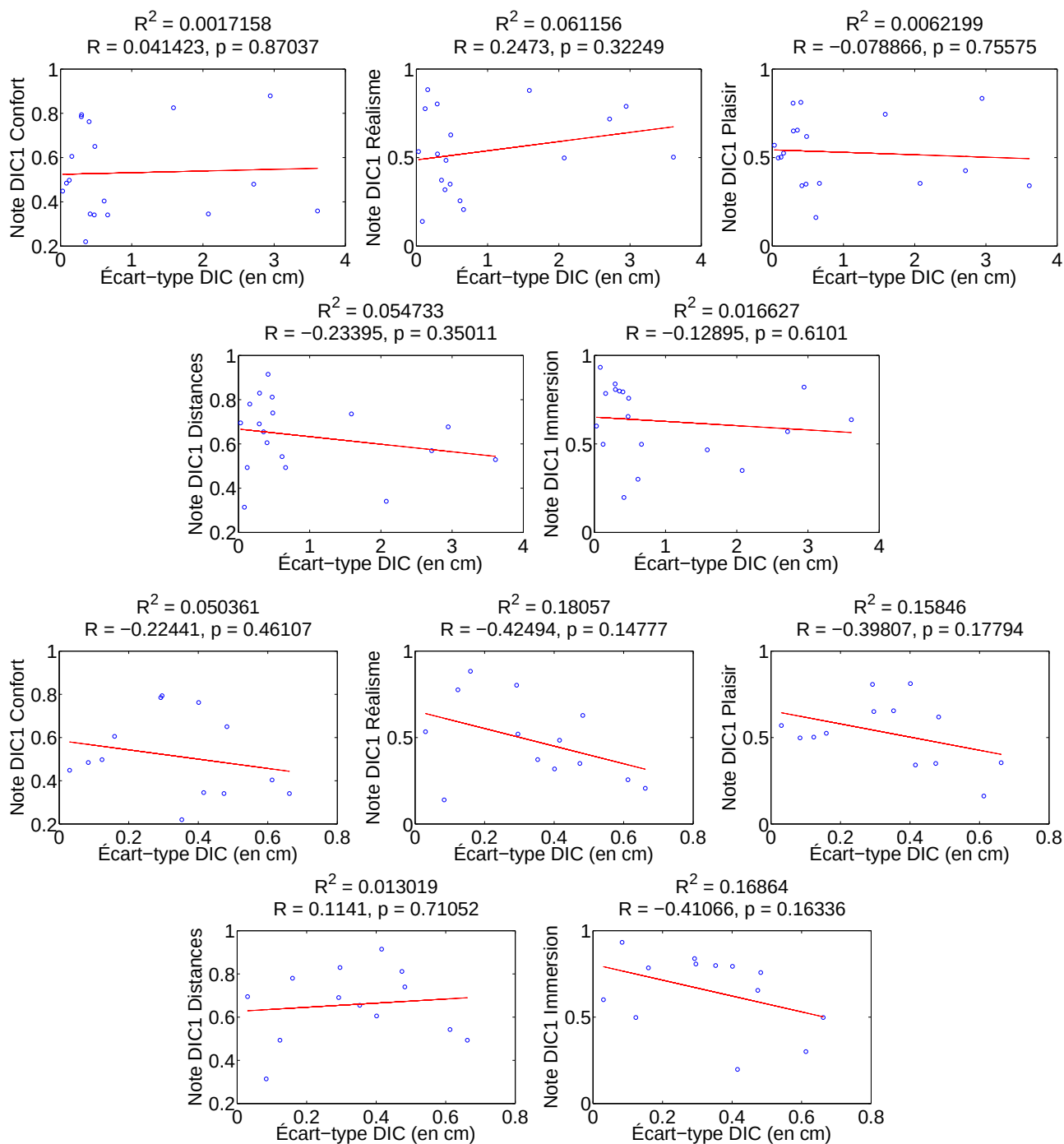


Figure D.7 Notes obtenues par **DIC1** en fonction de l'écart-type entre les 3 DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).

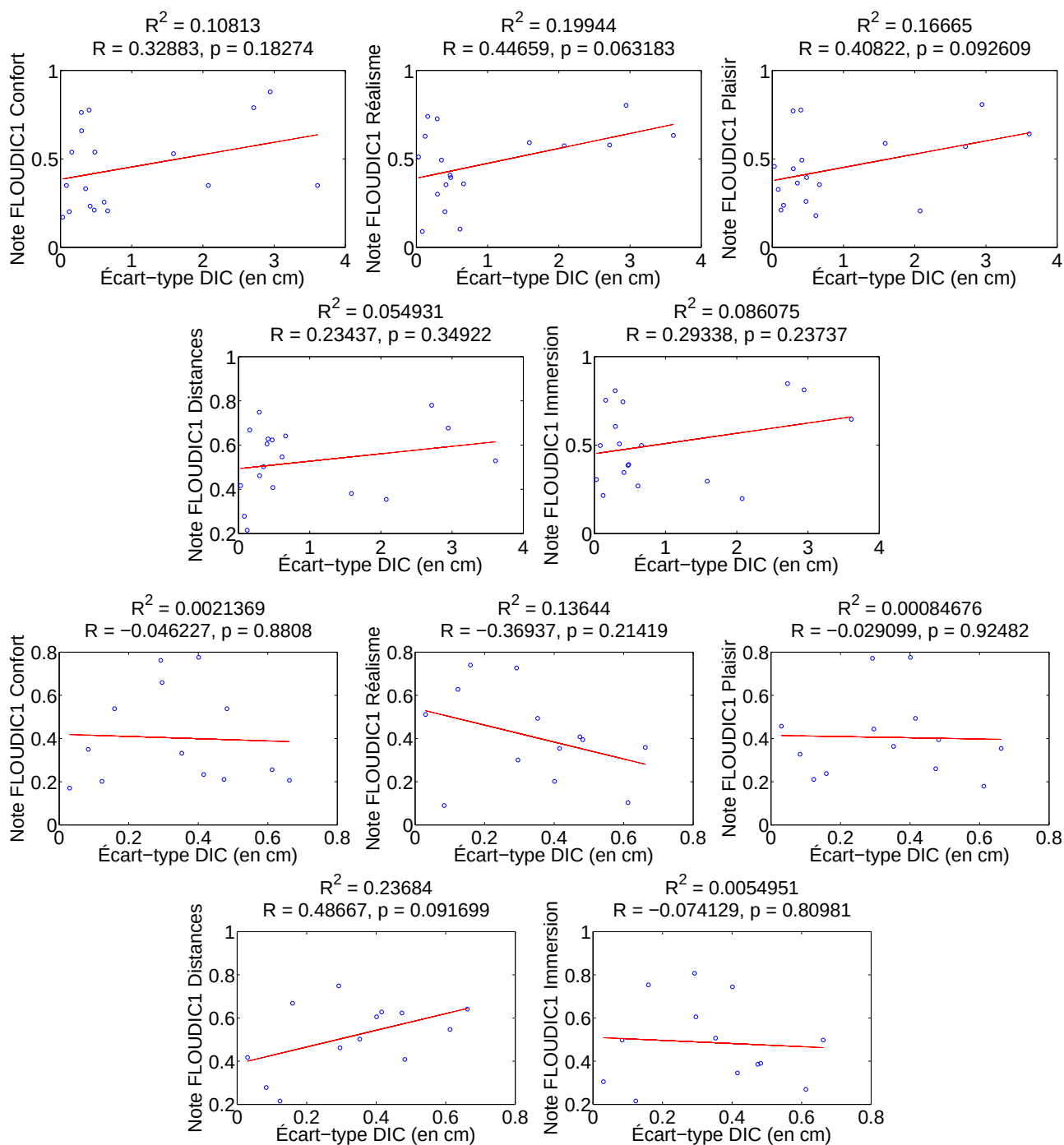


Figure D.8 Notes obtenues par **FLOUDIC1** en fonction de l'écart-type entre les 3 DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).

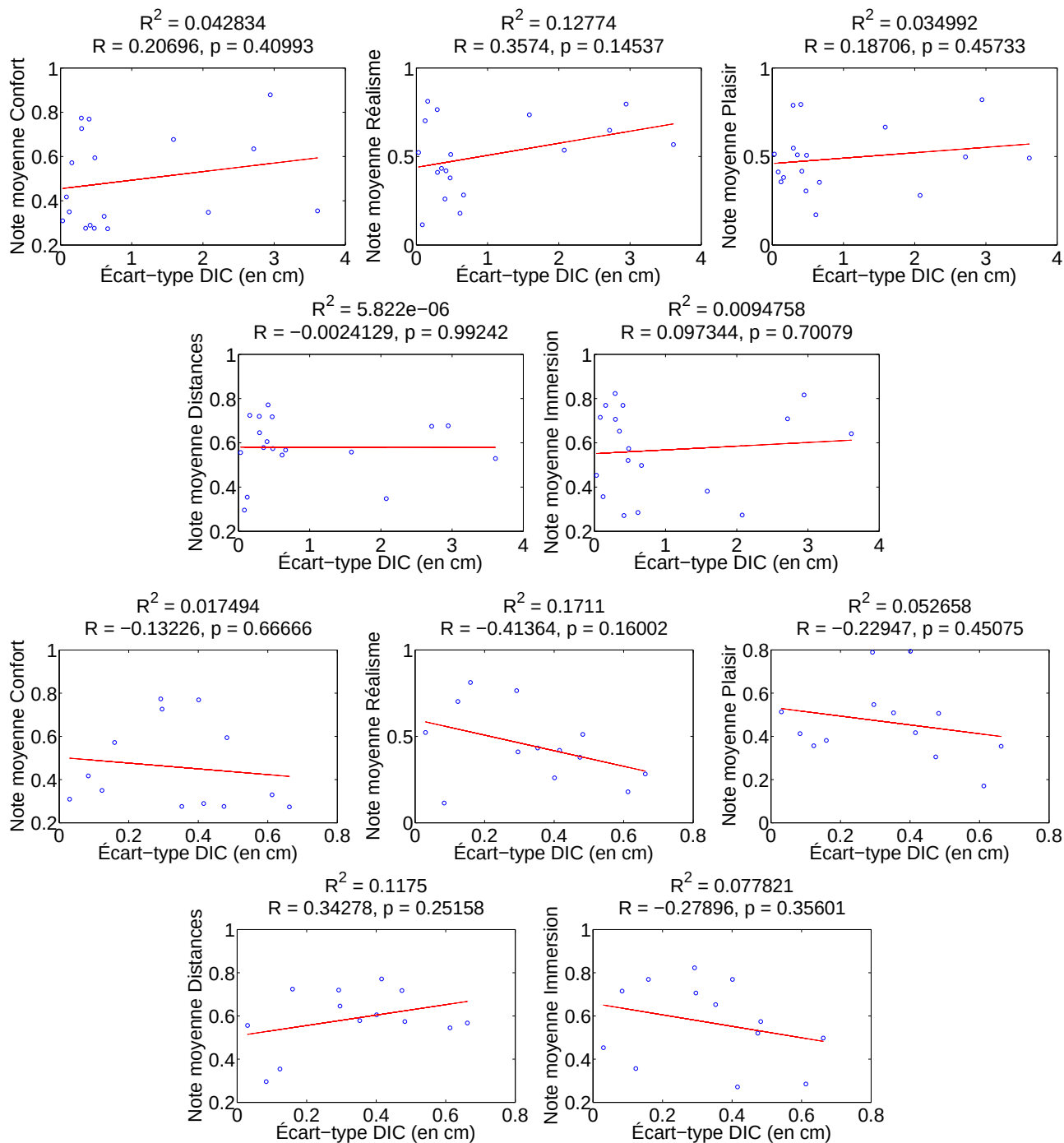


Figure D.9 Notes moyennes **DIC1** et **FLOUDIC1** en fonction de l'écart-type entre les DIC1 choisies (5 graphiques du haut), puis en écartant les 5 écarts-types extrêmes (5 graphiques du bas).

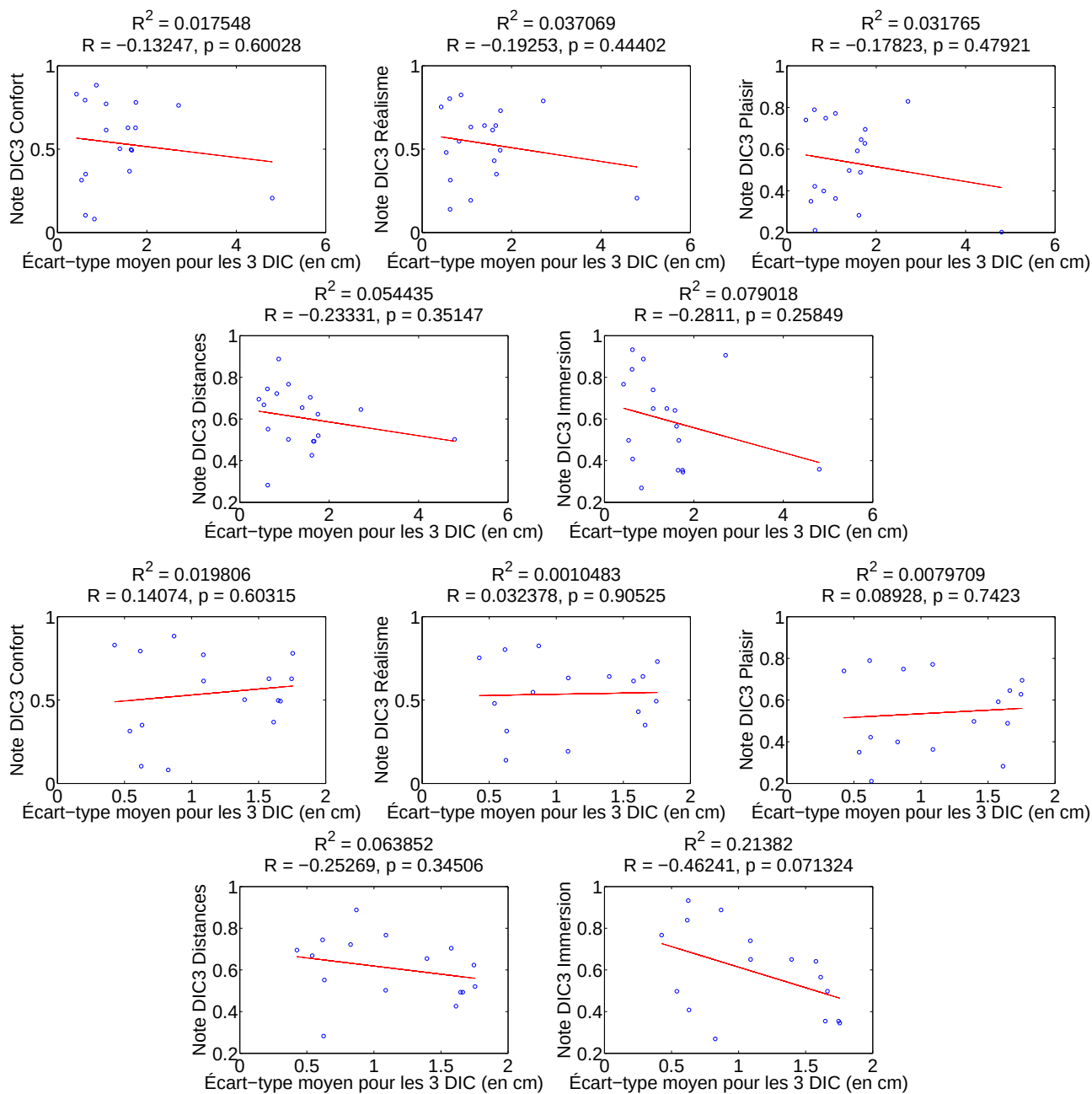


Figure D.10 Notes obtenues par **DIC3** en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).

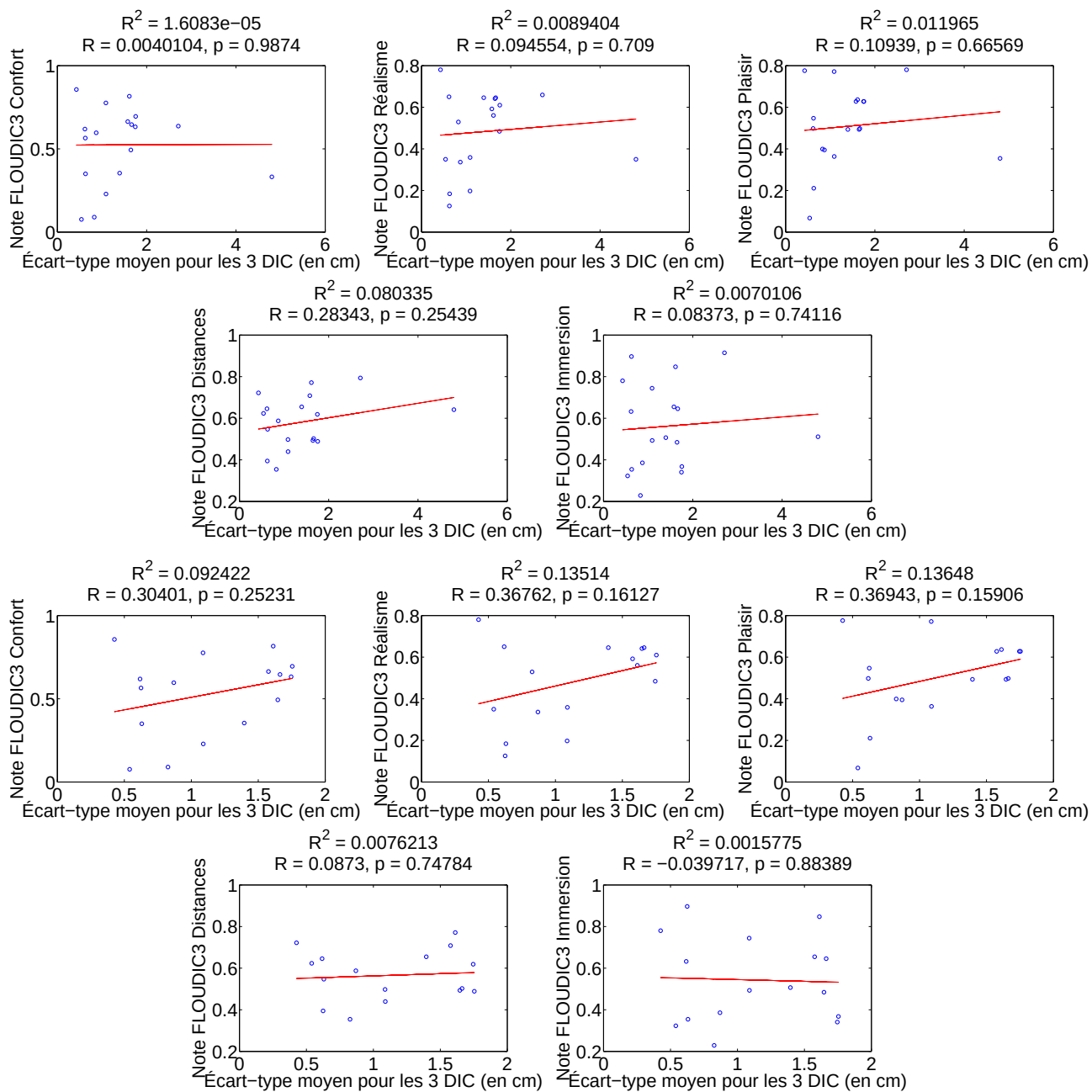


Figure D.11 Notes obtenues par **FLOUDIC3** en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).

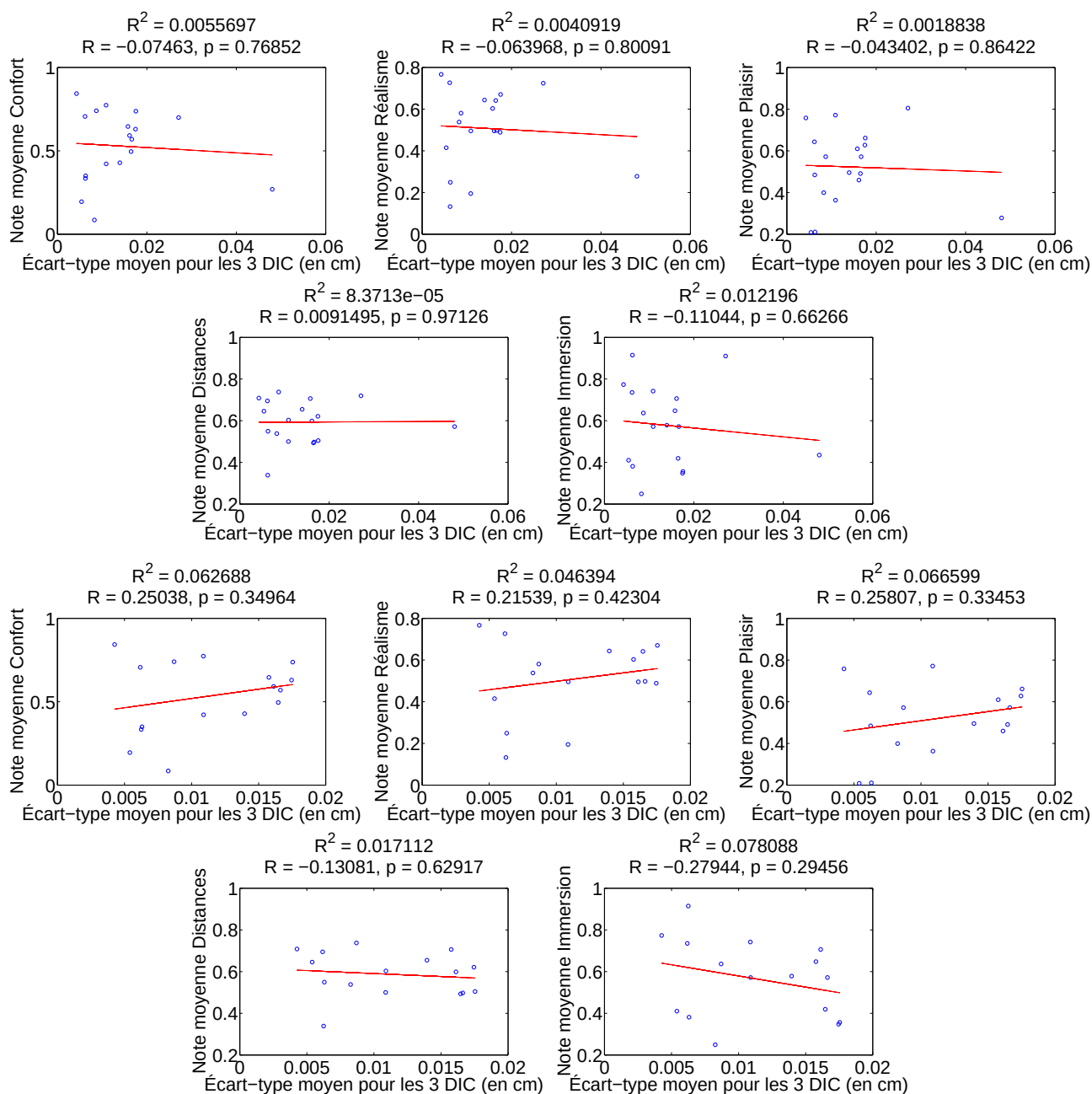


Figure D.12 Notes moyennes **DIC3** et **FLOUDIC3** en fonction de la moyenne des écarts-types des DIC3 choisies (5 graphiques du haut), puis en écartant les 2 écarts-types extrêmes (5 graphiques du bas).

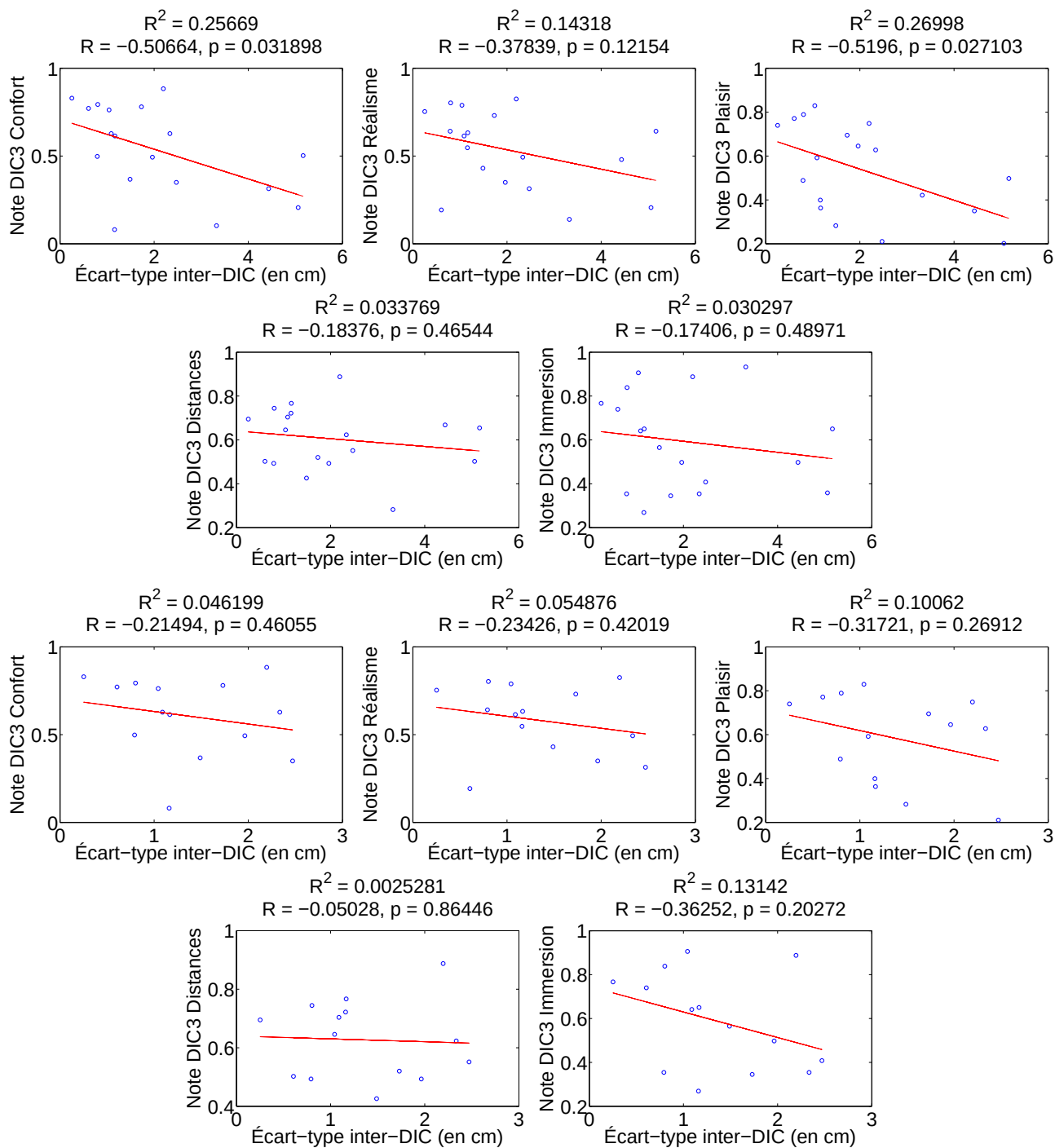


Figure D.13 Notes obtenues par **DIC3** en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).

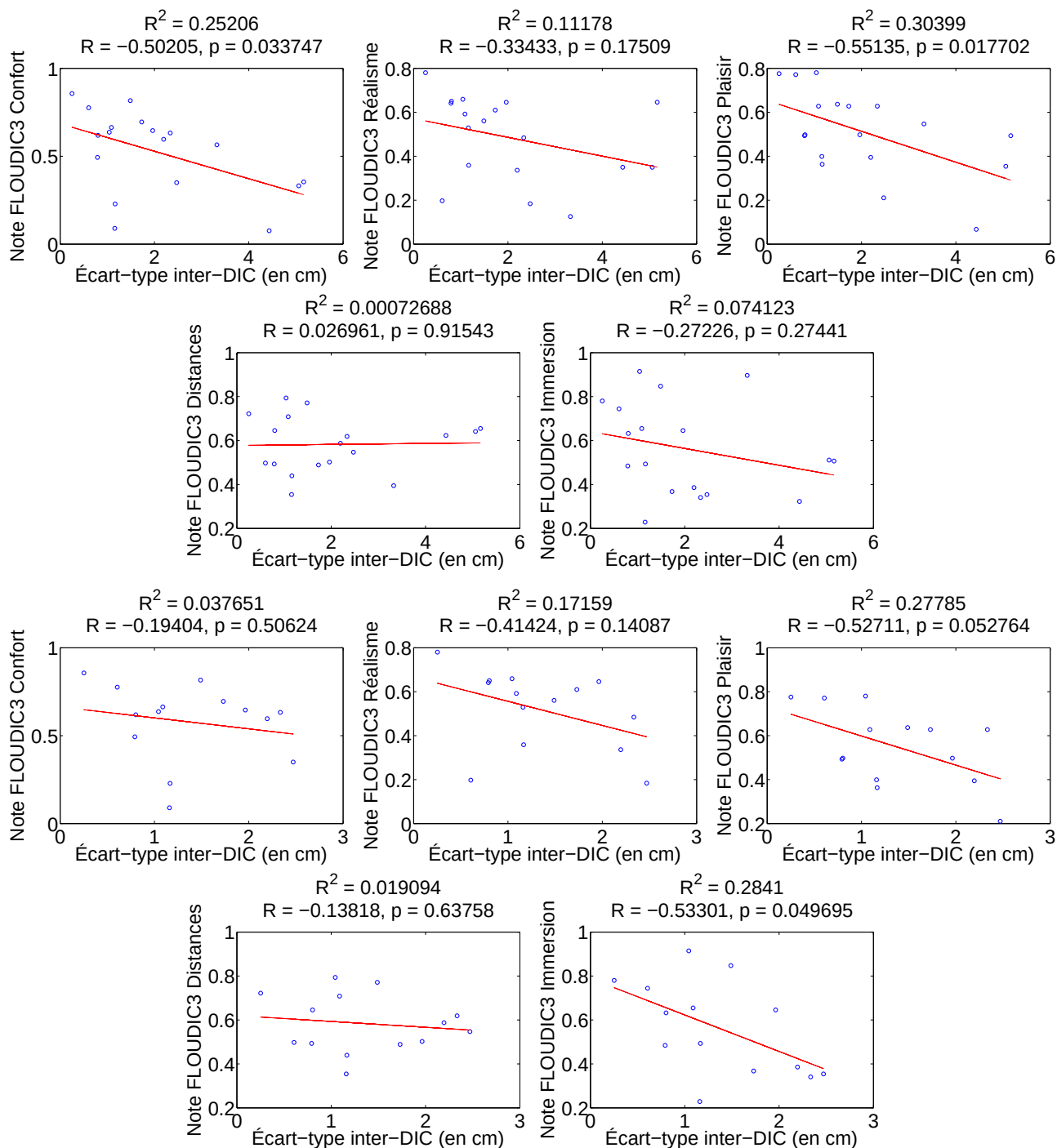


Figure D.14 Notes obtenues par **FLOUDIC3** en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).

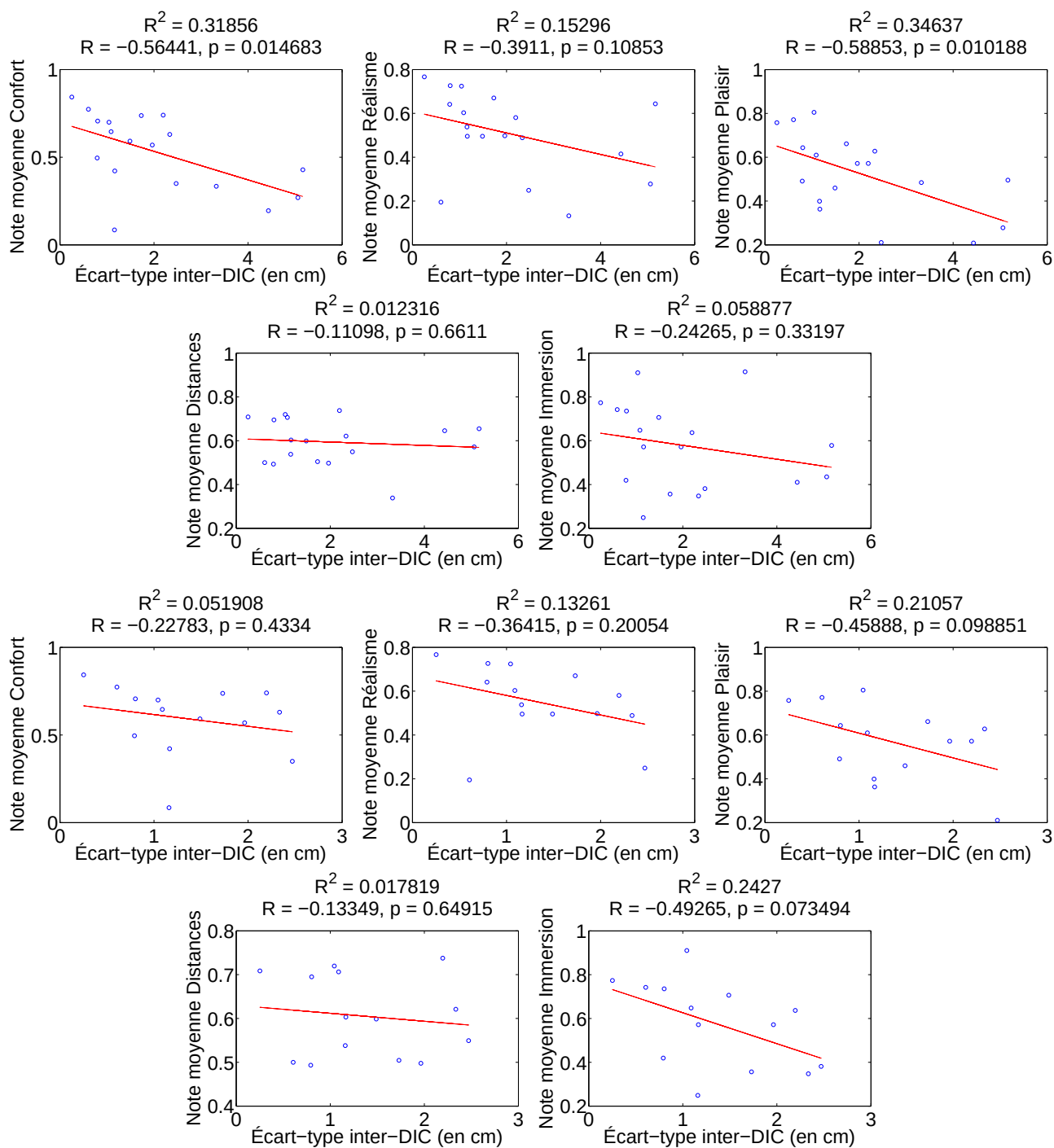


Figure D.15 Notes moyennes **DIC3** et **FLOUDIC3** en fonction de la moyenne l'écart inter-DIC3 (5 graphiques du haut), puis en écartant les 4 écarts-types extrêmes (5 graphiques du bas).

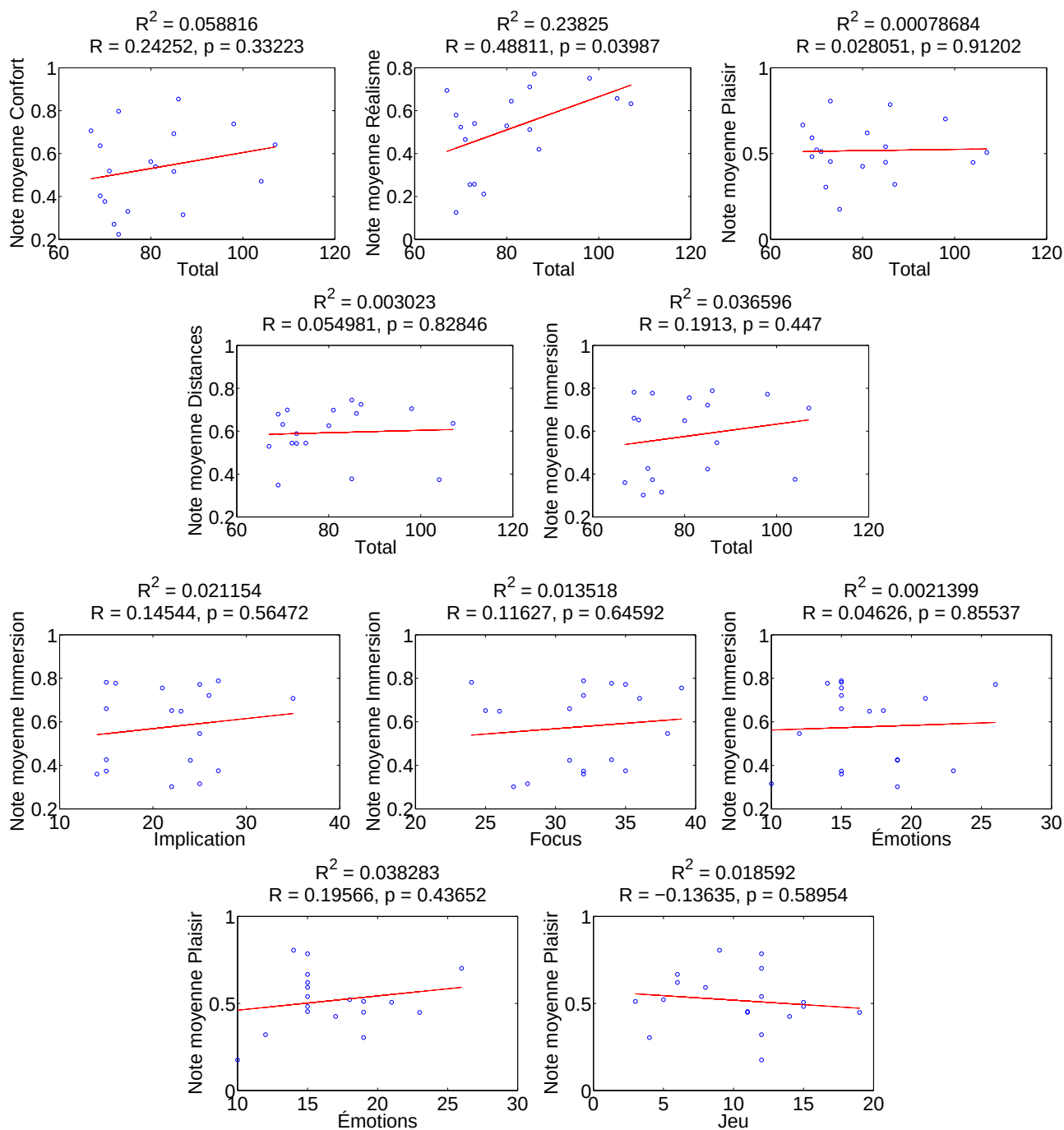


Figure D.16 Notes moyennes en fonction du score total obtenu au QPI (5 graphiques du haut). Liens implication-immersion, focus-immersion, émotion-immersion, émotion-plaisir, jeu-plaisir (5 graphiques du bas).

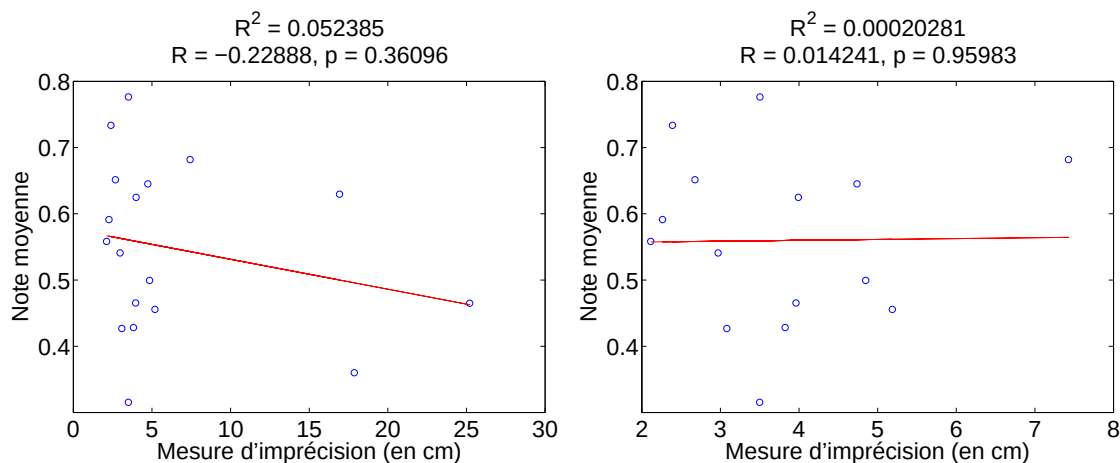


Figure D.17 Note moyenne en fonction de la mesure d'imprécision, en considérant l'ensemble des participants (à gauche) ou en écartant les 3 plus hautes imprécisions (à droite).

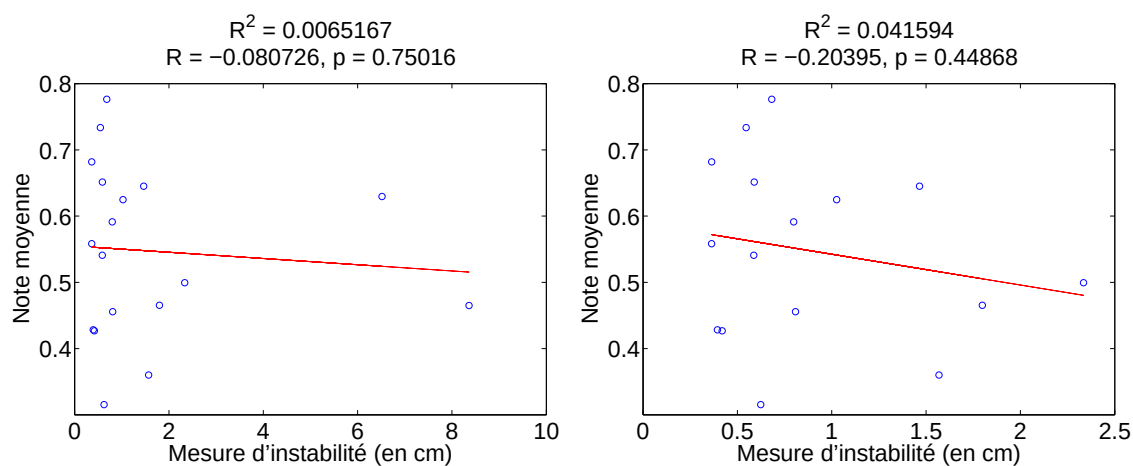


Figure D.18 Note moyenne en fonction de la mesure d'instabilité, en considérant l'ensemble des participants (à gauche) ou en écartant les 2 plus hautes instabilités (à droite).

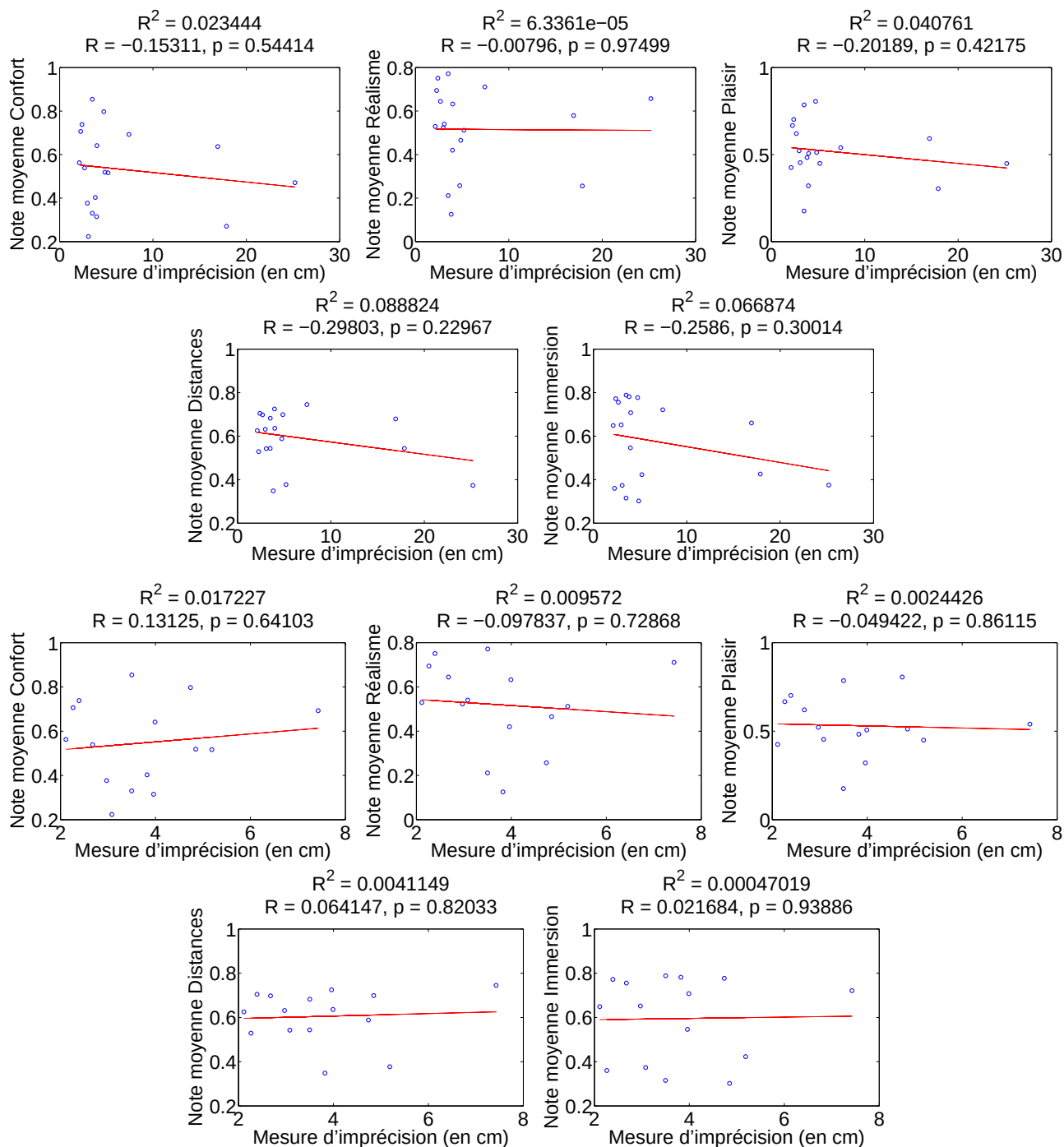


Figure D.19 Notes moyennes en fonction de la mesure d'imprécision du point de regard (5 graphiques du haut), puis en écartant les 3 plus hautes imprécisions (5 graphiques du bas).

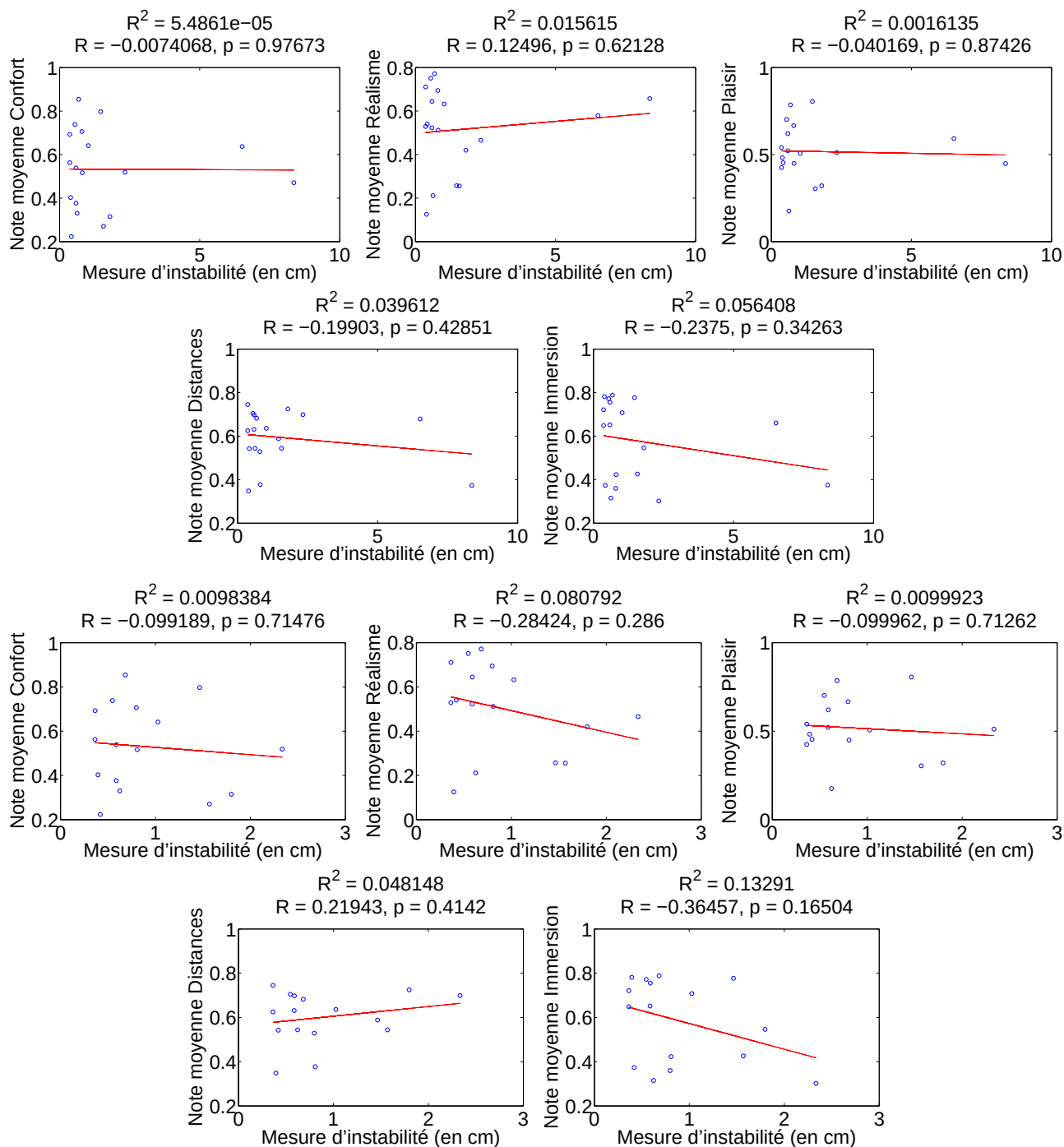


Figure D.20 Notes moyennes en fonction de la mesure d'instabilité du point de regard (5 graphiques du haut), puis en écartant les 2 plus hautes instabilités (5 graphiques du bas).

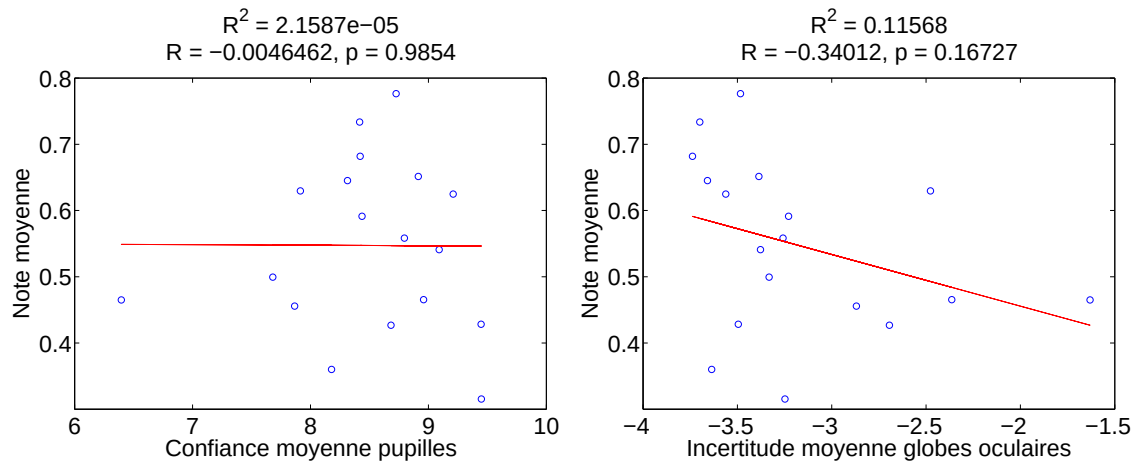


Figure D.21 Impact de l'incertitude du repérage des pupilles (à gauche) et des globes oculaires (à droite) sur les notes.

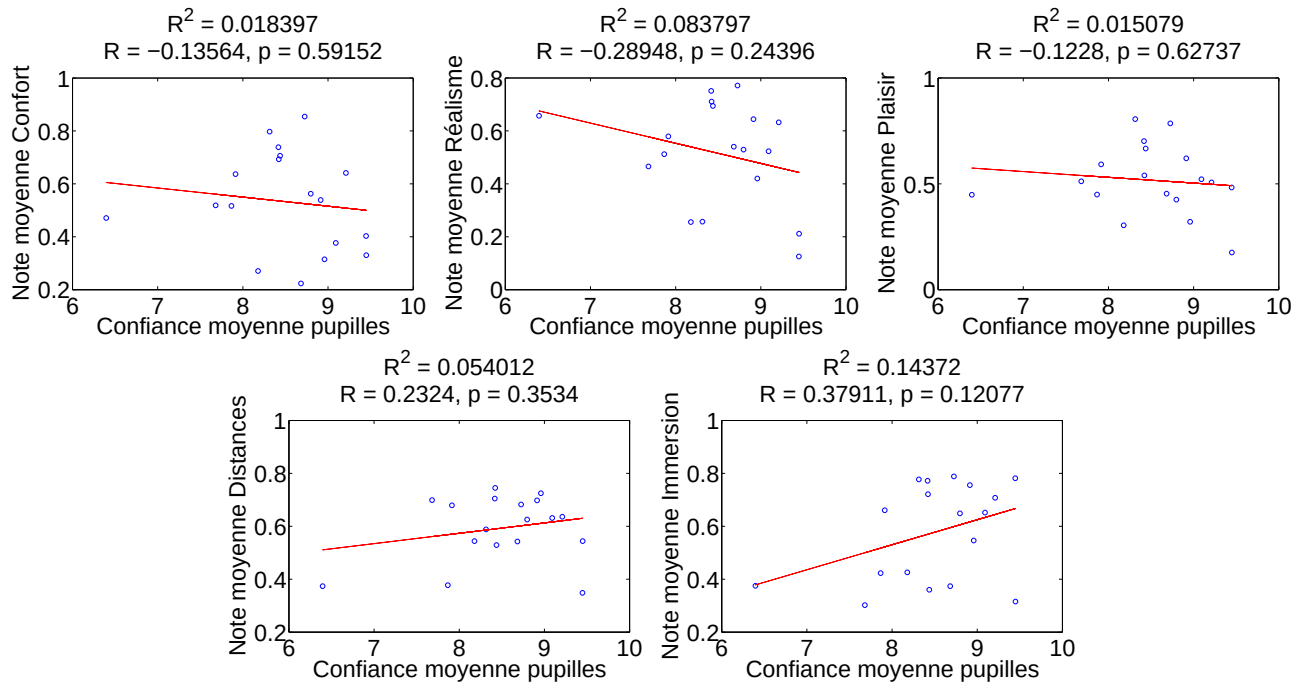


Figure D.22 Notes moyennes en fonction de l'incertitude sur la détection des pupilles.

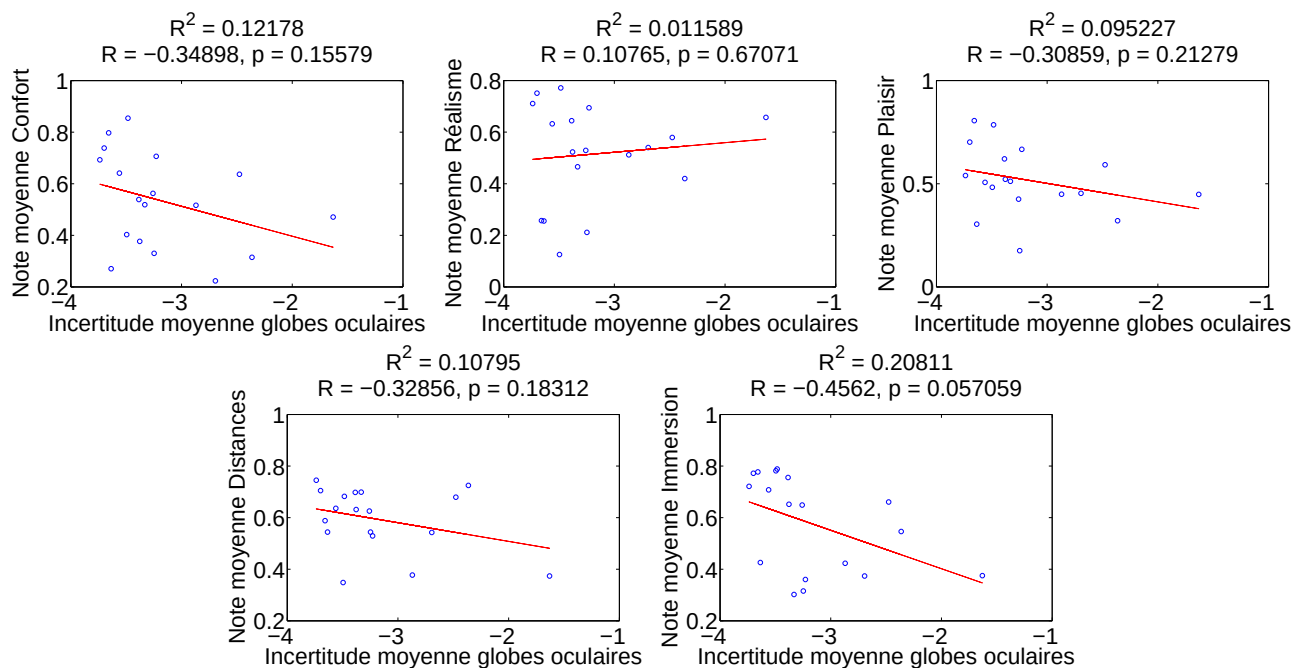


Figure D.23 Notes moyennes en fonction de l'incertitude sur la détection des globes oculaires.

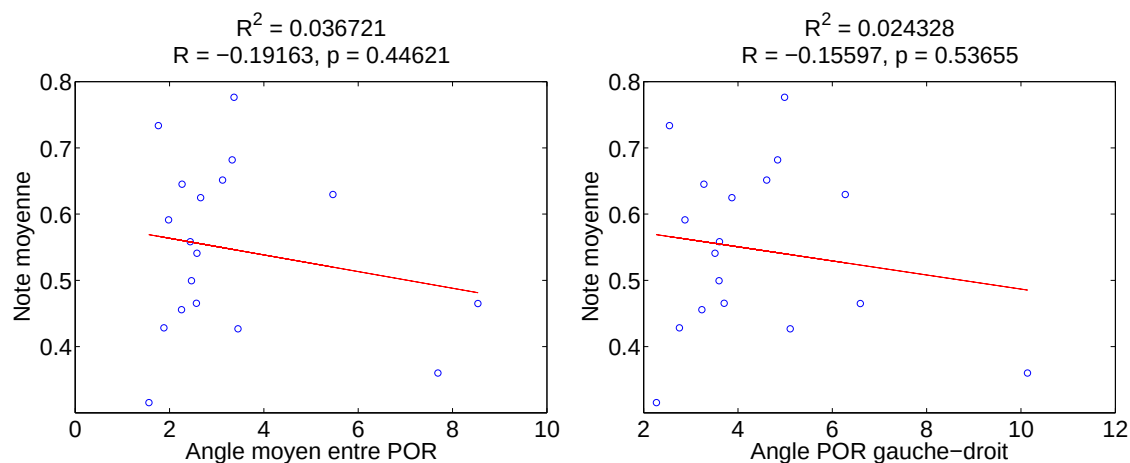


Figure D.24 Notes en fonction de l'angle moyen entre les 3 POR (à gauche) et de l'angle entre PORG et PORD (à droite).

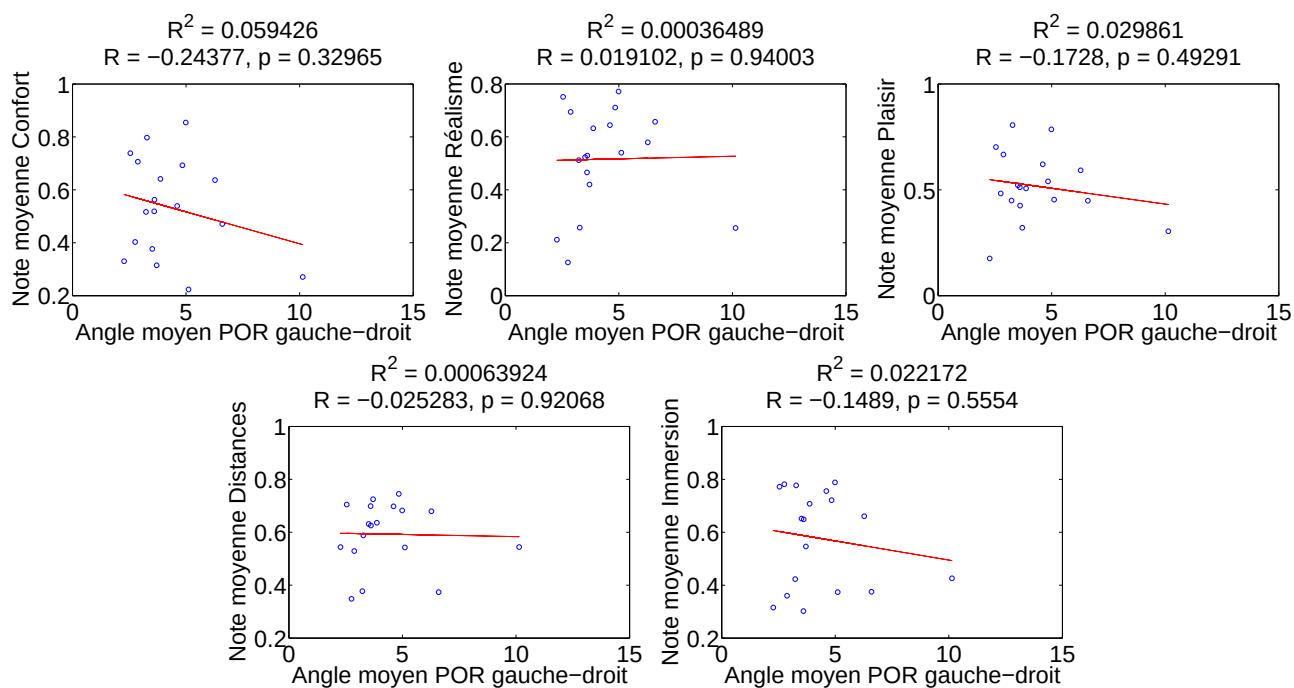


Figure D.25 Notes moyennes en fonction de l'angle entre PORG et PORD.

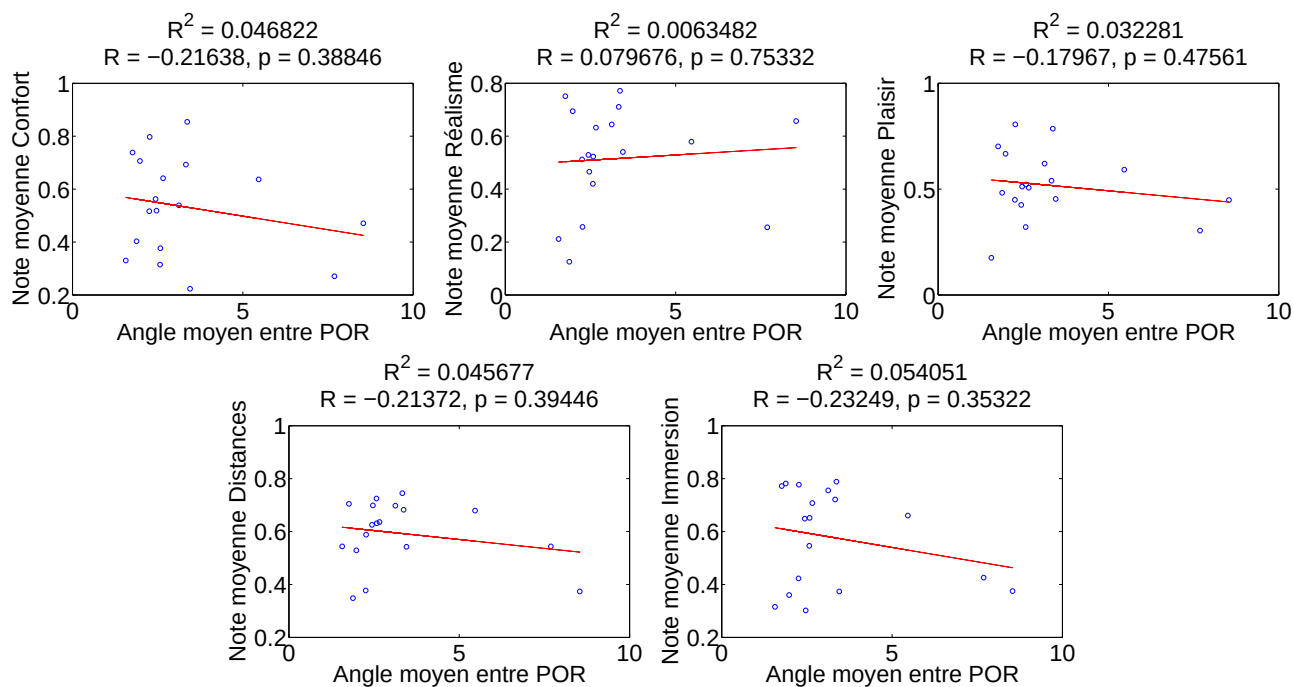


Figure D.26 Notes moyennes en fonction de la moyenne des 3 angles entre les 3 POR.

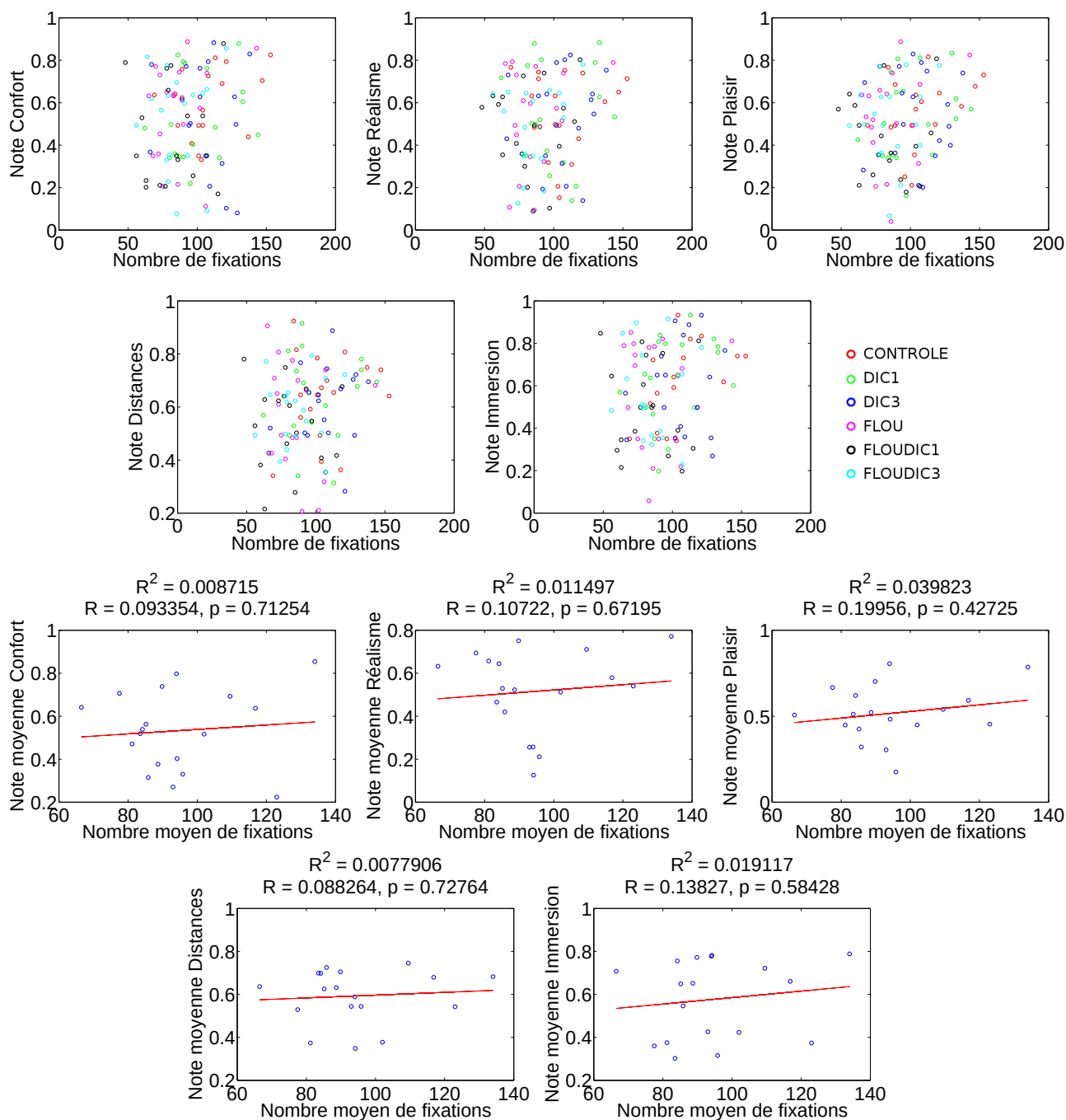


Figure D.27 Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre de fixations.

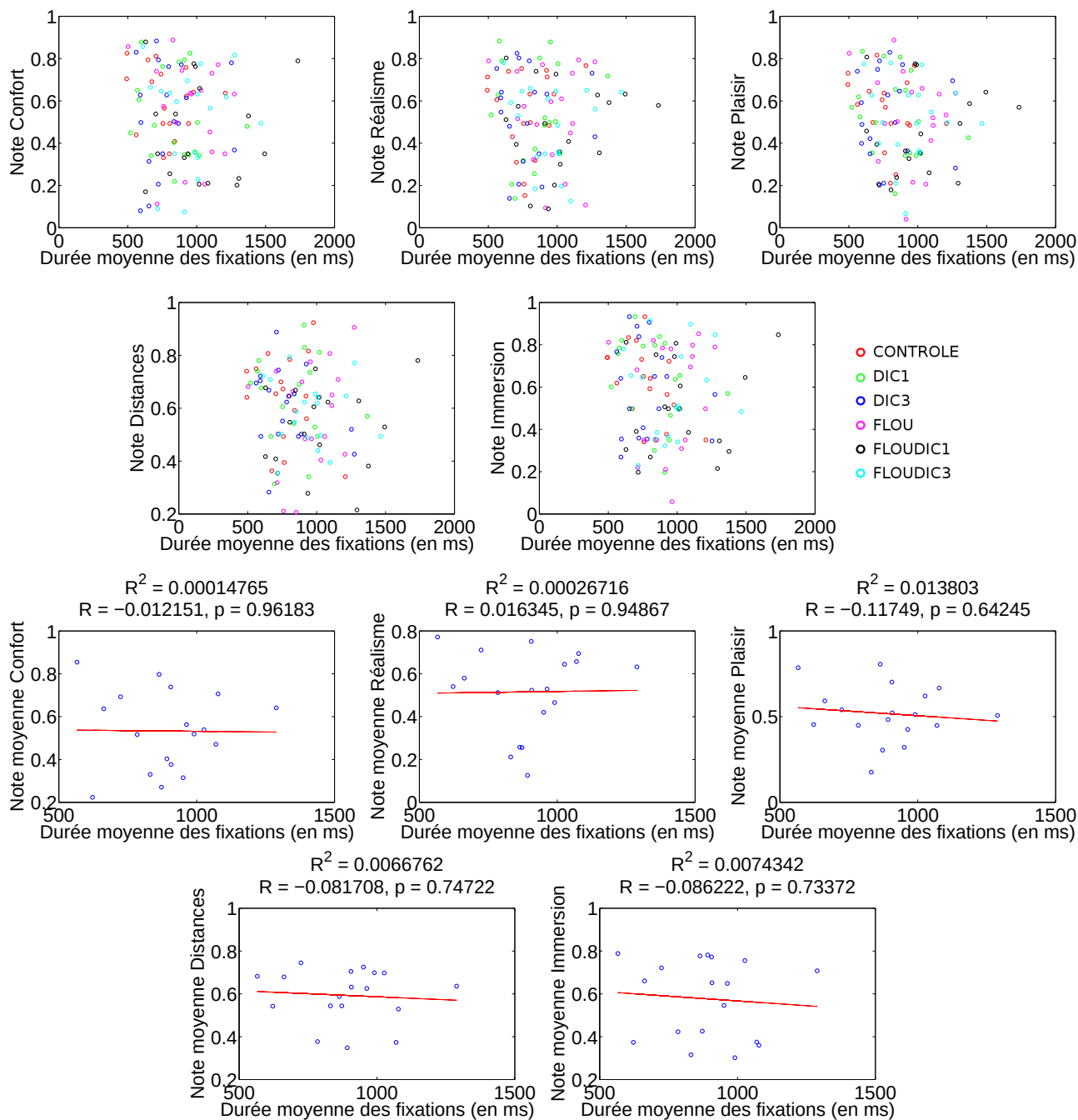


Figure D.28 Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction de la durée moyenne des fixations.

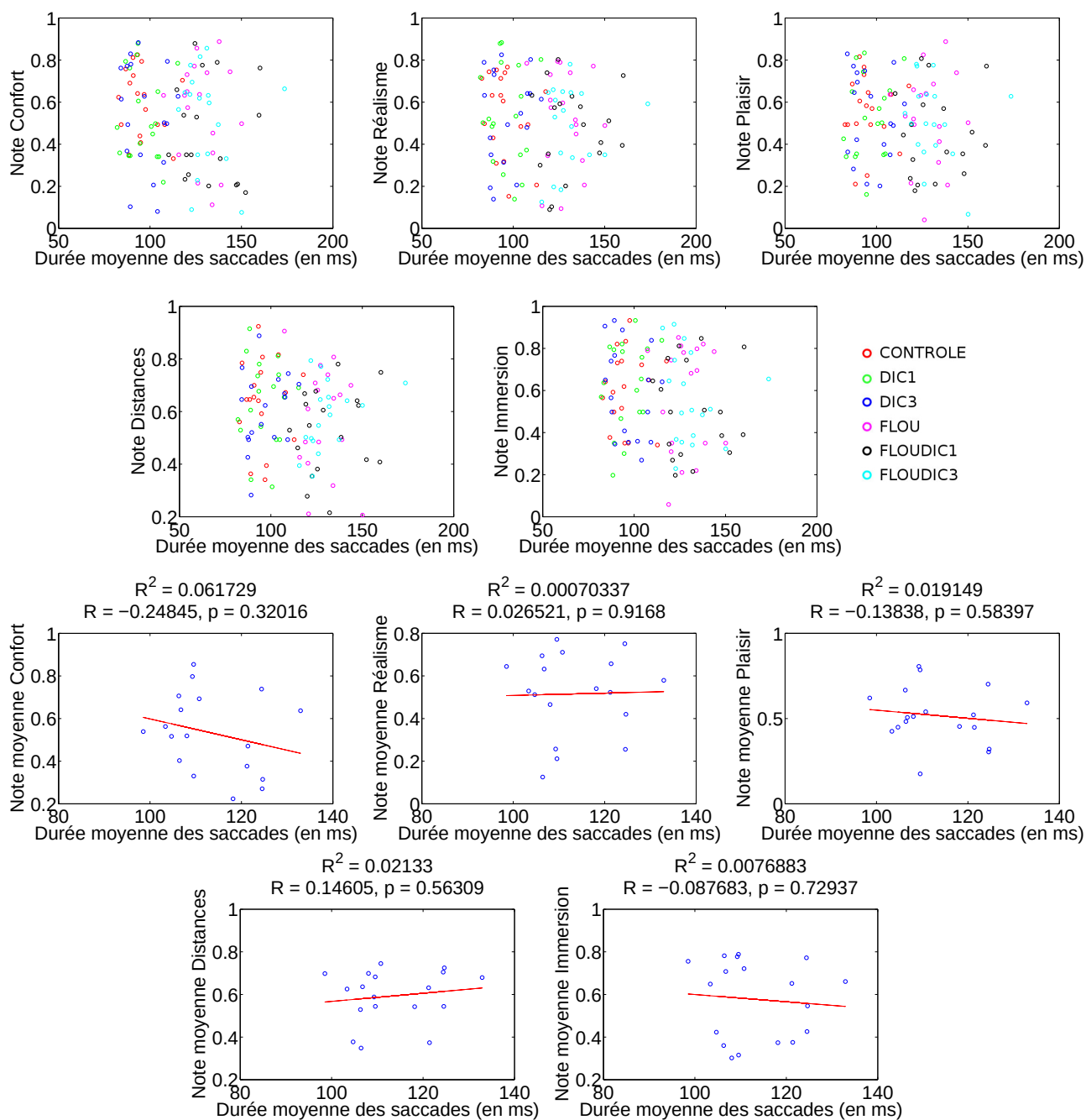


Figure D.29 Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction de la durée moyenne des saccades.

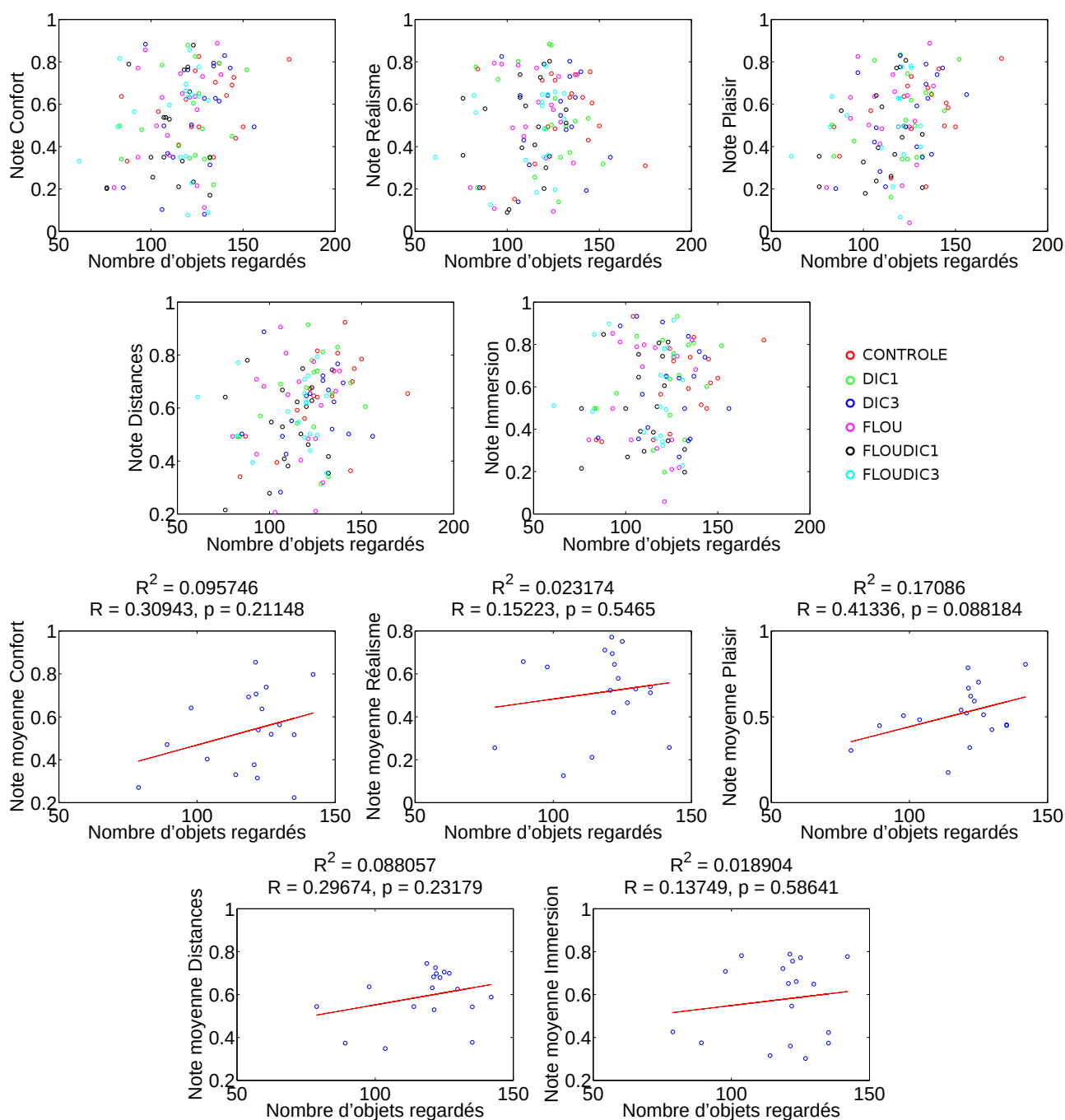


Figure D.30 Notes par configuration (5 graphiques du haut) et notes moyennes (5 graphiques du bas) en fonction du nombre d'objets regardés.

ANNEXE E

Limites de fusion

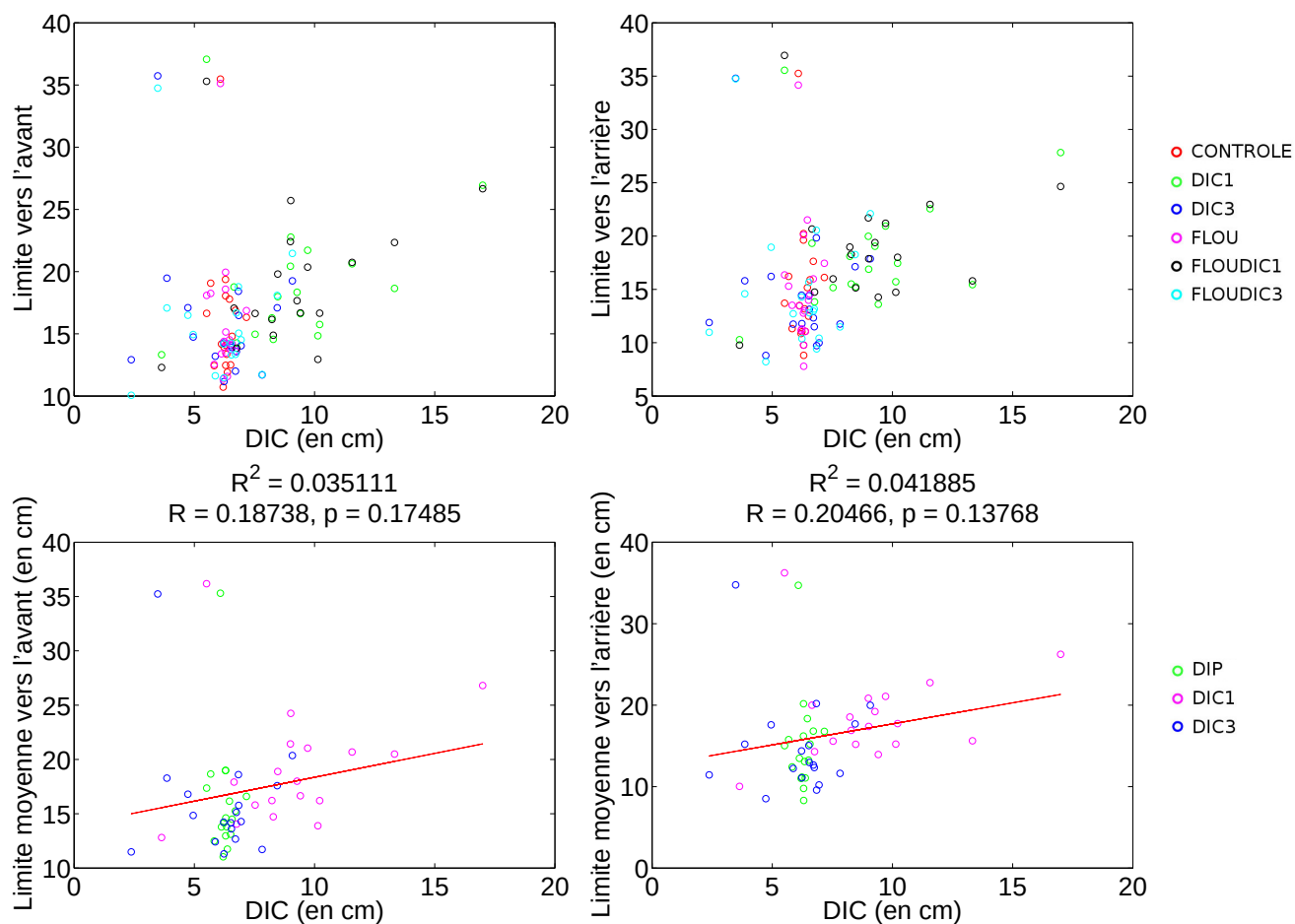


Figure E.1 Limites de fusion obtenues en fonction de la DIC, par configuration (en haut) en regroupant les configurations partageant la même DIC (en bas).

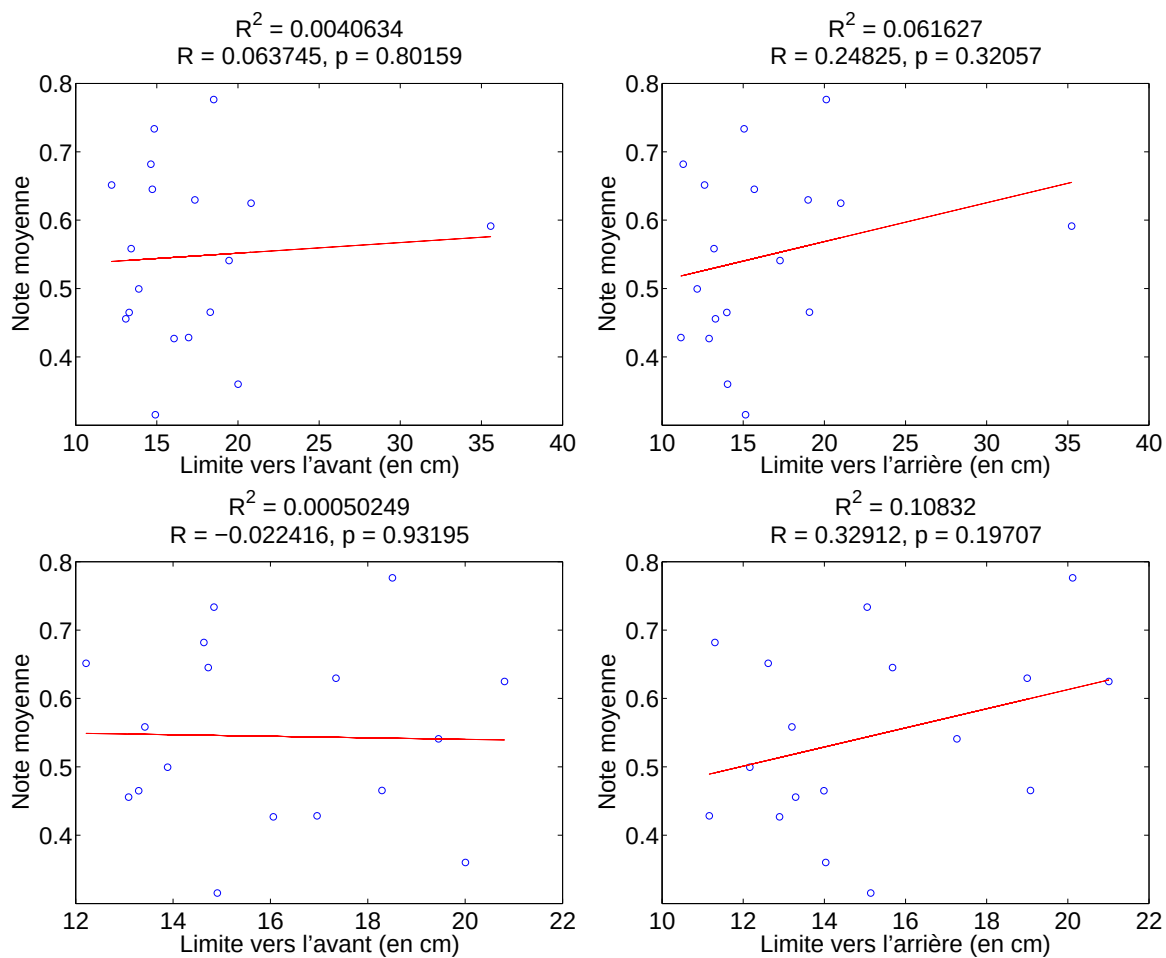


Figure E.2 Notes moyennes en fonction des limites de fusion (en haut), puis en écartant la limite de fusion extrême (en bas).

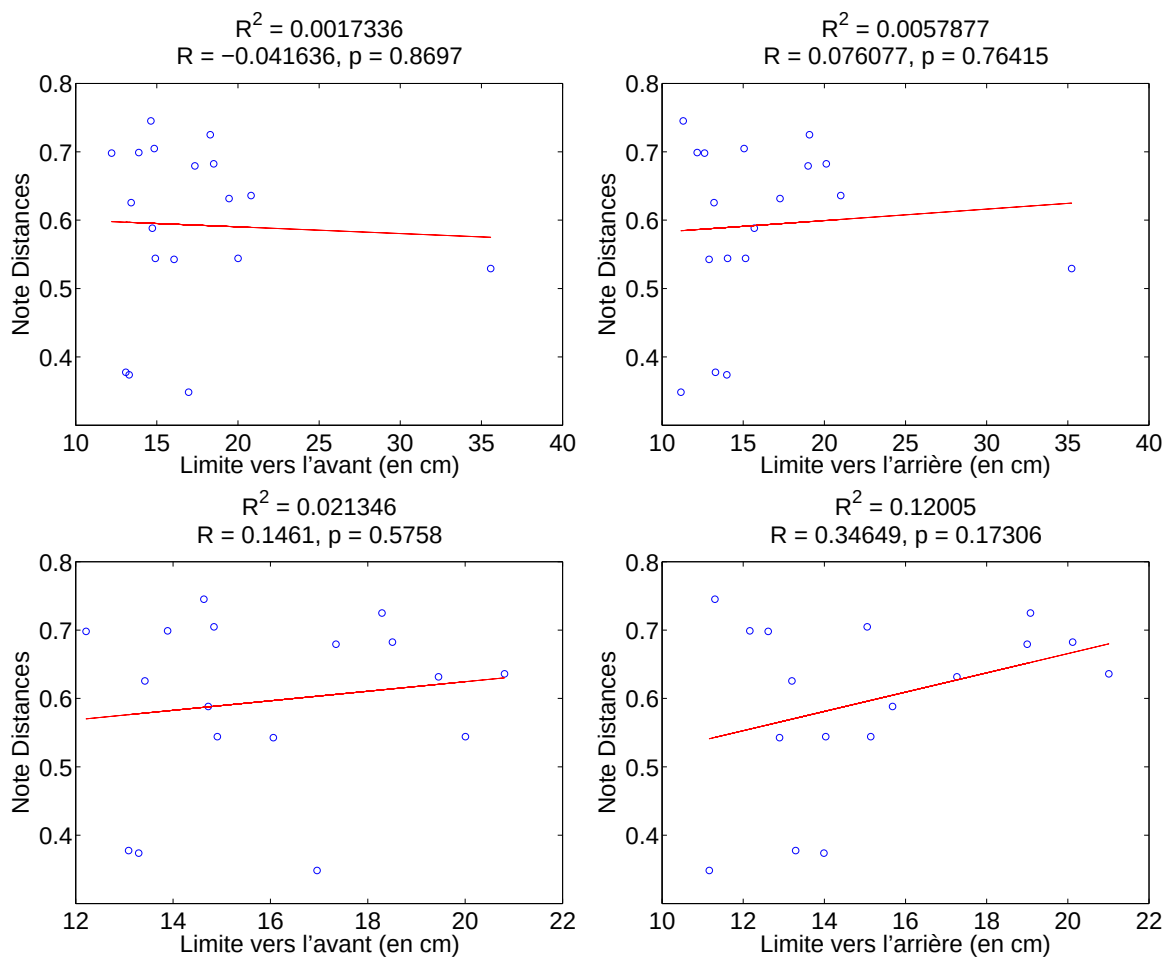


Figure E.3 Notes de perception des distances en fonction des limites de fusion (en haut), puis en écartant la limite de fusion extrême (en bas).

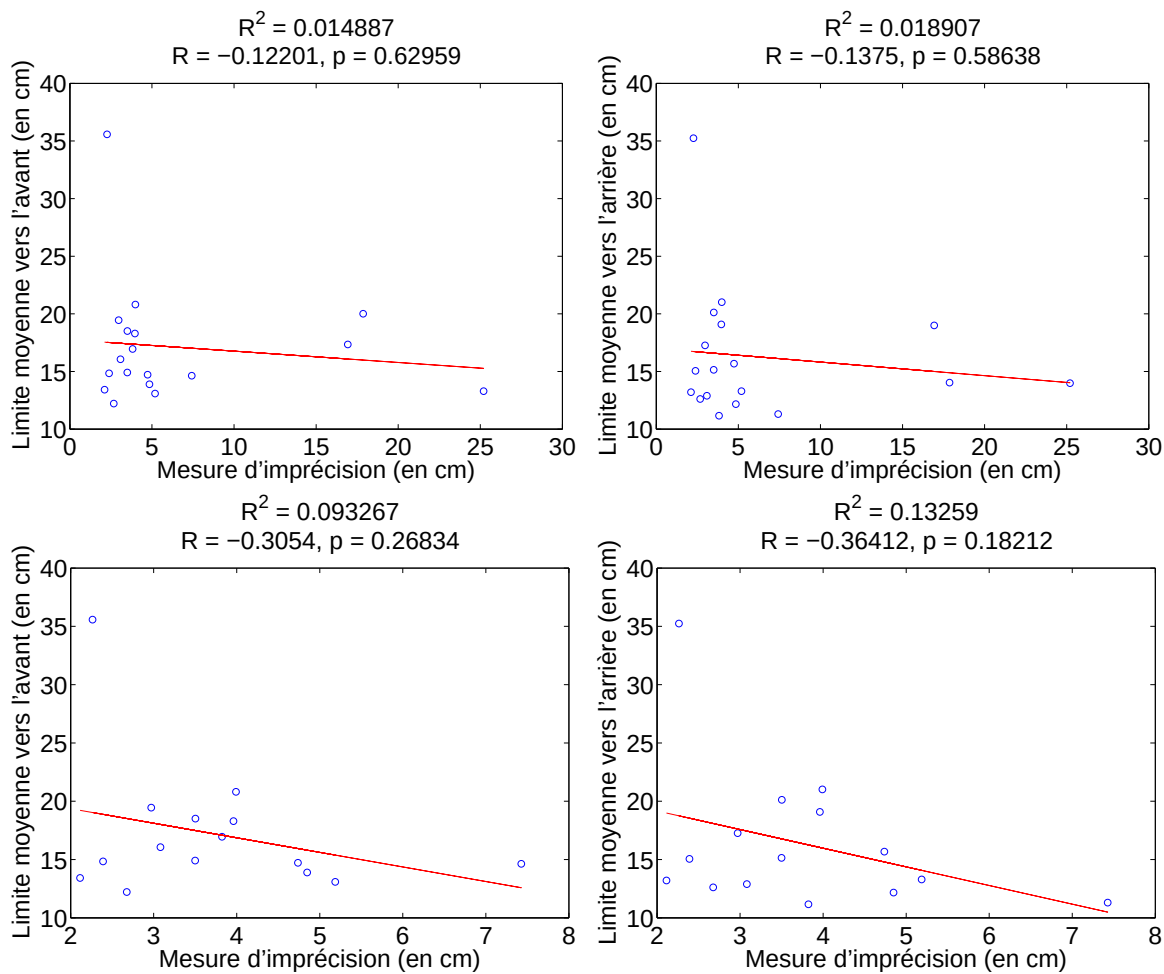


Figure E.4 Limites obtenues en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).

ANNEXE F

Comparaison des mouvements oculaires

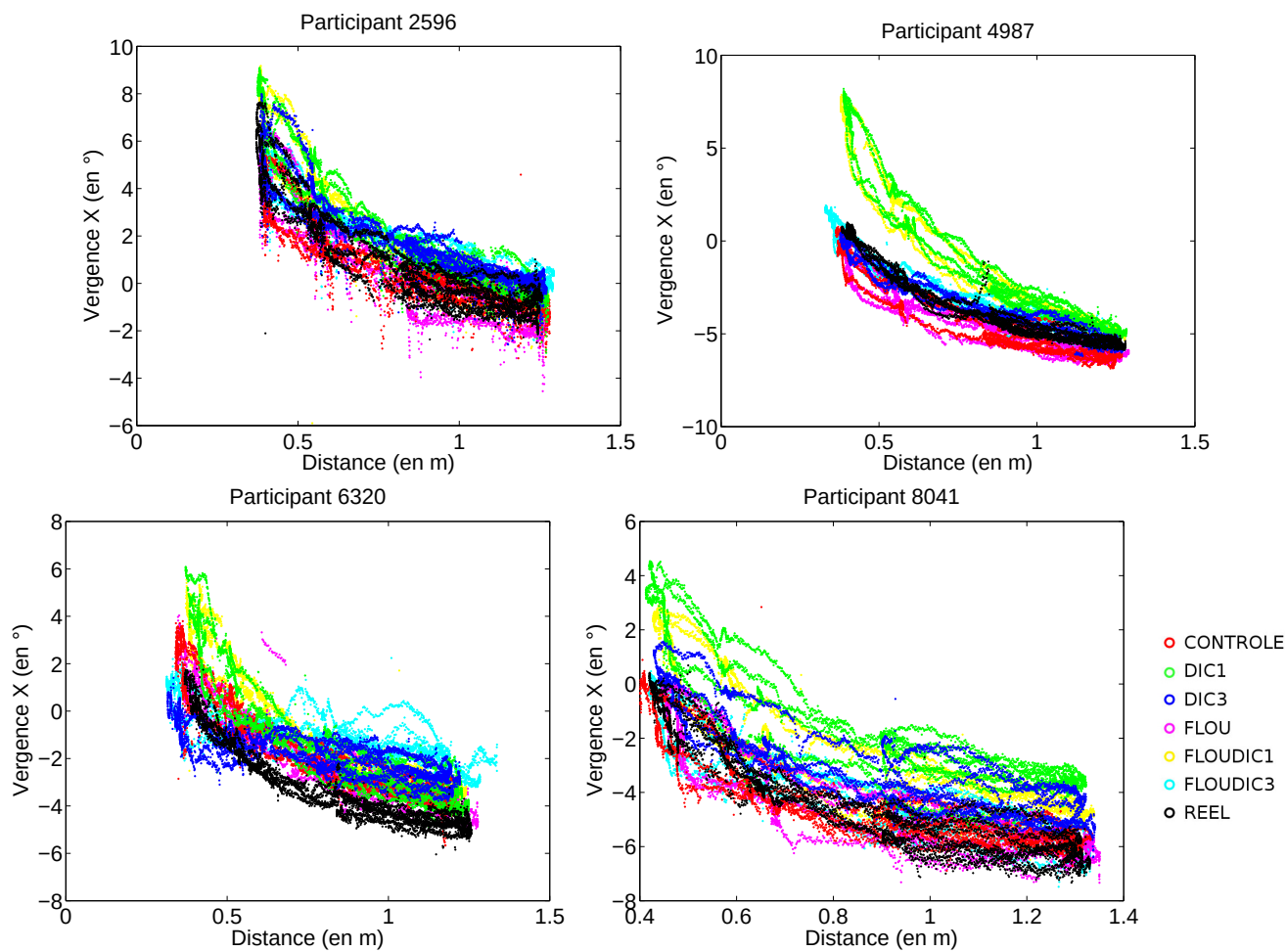


Figure F.1 Exemples de graphiques représentant la vergence en X en fonction de la distance montrant les angles supérieurs induits par les configurations **DIC1** et **FLOUDIC1**.

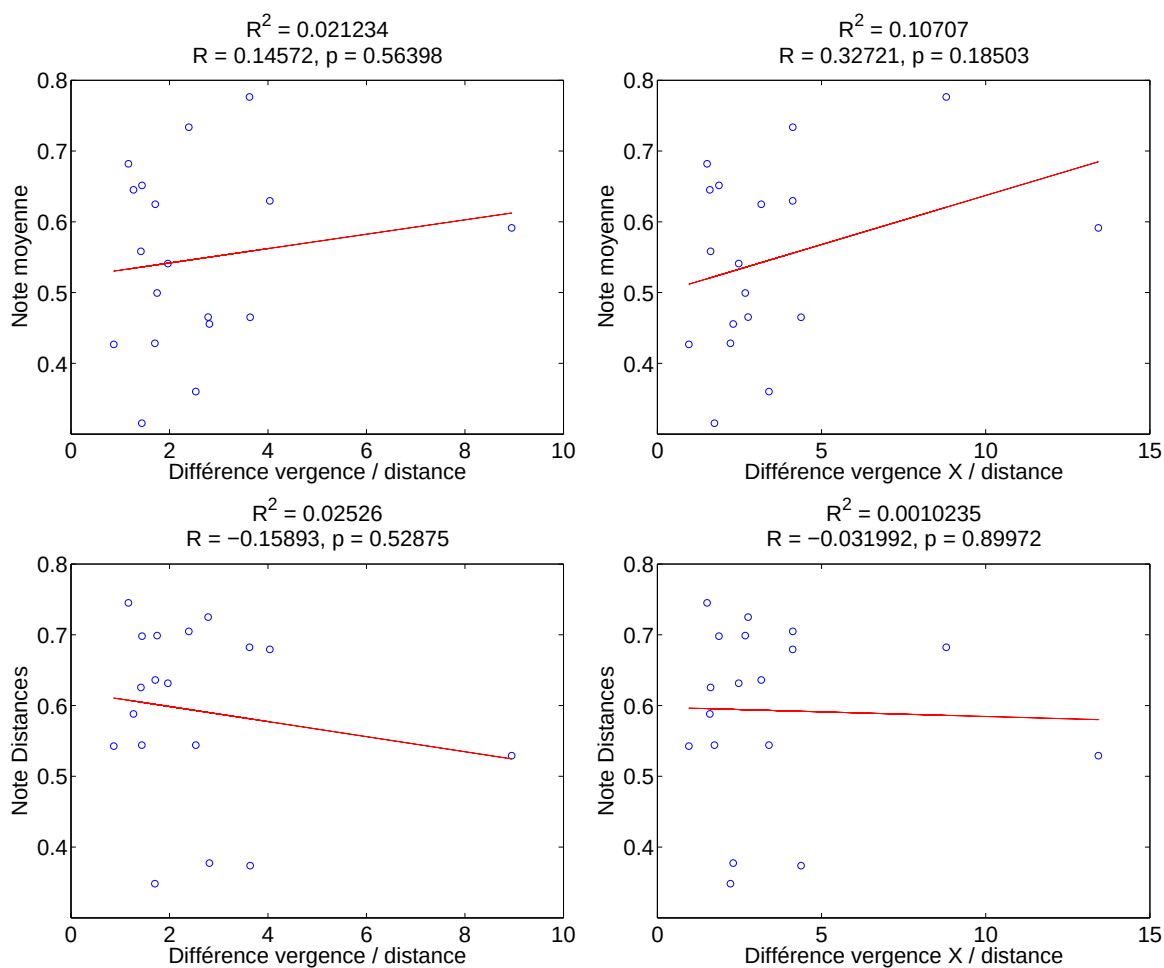


Figure F.2 Notes moyennes (en haut) et de perception des distances (en bas) en fonction de la différence de vergence/distance et vergence en X/distance par rapport à la configuration réelle.

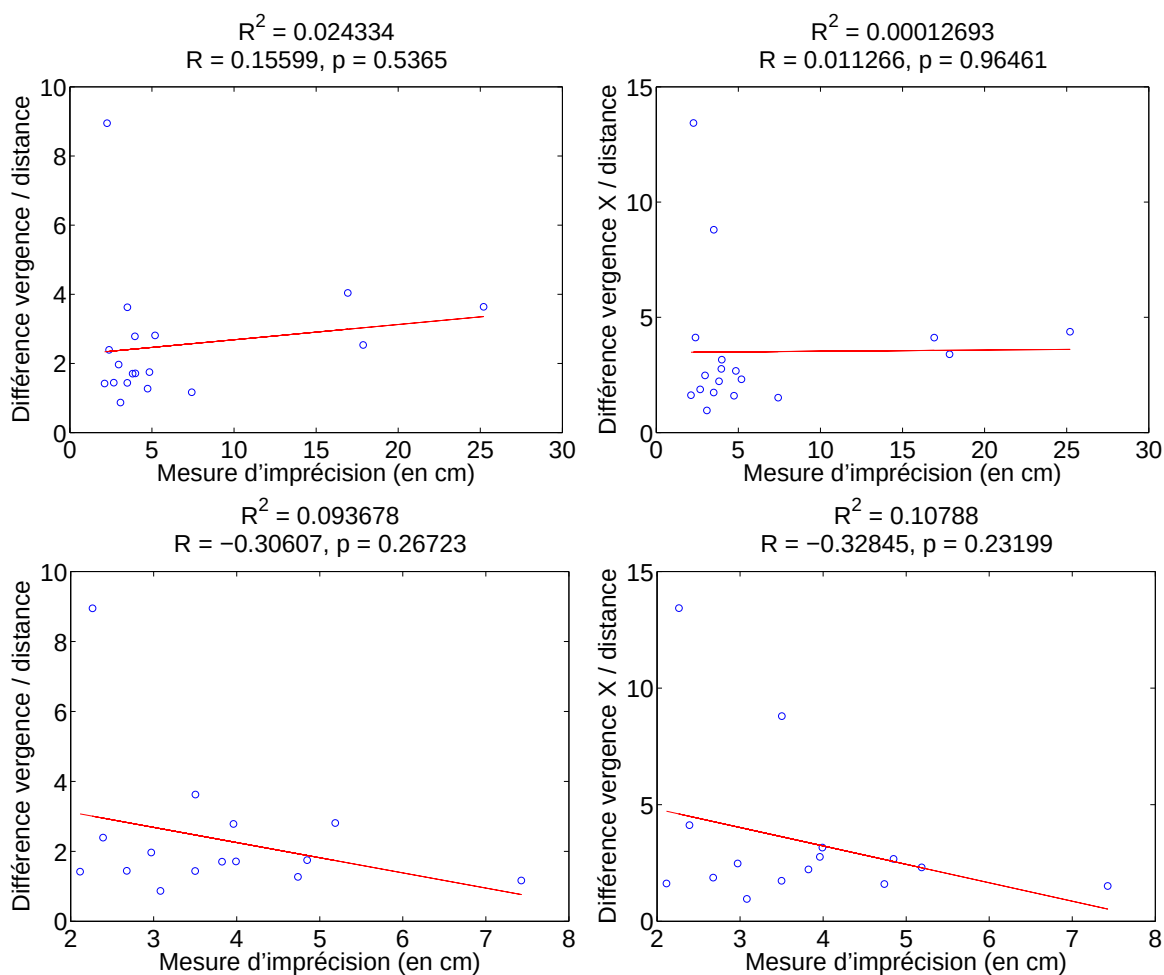


Figure F.3 Différence de vergence/distance et vergence en X/distance par rapport à la configuration réelle en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).

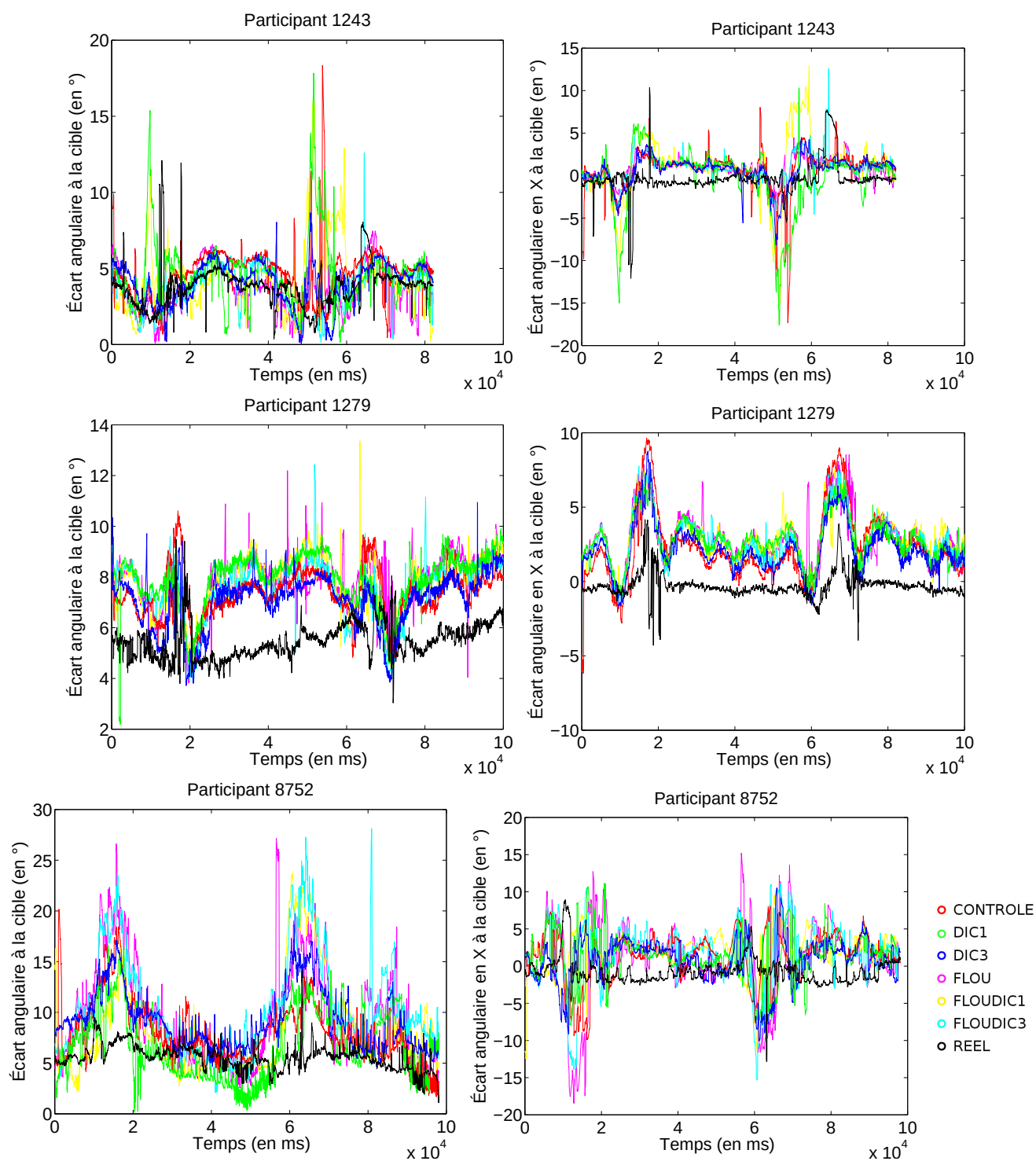


Figure F.4 Exemples de graphiques représentant les écarts angulaires total et en X à la cible en fonction du temps, montrant un écart angulaire tournant autour de 0 pour la configuration réelle.

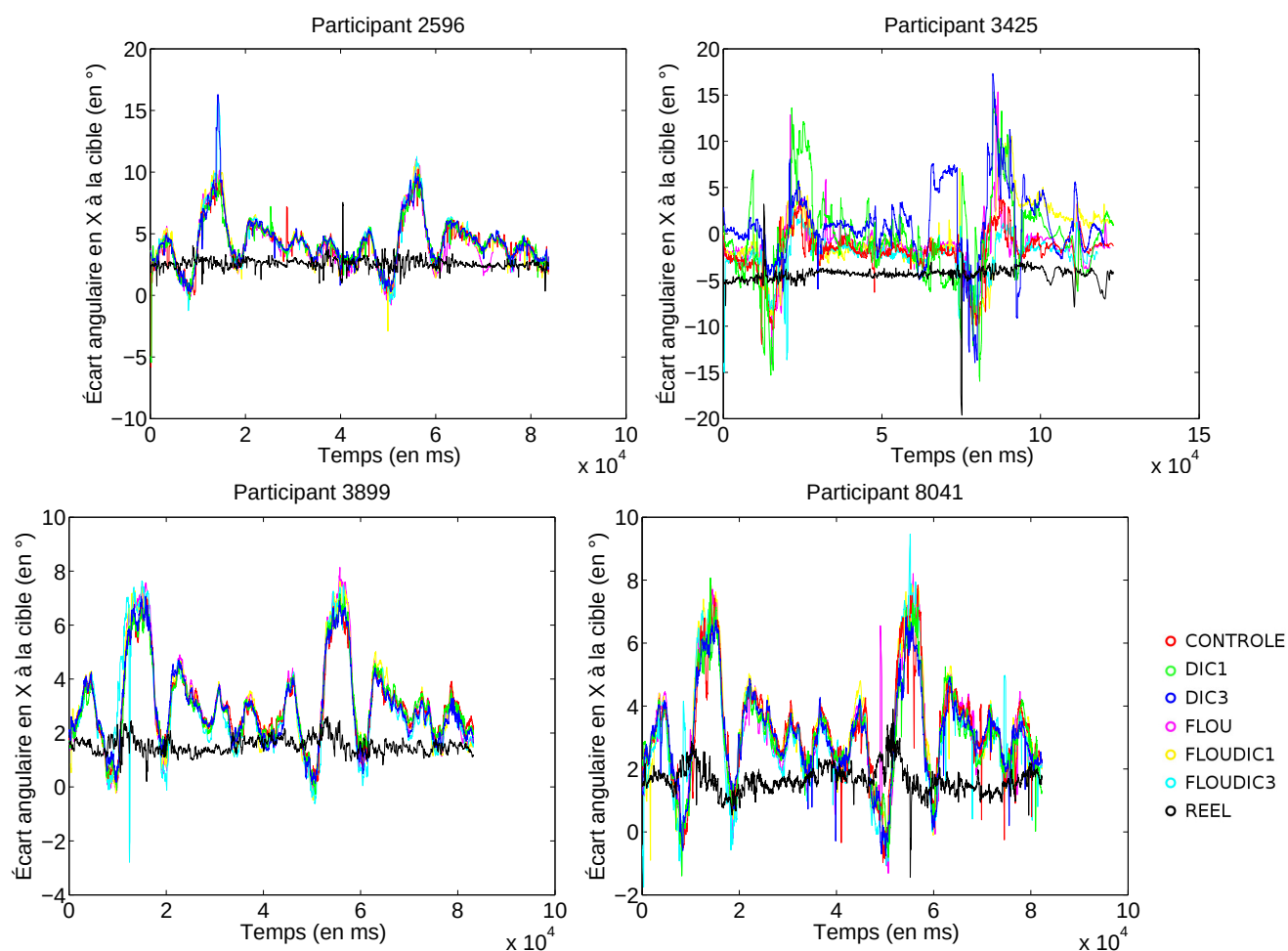


Figure F.5 Exemples de graphiques représentant les écarts angulaires total et en X à la cible en fonction du temps, montrant un écart angulaire différent de 0 pour la configuration réelle.

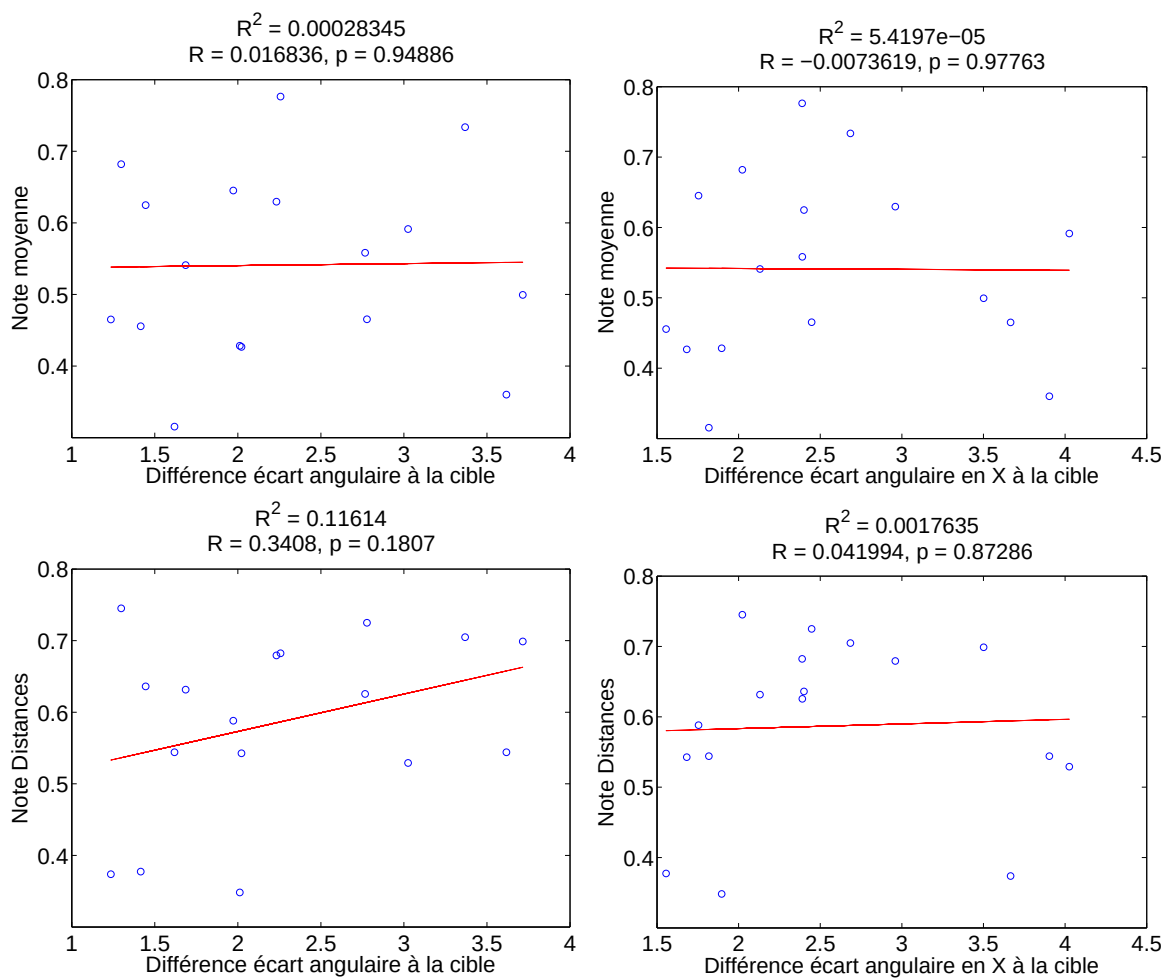


Figure F.6 Notes moyennes (en haut) et de perception des distances (en bas) en fonction de la différence d'écarts angulaires total et en X à la cible par rapport à la configuration réelle.

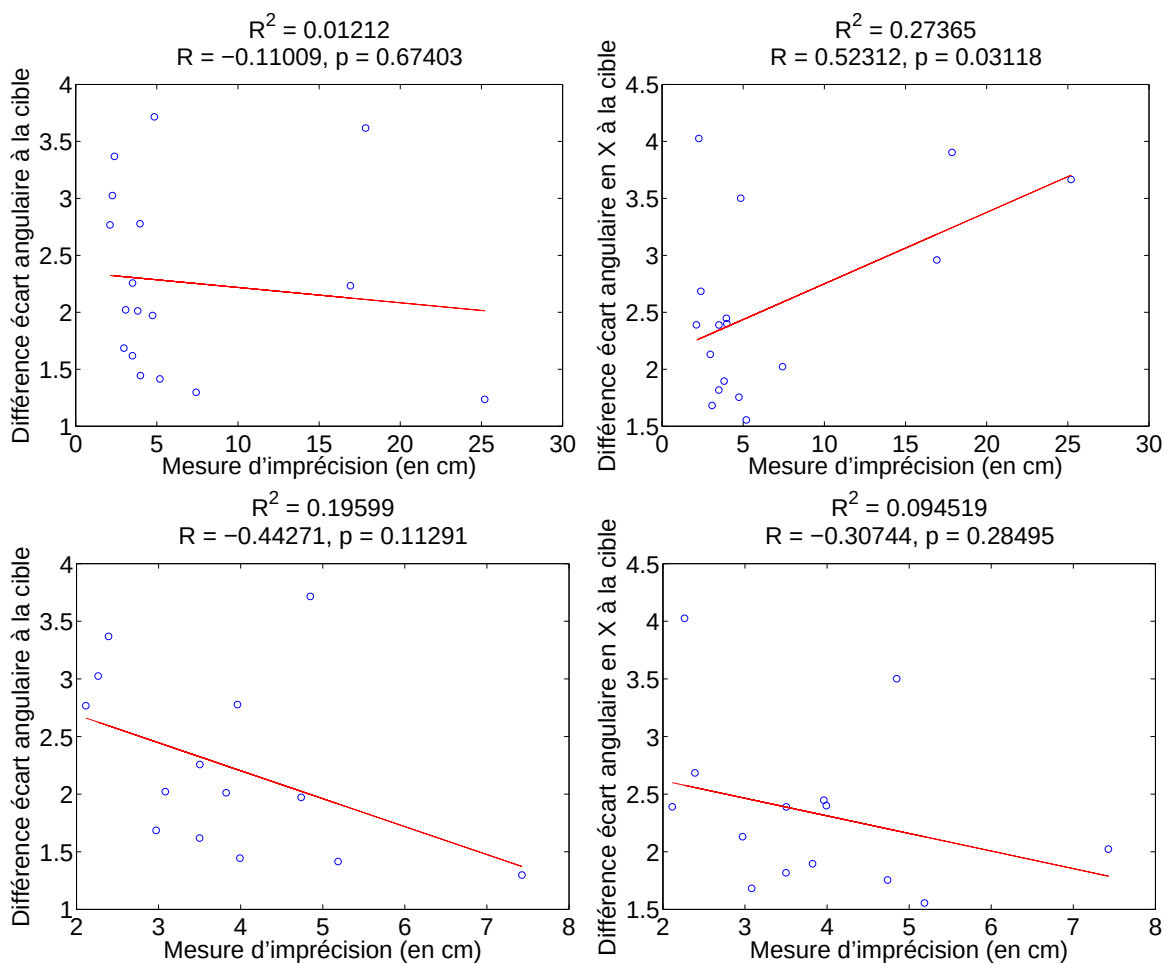


Figure F.7 Différence d'écarts angulaires total et en X à la cible par rapport à la configuration réelle en fonction de la précision du point de regard (en haut), puis en écartant les 3 plus mauvaises précisions (en bas).