



**HAL**  
open science

# Conception et évaluation d'un système transparent de capture de mouvements des mains pour l'interaction 3D temps réel en environnements virtuels

Franck Hernoux

► **To cite this version:**

Franck Hernoux. Conception et évaluation d'un système transparent de capture de mouvements des mains pour l'interaction 3D temps réel en environnements virtuels. Interface homme-machine [cs.HC]. Arts et Métiers ParisTech; Ecole nationale supérieure d'arts et métiers - ENSAM, 2011. Français. NNT : 2011ENAM0039 . pastel-00651084

**HAL Id: pastel-00651084**

**<https://pastel.hal.science/pastel-00651084>**

Submitted on 12 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École doctorale n° 432 : Sciences des Métiers de l'Ingénieur

**Doctorat ParisTech**

**T H È S E**

pour obtenir le grade de docteur délivré par

**l'École Nationale Supérieure d'Arts et Métiers**

**Spécialité " Informatique – Traitement du signal (AM) "**

*présentée et soutenue publiquement par*

**Franck HERNOUX**

le 7 décembre 2011

**Conception et évaluation d'un système transparent de capture de  
mouvements des mains pour l'interaction 3D temps réel en  
environnements virtuels**

Directeur de thèse : **Simon RICHIR**  
Co-encadrement de la thèse : **Olivier CHRISTMANN**

**Jury**

**M. Daniel Mestre**, Directeur de Recherche, CNRS  
**M. Guillaume Moreau**, Professeur des Universités, CNRS, Ecole Centrale de Nantes  
**Mme Sabine Coquillart**, Directrice de Recherche, INRIA Rhône-Alpes  
**M. Alexis Paljic**, Chargé de Recherche, Centre de Robotique, Mines ParisTech  
**M. Lionel Dominjon**, Docteur, Clarté  
**M. Simon Richir**, Professeur, LAMPA, Arts et Métiers ParisTech Angers  
**M. Olivier Christmann**, Maître de Conférences, LAMPA, Arts et Métiers ParisTech Angers

Rapporteur  
Rapporteur  
Examinatrice  
Examineur  
Examineur  
Examineur  
Examineur



*À mes parents et mon frère*



*“It would appear that we have reached the limits of what it is possible to achieve with computer technology, although one should be careful with such statements, as they tend to sound pretty silly in 5 years.”*

John Von Neumann (1903-1957)

*« Il semblerait que nous ayons atteint les limites du possible avec la technologie informatique, bien qu'il faille rester prudent sur de telles déclarations, car elles ont tendance à paraître assez ridicules 5 ans plus tard. »*



## Remerciements

Je souhaite, tout d'abord, remercier Simon Richir de m'avoir permis de faire cette thèse au sein de l'équipe Présence et Innovation et d'avoir accepté d'en être mon directeur.

Je tiens, par ailleurs, à remercier vivement messieurs Mestre et Moreau d'avoir accepté d'être les rapporteurs de ce travail et d'avoir consacré de leur temps précieux à son évaluation. Merci également à Madame Coquillart et à messieurs Paljic et Dominjon pour leur participation au jury de soutenance.

Un merci particulier à Olivier Christmann pour son amitié et toute l'aide précieuse qu'il a pu m'apporter dans la réalisation de cette thèse, qui ne serait pas ce qu'elle est aujourd'hui sans son implication, sa disponibilité et ses conseils avisés. Merci d'avoir été là quand j'en avais besoin, pour sa bonne humeur et sa gentillesse au quotidien, on ne peut rêver d'un meilleur ami !

Je remercie également tous les membres de l'équipe Présence et Innovation, avec qui j'ai eu la chance de travailler, avec une pensée particulière pour les doctorants : Daliah, Jayesh, Pierrick et Erik (merci pour la modélisation 3D de la main) ainsi que pour Elise et Majid de l'équipe HIT. Un merci spécial à Emilie avec qui j'ai passé d'excellents moments (plus que je ne saurai les compter) : pour sa joie de vivre, ses petites crises d'angoisse et sa douce folie qui font d'elle quelqu'un de vraiment unique ! Merci aussi à Gill (qui m'a beaucoup dépanné avec Virtools !) et à Guillaume pour tous les moments agréables, les soirées et les repas du midi partagés ensemble. Merci à tous d'avoir passé ces 3 années de thèse avec moi et pour l'excellente ambiance de travail.

Je souhaite également remercier l'agglomération de Laval qui a financé ma dernière année de thèse.

Je tiens, par ailleurs, à adresser mes remerciements à M. Gracanin, directeur du « Distributed Virtual Environments Research Laboratory » du département « Computer Science » de Virginia Tech, pour avoir accepté que je passe 4 mois dans son laboratoire. Merci pour son aide dès mon arrivée à VTech, ses conseils éclairés et d'avoir pris de son précieux temps pour m'encadrer pendant ces 4 mois fructueux.

Un merci, au passage, à trois étudiants de VTech : Yannick et Joshi ainsi qu'à Alice pour les excellentes soirées passées ensemble et tous les bons moments qui ont fait de mon passage à VTech, un excellent souvenir ! Thank you so much !

Enfin, merci à 3 amis très importants pour moi : Caterina, Sébastien et Geoffrey ! J'aimerais vous voir plus souvent et autrement que sur msn, facebook ou gtalk !

Enfin, merci à tous ceux qui ont contribué de près ou de loin à la réalisation de cette thèse.

Pendant toute ma scolarité j'ai eu l'occasion de rencontrer de nombreux enseignants, parmi lesquels deux personnes en particulier ont su se détacher du lot. Je tiens à remercier : M. Papin (professeur en Électronique et Physique-Chimie au lycée) ainsi que M. Crison (professeur d'Électronique en école d'ingénieur à l'ESIEA), qui par leurs enseignements m'ont transmis tous deux leur passion pour l'électronique. M. Crison m'a également fait découvrir le domaine de la réalité virtuelle qui me fascine, je l'en remercie chaleureusement ! Leurs connaissances, leurs conseils, leur disponibilité et leur gentillesse m'ont marqué plus que je ne saurais le dire.

Enfin, et de loin le plus important à mes yeux et à mon cœur, je tiens à remercier mes parents et mon frère. Merci de m'avoir toujours incité à aller plus loin et pour votre indéfectible soutien. Merci d'avoir été à tout moment à mes côtés et de m'avoir remonté le moral dans les moments difficiles. Merci d'avoir toujours été là pour moi. On ne peut choisir sa famille et quand bien même ce serait possible, je n'échangerais la mienne pour rien au monde. Je vous aime.



# Table des matières

<b>Chapitre 1 Présentation générale</b> .....	<b>1</b>
<b>Chapitre 2 Contexte du travail de thèse</b> .....	<b>7</b>
<b>1 Des applications 2D à l'interaction 3D temps réel</b> .....	<b>9</b>
<b>2 L'interaction manuelle, une interaction riche</b> .....	<b>12</b>
2.1 La main : caractéristiques et potentialités .....	12
2.2 Les prises .....	13
<b>3 Une taxonomie des interfaces actuelles de capture des mouvements</b> .....	<b>18</b>
3.1 Critères de catégorisation .....	18
3.2 Solutions à base de capteurs .....	19
3.2.1 Type de capteurs .....	19
3.2.2 Les capteurs « Very High Degree of Freedom » .....	20
3.2.3 Les capteurs « High Degree of Freedom » .....	30
3.3 Solutions basées sur le traitement d'images .....	33
3.3.1 Vision monoscopique .....	35
3.3.2 Vision stéréoscopique / multi-vues .....	41
<b>4 Discussion</b> .....	<b>47</b>
4.1 Synthèse des avantages et désavantages des solutions présentées .....	47
4.2 Faiblesse des solutions présentées .....	49
4.3 Émergences et promesses des caméras 3D .....	50
4.3.1 Triangulation .....	51
4.3.2 Temps de vol .....	54
4.3.3 Avantages .....	59
4.3.4 Comparatif des principales caméras 3D du marché .....	62
<b>5 Objectifs, Motivations et Hypothèses</b> .....	<b>62</b>
5.1 Le choix d'une technologie .....	62
5.2 Notre vision du modèle d'interaction de Norman .....	63
5.3 Hypothèses générales .....	64
5.4 Critères d'évaluation .....	65
5.5 Structuration du travail .....	66

## Chapitre 3 Souris vs. Cam3D pour la sélection en environnement virtuel : étude 1 .....67

<b>1</b>	<b>Méthodologie</b> .....	<b>69</b>
1.1	Présentation générale .....	69
1.2	Terminologie .....	70
1.3	Hypothèses opérationnelles .....	70
1.4	Tâche de sélection .....	71
1.5	Environnement virtuel : conception .....	72
1.6	Une interaction équivalente mais adaptée à chaque périphérique.....	73
1.7	Implémentation et développements logiciels.....	73
<b>2</b>	<b>Protocole expérimental</b> .....	<b>75</b>
2.1	Généralités .....	75
2.2	Choix des participants.....	75
2.3	Conditions initiales et consignes .....	76
2.4	Déroulement de l'expérimentation .....	76
2.5	Données recueillies.....	77
2.6	Variables et mesures .....	78
2.7	Méthodes d'analyse et tests statistiques .....	78
2.7.1	<i>Expertise en réalité virtuelle</i> .....	78
2.7.2	<i>Taille de la main</i> .....	79
2.7.3	<i>Performances / Acceptabilité</i> .....	80
2.8	Validité interne et externe.....	80
2.8.1	<i>Appariement</i> .....	80
2.8.2	<i>Contre-balancement</i> .....	81
2.8.3	<i>Affectation aléatoire</i> .....	81
<b>3</b>	<b>Influence du périphérique sur les performances</b> .....	<b>81</b>
3.1	Tests préliminaires.....	81
3.1.1	<i>Normalisation</i> .....	82
3.1.2	<i>Etude de l'effet d'apprentissage</i> .....	82
3.2	Résultats globaux : temps d'exécution .....	82
3.2.1	<i>Influence du type de périphérique sur les performances</i> .....	82
3.2.2	<i>Discussion</i> .....	82
3.3	Interaction entre le sexe et le périphérique sur les performances .....	83
3.4	Interaction entre l'expertise en R.V. et le périphérique sur les performances.....	83
3.5	Interaction entre la taille de la main et le périphérique sur les performances .....	84
3.6	Discussion.....	85

<b>4</b>	<b>Influence du périphérique sur les préférences subjectives : étude de l'acceptabilité .....</b>	<b>86</b>
4.1	Résultats globaux.....	86
4.1.1	<i>Statistiques .....</i>	<i>86</i>
4.1.2	<i>Influence du périphérique sur le ressenti de l'acceptabilité .....</i>	<i>86</i>
4.2	Interaction entre le sexe et le périphérique .....	89
4.3	Interaction entre l'expertise en R.V. et le périphérique.....	90
4.4	Synthèse des résultats .....	91
<b>5</b>	<b>Conclusions générales .....</b>	<b>91</b>
<b>Chapitre 4 Gant de données vs. Cam3D pour la manipulation et la navigation en environnement virtuel : étude 2 .....</b>		
<b>93</b>		
<b>1</b>	<b>Méthodologie .....</b>	<b>95</b>
1.1	Présentation générale.....	95
1.2	Terminologie .....	96
1.3	Hypothèses.....	96
1.4	Tâches de manipulation et de navigation .....	97
1.5	Environnements virtuels et tâches à réaliser : conception .....	98
1.6	L'utilisation d'un pie-menu adapté à l'environnement .....	100
1.7	Manipulation : pistes envisagées et solutions retenues.....	101
1.7.1	<i>Tâches simples .....</i>	<i>102</i>
1.7.2	<i>Tâches complexes .....</i>	<i>102</i>
1.8	La métaphore du joystick pour la navigation .....	103
1.9	Implémentation et développements logiciels.....	104
<b>2</b>	<b>Protocole expérimental .....</b>	<b>106</b>
2.1	Généralités.....	106
2.2	Choix des participants.....	106
2.3	Conditions initiales et consignes .....	107
2.4	Déroulement de l'expérimentation .....	108
2.5	Sources de données.....	109
2.6	Variables étudiées .....	110
2.7	Conditions de passation.....	112
2.8	Validité interne et externe .....	113
2.8.1	<i>Appariement .....</i>	<i>113</i>
2.8.2	<i>Contre-balancement .....</i>	<i>113</i>
2.8.3	<i>Affectation aléatoire .....</i>	<i>113</i>

<b>3</b>	<b>Analyse comparative des performances des participants.....</b>	<b>113</b>
3.1	Tests préliminaires et tests statistiques.....	114
3.2	Résultats généraux .....	114
3.3	Gants vs. Cam3D selon chaque tâche .....	115
3.3.1	<i>Tâche 1 – Puits</i> .....	115
3.3.2	<i>Tâche 2 – Cheval</i> .....	115
3.3.3	<i>Tâche 3 – Maisonnette</i> .....	116
3.3.4	<i>Tâche 4 – Navigation 1</i> .....	116
3.3.5	<i>Tâche 5 – Globe</i> .....	116
3.3.6	<i>Tâche 6 – Horloge</i> .....	116
3.3.7	<i>Tâche 7 – Ordinateur portable</i> .....	117
3.3.8	<i>Tâche 8 – Navigation 2</i> .....	117
3.3.9	<i>Tâche 9 – Bateau</i> .....	117
3.3.10	<i>Synthèse</i> .....	118
3.4	Étude de l’erreur d’orientation et de positionnement décomposée selon les axes X, Y et Z ... .....	118
3.4.1	<i>Déplacement</i> .....	118
3.4.2	<i>Rotation</i> .....	119
3.4.3	<i>Synthèse</i> .....	120
3.5	Étude de l’impact d’effet d’apprentissage .....	121
3.5.1	<i>Données globales</i> .....	121
3.5.2	<i>Données relatives à chaque tâche</i> .....	121
3.6	Étude du rapport précision / temps .....	122
3.7	Discussion sur les performances .....	123
<b>4</b>	<b>Influence du périphérique sur les préférences subjectives .....</b>	<b>124</b>
4.1	Analyse statistique.....	124
4.2	Etude de l’acceptabilité .....	125
4.3	Étude des préférences globales .....	127
4.4	Synthèse.....	128
<b>5</b>	<b>Évaluation des modalités d’interaction proposées .....</b>	<b>129</b>
5.1	Analyse statistique.....	129
5.2	Évaluation des modalités d’interaction .....	129
5.2.1	<i>Efficacité</i> .....	130
5.2.2	<i>Rapidité</i> .....	130
5.2.3	<i>Précision</i> .....	131
5.2.4	<i>Simplicité</i> .....	131
5.2.5	<i>Intuitivité</i> .....	132
5.2.6	<i>Synthèse</i> .....	132

5.3	Évaluation des modalités d'interaction .....	134
5.4	Synthèse.....	135
<b>6</b>	<b>Suggestions d'améliorations.....</b>	<b>136</b>
<b>7</b>	<b>Conclusions générales .....</b>	<b>137</b>
7.1	Synthèse des résultats .....	137
7.2	Conclusion.....	138
<b>Chapitre 5 Vers des préconisations pour le choix d'une interaction : une étude complémentaire .....</b>		<b>141</b>
<b>1</b>	<b>Retour sur l'étude 2 .....</b>	<b>143</b>
1.1	Analyse des commentaires et suggestions des participants concernant l'interaction.....	143
1.2	Suggestions prises en compte et modifications implémentées .....	144
1.2.1	<i>Modalités proposées pour le déplacement.....</i>	<i>144</i>
1.2.2	<i>Modalités proposées pour le redimensionnement.....</i>	<i>144</i>
1.2.3	<i>Modalités proposées pour la rotation.....</i>	<i>145</i>
<b>2</b>	<b>Méthodologie .....</b>	<b>146</b>
2.1	Présentation générale.....	146
2.2	Terminologie .....	146
2.3	Hypothèses.....	147
2.4	Tâches de manipulation.....	147
2.5	Environnement virtuel : conception .....	147
2.6	Implémentation et développements logiciels.....	148
<b>3</b>	<b>Protocole expérimental .....</b>	<b>149</b>
3.1	Généralités .....	149
3.2	Choix des participants.....	149
3.3	Conditions initiales et consignes .....	150
3.4	Déroulement de l'expérimentation.....	151
3.5	Données recueillies, variables, mesures et méthodes d'analyse .....	151
3.6	Validité interne et externe .....	153
3.7	Tests préliminaires et tests statistiques.....	153
<b>4</b>	<b>Influence de la participation à la seconde expérimentation .....</b>	<b>154</b>
<b>5</b>	<b>Étude des performances et préférences subjectives des participants concernant les interactions .....</b>	<b>155</b>
5.1	Le déplacement .....	155

5.2	Le redimensionnement.....	156
5.3	La rotation.....	157
5.4	Synthèse des résultats.....	159
<b>6</b>	<b>Influence de critères interindividuels sur les performances et les préférences .....</b>	<b>159</b>
6.1	Le sexe .....	159
6.2	L'expertise en RV .....	160
6.3	Synthèse.....	160
<b>7</b>	<b>Discussion et préconisations.....</b>	<b>161</b>
<b>Chapitre 6 Conclusions et perspectives .....</b>		<b>163</b>
<b>1</b>	<b>Conclusion .....</b>	<b>165</b>
<b>2</b>	<b>Perspectives .....</b>	<b>169</b>
<b>Annexes.....</b>		<b>173</b>
<b>Annexe A.....</b>		<b>175</b>
	Etude 1 – Fiche explicative pour la manipulation à la souris.....	175
	Etude 1 – Fiche explicative pour la manipulation avec la Cam3D .....	176
	Etude 1 – Questionnaire d'identification .....	177
	Etude 1 – Questionnaire post-passation .....	179
<b>Annexe B .....</b>		<b>183</b>
	Etude 2 – Fiche explicative sur la sélection / désélection .....	183
	Etude 2 – Fiche explicative sur l'appel du menu et sélection d'une action .....	184
	Etude 2 – Fiche explicative sur le déplacement d'un objet.....	185
	Etude 2 – Fiche explicative sur le redimensionnement d'un objet.....	186
	Etude 2 – Fiche explicative sur la rotation d'un objet.....	187
	Etude 2 – Fiche explicative sur la navigation en EV .....	188
	Etude 2 – Consigne et listes des tâches à effectuer.....	189
	Etude 2 – Questionnaire d'identification .....	192
	Etude 2 – Questionnaire post-passation .....	194
<b>Annexe C .....</b>		<b>200</b>
	Etude 3 – Fiche explicative sur la modalité 1 du déplacement .....	200
	Etude 3 – Fiche explicative sur la modalité 2 du déplacement .....	201
	Etude 3 – Fiche explicative sur la modalité 1 du redimensionnement .....	202

Etude 3 – Fiche explicative sur la modalité 2 du redimensionnement .....	203
Etude 3 – Fiche explicative sur la modalité 1 de la rotation.....	204
Etude 3 – Fiche explicative sur la modalité 2 de la rotation.....	205
Etude 3 – Questionnaire d’identification .....	206
Etude 3 – Questionnaire post-passation .....	208
Etude 3 – Questionnaire post-passation .....	208
<b>Bibliographie.....</b>	<b>213</b>



## Table des figures

Figure 1: Translation dans un repère cartésien ( $T_x, T_y, T_z$ ).....	9
Figure 2 : Rotation dans un repère cartésien ( $R_x, R_y, R_z$ ) .....	9
Figure 3 : Déplacement, redimensionnement et rotation d'un objet dans l'espace .....	10
Figure 4 : Les différentes articulations et parties de la main tiré de (Sturman, 1992) .....	13
Figure 5 : Mouvements des mains et des doigts : à gauche, supination et pronation, tirés de (Chalfoun et al., 2003) – au centre, abduction et adduction – à droite extension et flexion, tirés de (Bailly et al., 2008) .....	13
Figure 6 : Les différents mouvements que la main et les doigts peuvent effectuer en utilisant ses 23 degrés de liberté (Sturman, 1992).....	14
Figure 7 : A gauche : Modélisation mécanique de l'index, tirée de (Pouliquen, 2006), au milieu : modélisation mécanique de la main, à droite : axe de la main(Savescu, 2006) .....	15
Figure 8 : Exemples de classification de postures de la main, tirés de Schlesinger (MacKenzie & Iberall, 1994) .....	17
Figure 9 : Classification des posture de la main d'après Cutkosky (MacKenzie & Iberall, 1994) .....	17
Figure 10 : Capteurs regroupés par catégories .....	19
Figure 11 : à gauche, 3 bobines de cuivres disposées de façon orthogonales dans un capteur magnétique ; à droite, composants d'un système magnétique de tracking des mains (Miners, 2001).....	22
Figure 12 : Systèmes de capture de mouvements des mains « matériels » .....	32
Figure 13 : Techniques de traitement d'image abordées dans notre travail .....	35
Figure 14 : Principe de la technique Shape-from-silhouettes .....	44
Figure 15 : Systèmes de capture de mouvements des mains « logiciels » .....	46
Figure 16 : Principe de fonctionnement d'une triangulation passive (à gauche) et d'une triangulation active (à droite) .....	52
Figure 17 : Triangulation active à lumière structurée (visible).....	53
Figure 18 : Triangulation active à lumière structurée (invisible - infrarouge).....	54
Figure 19 : Principe de fonctionnement d'une caméra à temps de vol pulsé.....	55
Figure 20 : Principe de fonctionnement d'une caméra temps de vol à décalage de phase (May et al., 2007). .....	57
Figure 21 : Décalage de phase d'une onde lumineuse .....	58
Figure 22 : relation entre une distance $d$ et un décalage de phase $\phi d$ .....	59
Figure 23 : Caméras 3D .....	61
Figure 24 : Modèle de l'interaction homme-machine, tiré de (Norman, 1988).....	63
Figure 25 : Notre modèle de l'interaction homme-machine .....	64
Figure 26 : Environnement virtuel - espace de jeu de l'expérimentation 1.....	72
Figure 27 : Cartes de distances obtenues avec une SwissRanger SR4000. Le bruit est principalement présent au niveau des contours de la main. ....	74
Figure 28 : Une participante pendant l'expérimentation, avec la souris (à gauche) et la cam 3D (à droite) .....	77
Figure 29 : Méthode de mesure de la taille de la main .....	79
Figure 30 : Répartition des participants en fonction de la taille de leur main, en 3 groupes .....	79
Figure 31 : Différences de performances pour chaque périphérique selon le niveau d'expertises en RV. ....	84
Figure 32 : Distribution des notes jugeant l'apport de la Cam3D par rapport à la souris .....	87
Figure 33 : Distribution des notes concernant la fatigue perçue avec chaque périphérique .....	88

Figure 34 : Évaluation de l'apport de la Cam3D par les hommes et les femmes .....	89
Figure 35 : Fatigue perçue pour chaque périphérique, selon le sexe des participants .....	90
Figure 36 : Scène 1 avec les 3 objets à manipuler (puits, cheval et maisonnette) avant et après manipulation .....	98
Figure 37 : Scène 2 avec les 3 objets à manipuler (globe, horloge et ordinateur) avant et après manipulation .....	99
Figure 38 : Scène 3 avec l'objet à manipuler (bateau), avant et après manipulation .....	99
Figure 39 : Menu circulaire avec 4 choix possibles + un bouton validation pour la fin de l'expérimentation .....	100
Figure 40 : Procédure de sélection d'un objet (pour un droitier) .....	102
Figure 41 : Navigation entre les 3 scènes. Les flèches bleues indiquent l'itinéraire à suivre, les cadres marrons permettent au participant de se positionner avec précision sur la nouvelle scène.....	103
Figure 42 : Environnement virtuel de la phase d'apprentissage de l'expérimentation 2 .....	107
Figure 43 : Écran des consignes au début de l'expérimentation.....	108
Figure 44 : Participante passant l'expérimentation 2 avec le système Gants3D (à gauche) et le système Cam3D (à droite) .....	109
Figure 45 : représentation des scores moyens de chaque critère pour les 2 systèmes étudiés. Les variables pour lesquelles les différences sont significatives sont encadrées (en bleu, tendances).....	126
Figure 46 : Histogramme empilé de la répartition des participants selon le choix du participant pour chaque critère étudié.....	128
Figure 47 : Histogramme empilé de la répartition des participants selon leur évaluation de l'efficacité pour les 4 modalités d'interaction.....	130
Figure 48 : Histogramme empilé de la répartition des participants selon leur évaluation de la rapidité pour les 4 modalités d'interaction.....	130
Figure 49 : Histogramme empilé de la répartition des participants selon leur évaluation de la précision pour les 4 modalités d'interaction.....	131
Figure 50 : Histogramme empilé de la répartition des participants selon leur évaluation de la simplicité pour les 4 modalités d'interaction.....	131
Figure 51 : Histogramme empilé de la répartition des participants selon leur évaluation de l'intuitivité pour les 4 modalités d'interaction.....	132
Figure 52 : représentation des scores moyens de chaque critère pour les 5 modalités d'interaction ....	132
Figure 54 : Histogramme empilé de la répartition des participants selon le choix du participant pour chaque critère étudié et chaque modalité d'interaction .....	134
Figure 55 : fonctionnement de la rotation avec la modalité 1.....	146
Figure 56 : Scène et actions de déplacement, redimensionnement et rotation.....	148
Figure 57 : plan expérimental.....	150
Figure 58 : Participant passant l'expérimentation 3 - tâche de déplacement .....	151
Figure 59 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées .....	155
Figure 60 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées .....	157
Figure 61 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées en noir(en gris pour les tendances).....	158

## Liste des tableaux

Tableau 1 : Tableau récapitulatif des 6 degrés de liberté « classiques » pour le déplacement et l'orientation d'un objet.....	10
Tableau 2 : Mouvements fonctionnels de la main .....	14
Tableau 3 : Limites articulaires des doigts de la main .....	15
Tableau 4 : Déplacements et actions de sélection avec la « Cam3D » et le souris .....	70
Tableau 5 : Hypothèses opérationnelles de l'expérimentation 1 .....	71
Tableau 6 : Niveau d'expertise en fonction du score calculé d'après les réponses aux échelles de Likert et au questionnaire d'identification. ....	79
Tableau 7 : Taille de la main en fonction du score calculé d'après les réponses aux échelles de Likert....	80
Tableau 8 : Temps d'exécution avec la Cam3D et la Souris – données originales et transformées (Log10) .....	82
Tableau 9 : Moyennes des temps d'exécution avec la Cam3D et la Souris selon le sexe – données originales et transformées (Log10).....	83
Tableau 10 : Résultats de l'analyse des performances en fonction d'expertise en RV pour chaque périphérique.....	84
Tableau 11 : Performances en fonction de la taille de la main pour chaque périphérique.....	85
Tableau 12 : Proportions en faveur de la Cam 3D et de la souris, ainsi que les réponses aux échelles de Likert, selon chaque critère.....	87
Tableau 13 : Enchaînement des tâches (D = déplacement, T = redimensionnement, R = rotation, Navi. = navigation) - Les cases grisées sont des actions non réalisables pour les tâches correspondantes .....	109
Tableau 14 : Niveaux d'expertises basés sur les scores calculés d'après les réponses aux échelles de Likert .....	110
Tableau 15 : Étude des systèmes – variables étudiées et données recueillies .....	112
Tableau 16 : Étude des modalités d'interaction – variables étudiées et données recueillies .....	112
Tableau 17 : Résultats concernant les temps totaux d'expérimentation et de manipulation.....	114
Tableau 18 : Différences significatives entre Cam3D et Gants3D pour la tâche 2 .....	116
Tableau 19 : Différences significatives entre Cam3D et Gants3D pour la tâche 6 .....	117
Tableau 20 : Différences significatives entre Cam3D et Gants3D pour la tâche 7 .....	117
Tableau 21 : Différence significative entre Cam3D et Gants3D pour la tâche 9 .....	117
Tableau 22 : Décomposition de l'erreur de positionnement selon chaque axe pour les Gants3D .....	119
Tableau 23 : Décomposition de l'erreur de positionnement selon chaque axe pour la Cam3D.....	119
Tableau 24 : Décomposition de l'erreur d'orientation selon chaque axe pour les Gants3D (* le test utilisé est le test de Friedman, pour cause de non-normalité) .....	119
Tableau 25 : Décomposition de l'erreur d'orientation selon chaque axe pour la Cam3D.....	120
Tableau 26 : Comparatif de l'effet d'apprentissage entre les 2 systèmes. TX = tâche n°x ; C = Cam3D ; G = Gants3D ; T = temps total de la tâche ; TD = temps de déplacement ; TR = temps de rotation ; TE = temps de redimensionnement ; TN = temps de navigation .....	122
Tableau 27 : Différences significatives entre Cam3D et Gants3D pour le ratio « précision / temps ».....	123
Tableau 28 : Valeurs moyennes pour chaque critère d'étude de l'acceptabilité, d'après les échelles de Likert à 5 modalités (1 = très mauvais ; 5 = très bon) .....	125
Tableau 29 : Effectifs de participants selon la préférence attribuée à chaque critère.....	127
Tableau 30 : Oppositions entre les modalités d'interaction pour chaque critère.....	133
Tableau 31 : Effectifs de participants selon la préférence accordée pour modalité d'interaction et chaque critère.....	134

Tableau 32 : Différences significatives de performances pour chaque modalité selon la participation à l'expérimentation 2 .....	154
Tableau 33 : Différences significatives de préférences pour chaque modalité selon la participation à l'expérimentation 2 .....	154
Tableau 34 : Différences significatives entre modalité 1 et modalité 2 pour le déplacement .....	155
Tableau 35 : Différences significatives entre modalité 1 et modalité 2 pour le redimensionnement .....	156
Tableau 36 : Différences significatives entre modalité 1 et modalité 2 pour la rotation .....	157
Tableau 37 : Différences significatives entre modalité 1 et modalité 2 pour la rotation .....	158
Tableau 38 : Différences significatives de performances pour chaque modalité selon l'expertise en RV .....	160
Tableau 39 : Différences significatives de préférences pour chaque modalité selon l'expertise en RV .....	160

## Liste des équations

Équation 1 : Norme L1 .....	39
Équation 2 : Norme L2 .....	39
Équation 3 : Calcul de distance par triangulation passive .....	52
Équation 4 : Calcul de distance par triangulation active.....	53
Équation 5 : Calcul du temps de retard de la lumière dans une caméra à temps de vol pulsé .....	54
Équation 6 : Calcul de la portée maximale d'une caméra à temps de vol pour une longueur d'impulsion de 50ns.....	54
Équation 7 : Calcul de la distance avec la formule de Medina.....	55
Équation 8 : Calcul de la distance avec $S1=0,66$ et $S2=0,33$ .....	55
Équation 9 : Calcul de la distance pour une caméra à temps de vol à décalage de phase.....	57
Équation 10 : Détail du calcul de la distance pour une caméra à temps de vol à décalage de phase .....	58

# Glossaire

---



**Acceptabilité** (en anglais : *acceptability*) : désigne la valeur positive ou négative de l'image que se fait un individu (ou un groupe) à propos de l'outil ou du système. En d'autres termes, l'acceptabilité se mesure au niveau des attitudes et opinions des utilisateurs potentiels vis-à-vis du produit lui-même, de son utilité et de son utilisabilité (Tricot et al., 2003).

**Capture de mouvements** (en anglais : *motion capture* ou *mocap*) : aussi appelée captation de mouvements est une technique permettant de capter les mouvements d'un élément réel afin de les renvoyer dans un univers virtuel : les mouvements sont enregistrés ou restitués en temps réel vers d'autres systèmes.

**Cinématique inverse** (en anglais : *inverse kinematics*) : le problème résolu par la cinématique inverse est de trouver une posture garantissant la position et/ou l'orientation dans l'espace monde d'une ou de plusieurs parties du corps (Harrouet et al., 2006). En ce qui concerne la cinématique inverse, l'utilisateur spécifie pour un certain nombre d'articulations (généralement les extrémités du squelette) des contraintes spatiales (position et/ou orientation) et temporelles. Les paramètres des articulations restantes sont alors déterminés automatiquement. Un algorithme de cinématique inverse calcule une pose qui essaie de satisfaire au mieux les contraintes spécifiées, tout en respectant les degrés de liberté et les limites angulaires des articulations (Welman, 1993).

**Environnement virtuel (EV)** (en anglais : *virtual environment (VE)*) : c'est un environnement interactif, qui peut être tridimensionnel (3D), et qui est construit par un ordinateur en vue d'une simulation du monde réel. Un environnement virtuel est une modélisation d'un environnement réel.

**Immersion** : peut être définie comme l'ensemble des caractéristiques objectives d'un environnement virtuel qui visent à donner à un utilisateur des stimulations sensorielles et des possibilités d'action dans cet environnement virtuel (Mestre et al., 2006).

**Interaction** : comprend la sélection, le contrôle, la manipulation et la modification d'objets de l'EV et la navigation dans celui-ci (Coquillart et al., 2006).

**Interaction 3D** : interaction avec des mondes virtuels 3D par opposition à l'interaction 2D qui est une interaction pour laquelle les entités manipulées sont en 2D (Coquillart et al., 2006).

**Interaction temps réel** : est obtenue si l'utilisateur ne perçoit pas de décalage temporel (latence) entre son action sur l'environnement virtuel et la réponse sensorielle de ce dernier. Cette contrainte est difficile à satisfaire. A défaut, on peut chercher à ne point infliger de perturbations au sujet par ce décalage temporel, même s'il le perçoit (Burkhardt & Fuchs, 2006)

**Interface utilisateur 3D** (en anglais : *3D User Interface (3DUI)*) : C'est une interface utilisateur qui implique des interactions homme-machine dans lesquelles les tâches de l'utilisateur sont effectuées directement dans un contexte spatiale en 3D (Bowman et al., 2004).

**Latence** : C'est le décalage temporel entre une action de l'utilisateur sur les interfaces motrices et la perception des conséquences de cette action sur l'environnement virtuel à travers les interfaces sensorielles (Burkhardt & Fuchs, 2006).

**Métaphore** : D'après Coquillart et al. (2006), une métaphore d'interaction peut être assimilée à une image symbolique d'une action ou d'une perception utilisée pour réaliser une tâche précise dans un EV. Il s'agit de la transposition d'un objet ou d'un concept réel dans le monde virtuel. On parle par exemple de métaphore de la main virtuelle (*virtual hand*) : il s'agit de la transposition de la main de l'utilisateur dans l'environnement virtuel (Coquillart et al., 2006). L'utilisation de métaphores peut demander plus

d'efforts cognitifs si la symbolique métaphorique n'est pas connue par les utilisateurs. Ils devront faire l'effort de la comprendre et de l'assimiler, pour qu'elle devienne progressivement un schème d'usage (Burkhardt & Fuchs, 2006).

**Schème** : D'après (Piaget & Chomsky, 1979), les schèmes constituent, pour le sujet, des moyens à l'aide desquels il peut assimiler les situations et les objets auxquels il est confronté. Un schème est l'ensemble structuré des caractères généralisables de l'action qui permet de répéter l'action ou de l'appliquer à de nouveaux contenus. Un schème est donc l'organisation mentale des actions telles qu'elles se transfèrent ou se généralisent lors de la répétition de cette action en des circonstances analogiques.

**Technique d'interaction 3D** (en anglais : 3D interaction technique (3DIT)) : c'est une méthode permettant à un utilisateur d'accomplir une tâche dans un EV via l'interface utilisateur 3D. L'implémentation d'une 3DIT se charge de transformer l'information provenant d'un périphérique d'entrée (ou de plusieurs périphériques) en une action au sein du système et de transformer l'information de sortie du système en une forme qui peut être affichée par un périphérique de sortie (ou des périphériques) (Bowman et al., 2004).

**Transparence des interfaces** : de manière idéale, les réponses motrices (dans le cadre de la capture de mouvements) doivent être transmises sans support matériel entre l'homme et la machine. Au niveau physique, l'interface est transparente si elle ne gêne pas, ou peu, physiquement l'utilisateur, en particulier dans ses mouvements (Fuchs & Mathieu, 2003).

**Utilisabilité** (en anglais : usability) : peut se définir comme la mesure (ou une estimation probable) de la performance des utilisateurs, dans le contexte et pour l'utilisation prévue. Il existe plusieurs composantes de l'utilisabilité, y compris la facilité d'apprentissage, l'efficacité de l'utilisation, la mémorisation, la résistance à l'erreur, et la satisfaction subjective dans l'usage (Nielsen, 1994). Il en découle que l'utilisabilité n'est pas la propriété absolue d'un dispositif ou de telle technique d'interaction : elle est toujours relative à une tâche donnée, un contexte de réalisation et une population ciblée d'utilisateurs (Burkhardt & Sperandio, 2004).

**Utilité** (en anglais : usefulness) : ce critère correspond à un avantage ou un agrément significatif pour l'utilisateur dans une activité précise (en termes d'efficacité, de coût, de rapidité, de précision, d'agrément...). Cet avantage est toujours relatif : relatif aux objectifs de l'utilisateur, relatif aux outils existants ou habituellement utilisés, relatif à l'environnement d'utilisation, et relatif aux dépendances avec les autres activités » (Burkhardt, 2006).

# Chapitre 1

## Présentation générale

---



La métaphore du bureau, apparue sur les ordinateurs Apple en 1984, marque les débuts du paradigme WIMP (Window Icon Menu Pointer) (Beaudouin-Lafon, 2004), qui va de pair avec l'utilisation de la souris. Depuis, beaucoup d'efforts ont été portés sur l'amélioration de la qualité graphique des logiciels, mais d'une manière générale, ceux-ci ont fondé et fondent toujours leur interaction sur le binôme clavier-souris. Pour les usages les plus courants, telle la bureautique, la navigation dans des pages web, la retouche d'images, ... les applications restent en deux dimensions (i.e. sans information de profondeur) ; la souris reste alors le périphérique incontournable : elle est peu coûteuse et peut même adopter des formes ergonomiques pour limiter la fatigue musculaire. La manipulation d'objets à deux dimensions se fait donc très communément (e.g. une zone de texte dans un logiciel de PAO). D'autres applications, plus spécifiques permettent de « manipuler » du contenu dans un environnement en 3 dimensions (ou de se déplacer/naviguer dans celui-ci) : à la fois des applications ludiques (jeux ou fonctionnalités de logiciels) et des applications à portée professionnelle (e.g. conception, modélisation, réalité virtuelle). Ces dernières sont d'ailleurs de plus en plus adoptées par le grand public (e.g. les configurateurs d'intérieur). La recherche pour la conception de visualisations 2D ½ ou 3D (Robertson et al., 1998) souligne aussi les potentialités offertes par la puissance de calcul graphique des ordinateurs, même les plus simples.

L'interaction dans des environnements en 3 dimensions n'est plus à 3 degrés de liberté (2 pour la position et 1 pour la rotation), mais à 6 degrés de liberté : trois degrés de liberté permettent de connaître la position d'un objet quelconque dans l'espace (on prend par exemple le centre de gravité de l'objet) et trois autres servent à décrire son orientation. Les périphériques classiques ont naturellement été adaptés : si les simples déplacements d'une souris ne permettent pas d'orienter et de positionner un objet dans un environnement 3D, l'association d'un déplacement dans une direction et de l'appui sur une touche (clavier ou souris) va permettre de réaliser un déplacement différent. Bien souvent, le recours à des métaphores tend à faciliter la compréhension et l'interaction, par exemple dans les actions de rotation : la manipulation d'une sphère fictive est la plus courante (Chen et al., 1988). Même si les performances ou la facilité d'utilisation des périphériques 2D sont discutées (e.g., (Berard et al., 2009)), ces derniers sont toutefois largement répandus.

De fait, aucun périphérique d'interaction pour la manipulation d'objets 3D n'a été largement adopté par le grand public, mis à part dans certains secteurs bien spécifiques comme les jeux vidéos avec les joysticks et autres pads. Les souris 3D, bien que devenues abordables (de l'ordre de 50€), ne sont que très peu diffusées. Leur domaine d'utilisation reste trop restreint, car elles ne se substituent que partiellement à la souris en visant majoritairement l'interaction bi-manuelle (TAG, 2008). Le développement de tels périphériques soulignent en creux la faiblesse des modalités actuelles.

Il est aujourd'hui possible de proposer des périphériques qui soient réellement adaptés aux tâches en environnement 3D et acceptables par une grande majorité d'utilisateurs. D'autant plus que les utilisateurs professionnels et issus du grand public sont plus à même d'accepter le changement qu'il y a quelques années voire dizaines d'années. Nous sommes loin de la phrase de J. C. Dvorak à propos de la souris : « There is no evidence that people want to use these things » (Dvorak, 1984). Il suffit de voir le succès fulgurant des interfaces tactiles pour s'en persuader. Les clés du succès restent simples : une technologie performante et abordable, une réelle plus-value pour l'utilisateur et des applications qui en tirent parti. Dans le monde qui nous entoure, la main, par la richesse fonctionnelle qu'elle offre, nous permet de manipuler et d'interagir avec notre environnement. Ces gestes et ces prises naturels sont évidents pour nous. Il semble simple de mettre à profit cette interaction naturelle et de la transposer aux interfaces 3D comme c'est le cas en 2D avec les écrans tactiles.

De nombreux systèmes ont certes été développés, mais tous souffrent de défauts rédhibitoires, le premier étant souvent un prix prohibitif. Par exemple, les gants de données se doivent d'être couplés à des systèmes de tracking de la position et de l'orientation, et sont sensibles à la taille de la main. D'autres systèmes, comme les systèmes optiques nécessitent un appareillage relativement important et imposent de porter des marqueurs. Cela rend difficile une capture fine de mouvements des doigts d'une part, et rend l'utilisation laborieuse d'autre part (mise en place et calibration).

L'enjeu serait de transposer l'interaction manuelle (ou bi-manuelle) du monde réel au monde virtuel, avec une totale transparence pour l'utilisateur. Comment pourrions-nous interagir en environnement virtuel, avec la même simplicité que nous le ferions dans le réel, sans avoir recours à des interfaces intrusives ? L'apparition des caméras 3D apporte une première réponse à cette question qui aurait pu paraître encore idéalisée il y a peu. Bien que nos travaux soient antérieurs, l'apparition de la Kinect de Microsoft est un signal fort dans cette direction. Pour autant, tout l'interfaçage et l'applicatif reste encore à développer avant de disposer du périphérique « universel ».

Au-delà des potentialités techniques, peu de travaux scientifiques se sont intéressés à l'apport réel de cette technologie, encore naissante. L'objectif de notre recherche est de concevoir un nouveau système permettant la capture de mouvements de la main sans capteurs, et ainsi de rendre possible une interaction 3D temps réel transparente pour l'utilisateur. La problématique générale est de démontrer l'intérêt de notre système relativement aux périphériques existants. Nous nous intéressons plus particulièrement à la mesure des performances d'une part, et à la mesure de l'acceptabilité et des préférences des participants, d'autre part. La finalité de notre travail est la mise à disposition d'un système efficace, efficient, confortable, performant (dans le sens qu'il permet une bonne précision tout en minimisant le temps d'utilisation) et améliorant le sentiment d'immersion. Le système que nous proposons se doit d'apporter une plus-value réelle à l'utilisateur. L'utilisation visée concerne les applications de bureau, et de manière privilégiée, les environnements immersifs de grandes dimensions où l'utilisateur peut se tenir debout et bénéficier d'une vision en stéréoscopie. Aujourd'hui les domaines d'application concernés sont nombreux. Les usages peuvent aller du domaine de la santé, notamment pour la rééducation (Movea, 2009; Zhou & Hu, 2008), à la manipulation d'objets dans des environnements virtuels (Schlattmann & Klein, 2009), en passant par la commande de systèmes à partir de mouvements dans le cas de la domotique. Dans le domaine du handicap, un tel système pourrait par exemple faciliter l'interaction avec les enfants et les adultes polyhandicapés qui ont des difficultés lors de la réalisation de mouvements volontaires des membres supérieurs. Les activités actuelles, réalisées à base de commutateurs (Lancioni et al., 2011) pourraient être avantageusement remplacées par un système sans marqueurs, avec un langage relativement simple et adapté à chaque enfant. Ce ne sont là que des exemples parmi d'autres et il convient maintenant d'explorer le champ des possibilités.

Nous avons adopté une approche expérimentale reposant sur deux études principales et une étude complémentaire. Ces trois études reposent sur des protocoles expérimentaux similaires, et des tâches à réaliser relativement proches. Nous nous sommes intéressés à trois des tâches d'interaction les plus courantes en environnements virtuels, décrites par (Coquillart et al., 2003) : la sélection et la manipulation d'objets ainsi que la navigation. La manipulation étant composée des actions d'orientation, de positionnement et de redimensionnement. Nous avons opposé notre système à la souris (tâche de sélection) et aux gants de données couplés aux capteurs magnétiques (tâches de manipulation et de navigation). Cette opposition a donc été réalisée indépendamment sur les deux études principales, les données de la première ayant par ailleurs servi à affiner notre système.

Le document est organisé de la manière suivante. Le chapitre 1 est une introduction. Le chapitre 2 présente une taxonomie des applications 2D et 3D et permet de mettre en lumière la complexité et la richesse de l'interaction manuelle. Nous dressons ensuite un panorama des moyens actuels d'interaction, en présentant leurs avantages et leurs inconvénients. Nous montrons dans un premier temps, au travers de notre état de l'art, que de nombreux périphériques, basés sur des technologies variées, existent pour retranscrire les mouvements des mains et des doigts dans un environnement virtuel ou simplement pour reconnaître des gestes et des signes dans un environnement réel. Partant de ces constats, nous montrons qu'aucun périphérique « universel » n'a encore été développé pour l'interaction manuelle en trois dimensions. L'universalité pouvant ici se décrire par une adoption large, motivée par une utilisation simple, performante, peu contraignante et pour un coût relativement faible. Nous présentons ensuite nos objectifs et formulons nos hypothèses de travail.

Le chapitre 3 présente la première étude, conduite auprès de 71 participants, qui a pour but de démontrer l'intérêt de notre système pour une tâche simple de sélection, comparativement à un périphérique usuel et adopté par tous : la souris. Ce périphérique bien qu'universellement utilisé, ne permet d'interagir de façon simple que dans des environnements 2D, des artifices devant être trouvés pour une interaction en 3D. Les participants doivent sélectionner successivement 10 cubes dans un environnement virtuel, avec chaque système. Cette première expérimentation s'appuie donc sur un langage simple (main ouverte, main fermée) ainsi que sur la prise en compte de la position d'une seule main, dans un repère cartésien à trois dimensions. L'analyse porte sur des données qualitatives (temps d'exécution) et quantitatives (questionnaires post-passation).

Le chapitre 4 présente les améliorations apportées à notre système ainsi que la mise à l'épreuve de celui-ci par l'intermédiaire d'une expérimentation conduite sur 20 participants. Pour cette expérimentation, nous nous sommes appuyés sur la « Kinect » commercialisée par Microsoft, basée sur une technologie relativement proche de celle de la caméra que nous avons utilisée dans notre première expérimentation et dotée d'un rapport prix / performances extrêmement favorable. Cela nous a permis de nous appuyer sur des bibliothèques open-source et d'améliorer plusieurs points cruciaux : la capture des mouvements se fait maintenant sur les 2 mains de l'utilisateur et notre système peut même suivre plusieurs utilisateurs en même temps. Les capacités de notre système ont ainsi été étendues à la manipulation d'objets en 3D, c'est-à-dire le déplacement, la rotation, le redimensionnement ainsi que la navigation en environnement virtuel. Cette seconde expérimentation nous a permis de nous rapprocher de tâches réelles et d'évaluer l'intérêt qualitatif et quantitatif de notre système de capture sans marqueurs, comparativement aux systèmes classiques utilisés en réalité virtuelle tels que les gants de données et les capteurs magnétiques. Nous présentons donc les améliorations de notre système, le protocole expérimental qui a été adapté à la plus grande richesse des tâches proposées, puis les résultats, analyses et conclusions.

Nous présentons une étude complémentaire dans le chapitre 5, qui vise à l'amélioration des modalités d'interaction développées précédemment. En effet, les préférences, commentaires et suggestions d'améliorations recueillis lors de la seconde étude nous ont permis d'identifier les faiblesses de notre système associé aux modalités d'interaction proposées. Nous en avons donc développé de nouvelles pour la rotation, le positionnement et le redimensionnement et les avons opposées aux modalités présentes dans l'étude 2. Cette étude porte sur 30 participants, dont 12 ayant déjà participé à l'étude précédente. La tâche consiste à positionner (ou orienter ou redimensionner) un cube dans un environnement en 3 dimensions. Nous avons recueilli des données quantitatives (temps d'exécution,

précision) ainsi que qualitatives, d'après des questionnaires post-passation. Le protocole expérimental, les résultats, discussions et conclusions sont ainsi présentés dans ce chapitre.

Le chapitre 6 expose les conclusions générales et les perspectives de notre travail. Nous présentons un apport théorique : la taxonomie des moyens de capturer les mouvements des mains pour l'interaction en environnement virtuel, sans s'y restreindre. Nous présentons également deux apports pratiques. D'une part, la démonstration de l'intérêt des caméras 3D pour la capture de mouvements de la main en temps réel. D'autre part, le système que nous avons développé, c'est-à-dire la caméra 3D, les algorithmes de traitement et les modalités d'interactions. Enfin, nous dressons les perspectives scientifiques, technologiques et applicatives envisagées ; celles-ci justifient la poursuite des efforts dans le domaine de la reconnaissance des mouvements de la main sans marqueurs.

## Chapitre 2

### Contexte du travail de thèse

---



Dans ce chapitre, nous allons aborder différents points cruciaux permettant de comprendre les motivations et l'intérêt de cette thèse. En premier lieu, nous expliquerons quelles sont les différences entre les applications 2D et les applications 3D en montrant que les interactions usuelles dans les applications de type bureautique ne sont plus adaptées lorsqu'il s'agit de travailler dans un environnement 3D. Nous expliquerons ensuite en quoi consiste une tâche de manipulation d'objet (aussi bien dans la vie « réelle » que dans un environnement virtuel) et nous expliquerons en quoi la transposition du réel vers le virtuel est une tâche difficile à réaliser en particulier pour ce qui est de la transposition des mouvements des mains. Nous verrons pourquoi cet objectif est difficile à atteindre en montrant que la complexité fonctionnelle des mains (tant au niveau de leurs caractéristiques, de leurs potentialités mais également des prises qu'elles permettent) constitue la principale raison. Nous réaliserons, dans un second temps, une taxonomie des solutions existantes qui permettent de capturer les mouvements de la main. Ces solutions peuvent être aussi bien matérielles que logicielles et nous mettrons en exergue les points forts et les points faibles de chacune de ces techniques. Pour cela, nous appuierons notre argumentaire sur des exemples concrets et illustrés de travaux traitant de ces solutions. Nous synthétiserons ensuite les avantages et les inconvénients liés à l'utilisation de ces systèmes et discuterons de l'émergence et des promesses d'une technologie récente : les caméras 3D. Enfin, nous exposerons les motivations qui nous ont amenées à diriger nos recherches sur ce sujet, les objectifs que nous nous sommes fixés et également les hypothèses que nous avons émises.

## 1 Des applications 2D à l'interaction 3D temps réel

Lorsque l'on manipule un objet (aussi bien réel que virtuel), on lui applique des déplacements, des rotations et des actions (déformations, changements d'état, etc.). En réalité, on ne fait que passer d'un état à un autre. Six degrés de liberté suffisent pour décrire la position et l'orientation d'un objet dans l'espace par rapport à un repère fixe (système de coordonnées cartésiennes  $x, y, z$ ). Un degré de liberté (ddl) correspond à une rotation autour d'un axe articulaire ( $R_x$  - Roll,  $R_y$  - Pitch,  $R_z$  - Yaw) ou à une translation le long d'un axe ( $T_x$  - Forward,  $T_y$  - Drift,  $T_z$  - Ascension). Il suffit donc de changer ses coordonnées pour le déplacer ou l'orienter.

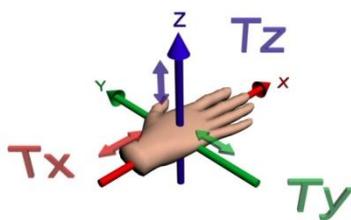


Figure 1: Translation dans un repère cartésien ( $T_x, T_y, T_z$ )

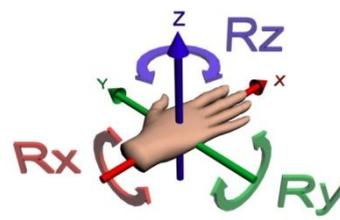


Figure 2 : Rotation dans un repère cartésien ( $R_x, R_y, R_z$ )

Le tableau suivant récapitule les mouvements que l'on peut appliquer à un objet (réel ou virtuel) avec les noms correspondant en français et en anglais ainsi que les axes concernés pour ces déplacements ou rotations.

Tableau 1 : Tableau récapitulatif des 6 degrés de liberté « classiques » pour le déplacement et l'orientation d'un objet

Degré	Axe	Mouvement (pour un véhicule par exemple)	Action réelle	Anglais
Tx	Longitudinal	Avance	Avant - arrière	<i>Forward</i>
Ty	Transversal	Dérive	Droite - Gauche	<i>Drift</i>
Tz	Vertical	Ascension	Haut – bas	<i>Ascension</i>
Rx	Longitudinal	Roulis	Basculer de droite à gauche	<i>Roll</i>
Ry	Transversal	Tangage	Basculer d'avant en arrière	<i>Pitch</i>
Rz	Vertical	Lacet	Pivoter comme les aiguilles d'une montre	<i>Yaw</i>

Par exemple, si l'on travaille dans un plan (x, y), ce qui est le cas dans les applications 2D courantes, un objet 2D peut être manipulé selon 2 ddl pour ce qui est des translations (Tx et Ty) et éventuellement 1 ddl en ce qui concerne les rotations (Rz). Ces actions sont réalisées au moyen de boutons, de liens et de menus, qui permettent de sélectionner des objets (par exemple des images), les modifier (en position, rotation, dimension), mais aussi de naviguer dans la page ou l'application. Le binôme clavier / souris est alors, dans ce cas, tout à fait suffisant.

Dans les applications 3D, la manipulation d'objet est différente car on ne travaille plus uniquement selon les axes x et y, mais on a aussi besoin de l'axe z. La souris étant un périphérique 2D, elle ne permet donc pas de travailler de façon directe selon le 3<sup>ème</sup> axe. Les logiciels nécessitant la manipulation d'objets 3D, comme par exemple les outils de modélisation (Catia, 3ds Max, Maya, ...), ont conservé l'usage du clavier et de la souris en adaptant la configuration de l'espace de travail : la scène est découpée sous différentes vues ou *viewports* (de haut, de gauche, de face,...). Pour positionner un objet dans la scène il est alors nécessaire de la positionner selon une vue (e.g. selon les axes X et Y) puis selon une deuxième (e.g. selon les axes Y et Z). Il n'est donc jamais possible d'agir sur les 3 axes simultanément.

Cette décomposition des actions en actions élémentaires constitue la première lacune. Celle-ci n'ayant pas été comblée par l'utilisation conjointe du clavier et de la souris. Le problème rencontré avec la manipulation se retrouve par ailleurs à l'identique pour l'orientation et le redimensionnement d'un objet en fonction des 3 axes. Cela fait donc, de façon optimale, six étapes différentes pour réaliser l'ensemble des actions possibles, et donc de permettre à un objet de passer d'une position A à une position B (voir Figure 3).

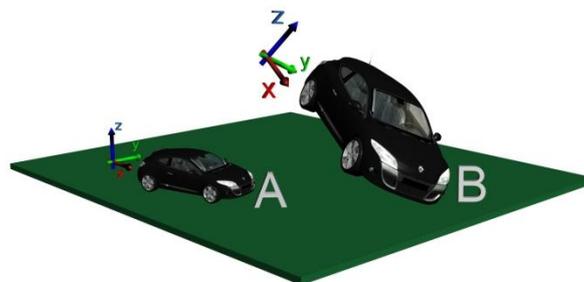


Figure 3 : Déplacement, redimensionnement et rotation d'un objet dans l'espace

Cela est réhhibitoire pour les applications de réalité virtuelle où la manipulation d'objet devrait pouvoir se faire, comme dans la vie réelle, selon les 3 axes en même temps et en temps réel. De plus, dans le

domaine de la réalité virtuelle, la manipulation d'objets ne se résume pas uniquement aux actions énoncées précédemment. De nombreux autres paramètres doivent pouvoir être appliqués aux objets, comme par exemple la vélocité ou la force pour imprimer un mouvement à l'objet ou encore connaître la pression exercée sur cet objet. Les périphériques courants montrent donc rapidement leurs limites en réalité virtuelle. Celles-ci sont multiples ; en premier lieu, elles sont d'ordre technique, comme par exemple l'impossibilité de capturer un geste professionnel, un savoir faire, lorsqu'on veut étudier les mouvements d'un sportif ou encore pour évaluer ergonomiquement des postures de travail. Il existe également des limites cognitives, en raison de l'obligation de décomposer une action en actions élémentaires qui demande plus de réflexion, d'attention, et d'efforts. Dans le monde « réel », nous n'avons pas à décomposer des mouvements simples (Bach & Scapin, 2003). De plus, l'utilisation de périphériques entre l'utilisateur et l'ordinateur peuvent diminuer le caractère immersif des applications et donc le sentiment d'immersion qui est pourtant primordial en réalité virtuelle (Coquillart et al., 2006). Enfin, d'un point de vue ludique, la jouabilité devient rapidement limitée, les jeux actuels étant en 3 dimensions. Les interactions restent alors simples et n'offrent pas une expérience de jeu proche de la réalité. Le joueur reste assis devant son écran d'ordinateur et se contente seulement d'un clavier et d'une souris.

La réalité virtuelle trouve une de ses motivations lorsque certaines applications ne peuvent plus être simplement adaptées au moyen d'interactions 2D existantes. Les moyens d'interactions, aussi bien logiciels que matériels, doivent être totalement repensés et adaptés aux tâches à exécuter quand il s'agit de travailler en environnements virtuels. Il devient alors nécessaire d'utiliser de nouvelles techniques ou périphériques pour parvenir au but recherché.

Un des principaux problèmes (et souvent difficile à résoudre), auquel doivent faire face les développeurs d'environnements virtuels (EVs) et d'applications de réalité virtuelle, est de trouver pour les utilisateurs des moyens d'interagir avec le monde virtuel. Les concepteurs d'applications doivent déterminer comment relier les entrées disponibles (les pressions de boutons, les gestes des mains, etc) à des actions dans l'environnement virtuel. Par conséquent, l'interaction au sein d'un EV est sans doute le facteur le plus limitant pour le développement d'applications complexes en réalité virtuelle (RV). Par exemple, les interactions utilisées avec de grandes quantités de données, des informations alphanumériques, ou des opérations abstraites peuvent ne pas bien être adaptées aux méthodes courantes d'interaction utilisées en réalité virtuelle, qui sont principalement spatiales (Christmann et al., 2010; Hartling et al., 2002).

D'après (Coquillart et al., 2003) « on nomme interaction 3D, l'interaction avec des mondes virtuels 3D par opposition à l'interaction 2D qui est une interaction pour laquelle les entités manipulées sont en 2D (dessin 2D,...) ». Ce type d'interaction regroupe de nombreux aspects. Selon (Hand, 1997) et (Bowman, 1999) trois tâches principales sont considérées : le déplacement et la recherche d'itinéraire, la sélection-manipulation d'objets et le contrôle d'applications. Cette dernière tâche correspond aux commandes requises pour modifier l'état du système ou encore changer les paramètres. Ces trois tâches se doivent d'être exécutées sans que l'utilisateur ne perçoive « de décalage temporel (latence) entre son action sur l'environnement virtuel et la réponse sensorielle de ce dernier » ; interaction alors qualifiées de temps réel. D'après (Arnaldi et al., 2003) « cette contrainte est difficile à satisfaire et si ce n'est pas possible, on doit chercher à ne pas infliger de perturbations au sujet par ce décalage temporel, même s'il le perçoit ».

Quelque soit le périphérique d'interaction, l'utilisateur se sert avant tout de ses mains et de ses bras pour interagir. Différents périphériques et techniques ont été développés pour récupérer et traduire les mouvements effectués par l'utilisateur. Bien que créée après la trackball (en 1952), la souris est l'un des

plus anciens systèmes de capture de mouvements (créée en 1963 à Stanford). Cependant, de nombreux autres ont été créés afin d'enrichir le nombre d'actions possibles et la précision avec laquelle on souhaite transposer ces mouvements sur ordinateur. Ainsi, sont apparus de nouveaux périphériques comme les joysticks, les souris 3D, les gants de données, les exosquelettes,...

Avant de faire un état de l'art reprenant ces différents périphériques de capture de mouvements et comparer leurs avantages et inconvénients, nous allons brièvement montrer la richesse fonctionnelle et la complexité de la main.

## 2 L'interaction manuelle, une interaction riche

La cinématique du corps humain comporte plus de 200 degrés de liberté (Bowman et al., 2004; Bowman & McMahan, 2007). Il apparaît clair qu'il est impossible de prendre en compte tous ces degrés de liberté dans le cadre d'une application avec les contraintes temporelles du temps réel. Une réduction du nombre de degrés de liberté n'est néanmoins pas neutre, car la construction d'un modèle simplifié entraîne forcément une perte de réalisme. Il est alors nécessaire de trouver un compromis entre « réalisme », simplicité de construction et d'utilisation, et efficacité par rapport à l'objectif visé (Lempereur, 2008).

À elle seule, la main est capable d'assurer 70% des capacités motrices de l'homme et est ainsi dotée d'une très grande richesse fonctionnelle (Lempereur, 2008). Nous nous sommes, par conséquent, focalisés sur les mains et les avant-bras, puisqu'ils sont suffisants dans le cadre de tâches d'interaction et de manipulation d'objets 3D.

### 2.1 La main : caractéristiques et potentialités

Krout a identifié plus de 5000 gestes différents de la main (Krout, 1935). Celle-ci étant, en effet, capable d'effectuer une grande multiplicité de tâches de saisie grâce à un réseau élaboré de muscles et de tendons (Lempereur, 2008). Elle se compose de deux parties : la main proprement dite (constituée de la paume et du talon) et les doigts (un doigt court : le pouce, et 4 doigts longs). Ces derniers lui permettent d'effectuer une grande variété de gestes. Les doigts comportent 15 articulations (Figure 4) lui permettant ainsi d'avoir 22 degrés de liberté ou 23 si l'on considère que le métacarpe (CMC) du pouce possède 3 degrés de liberté au lieu de 2 (Buchholz & Armstrong, 1992). Le pouce est différent des autres doigts car il ne possède pas de phalange intermédiaire. Du fait de la complexité de ses mouvements, l'accord n'a pas encore été fait sur une terminologie commune pour la définition et la mesure des degrés de liberté qu'il autorise. Le nombre de degré de liberté que possède la main ne fait donc pas l'unanimité.

La main prise dans son ensemble, il est donc possible d'avoir 29 degrés de libertés : 23 pour les doigts (en admettant que le métacarpe du pouce possède 3 ddl) et 6 pour le poignet (3 rotations et 3 translations).

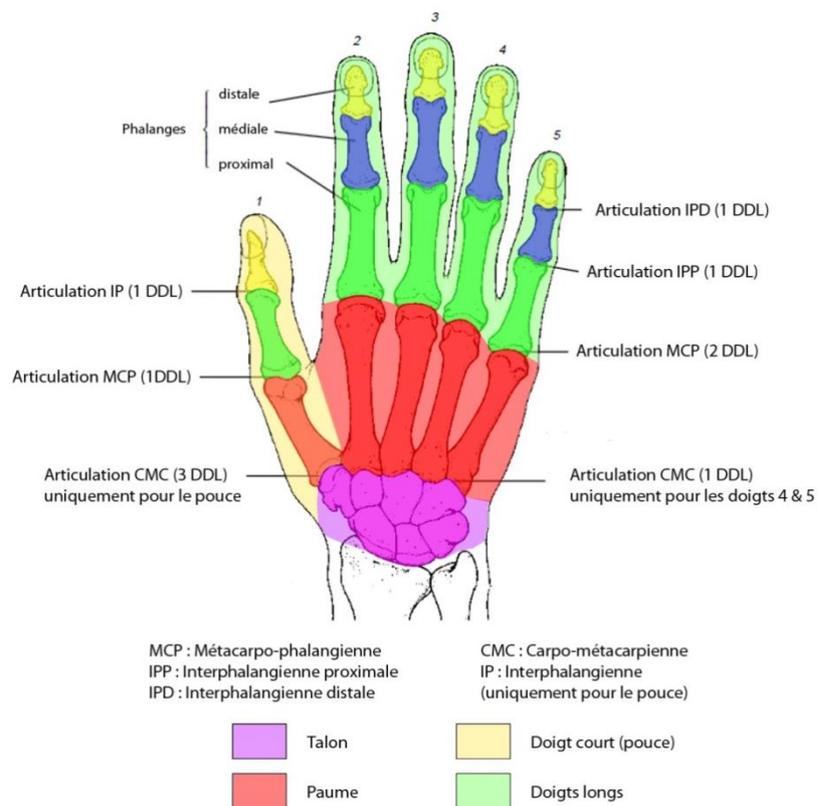


Figure 4 : Les différentes articulations et parties de la main tiré de (Sturman, 1992)

## 2.2 Les prises

Les mouvements des doigts sont contraints et organisés en 2 types (Lin et al., 2000) : les contraintes statiques et les contraintes dynamiques. Les contraintes statiques limitent les mouvements des doigts et correspondent aux butées articulaires. Par exemple, l'angle d'abduction / adduction de l'index et l'auriculaire est toujours compris entre  $-15^\circ$  et  $15^\circ$  (Lin et al., 2000). La Figure 5 représente les mouvements de supination, de pronation, d'abduction, d'adduction, d'extension et de flexion.

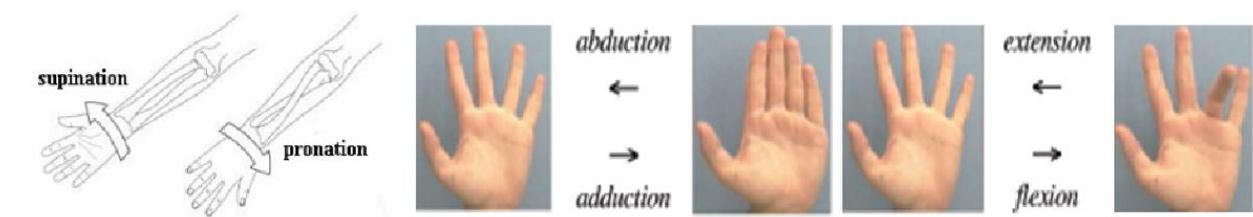
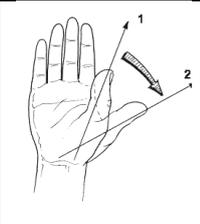
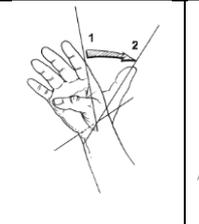
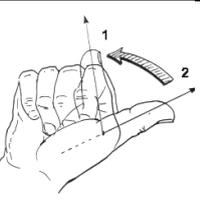
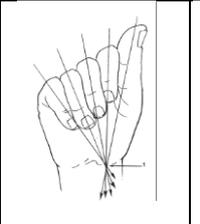
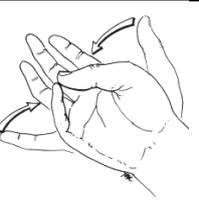


Figure 5 : Mouvements des mains et des doigts : à gauche, supination et pronation, tirés de (Chalfoun et al., 2003) – au centre, abduction et adduction – à droite extension et flexion, tirés de (Bailly et al., 2008)

Les contraintes dynamiques concernent la dépendance intra et extra-doigt. Par exemple, le majeur et l'annulaire sont très liés (Häger-Ross & Schieber, 2000) : un mouvement de forte amplitude de l'un de ces deux doigts entraîne un mouvement de faible amplitude de l'autre doigt. A l'opposé, le pouce et l'index peuvent être manipulés indépendamment l'un de l'autre et permettent un contrôle fin.

Par convention, les doigts sont numérotés de 1 à 5 en partant du pouce.

Tableau 2 : Mouvements fonctionnels de la main

					
Adduction (1) Abduction (2) du pouce en CMC	Flexion (1) Extension (2) du pouce en CMC	Flexion (1) Extension (2) du pouce en MCP	Adduction (1) Abduction (2) des doigts II à V	Convergence des doigts vers le point scaphoïdien	Opposition du pouce et des doigts (l'auriculaire ici)

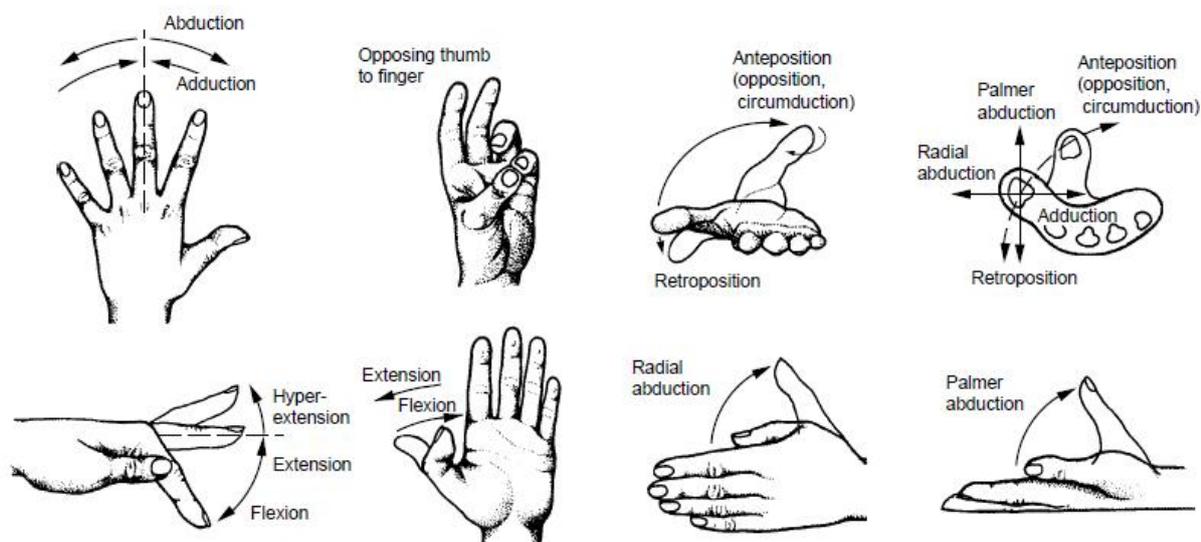


Figure 6 : Les différents mouvements que la main et les doigts peuvent effectuer en utilisant ses 23 degrés de liberté (Sturman, 1992)

L'articulation carpo-métacarpienne (CMC) de l'index et du médus est pratiquement immobile. Celles de l'annulaire et de l'auriculaire sont très légèrement mobiles, permettant à la main de se creuser. En revanche, l'articulation CMC du pouce est très complexe; elle permet des mouvements de flexion / extension (la flexion incline le pouce vers le creux de la paume, l'extension le place dans le plan de la paume) et d'abduction / adduction (le premier métacarpien se rapproche / écarte du deuxième) (Kamina & Francke, 1994) (cf. Tableau 2).

Les mouvements de flexion / extension (qui rapprochent ou éloignent le doigt de la paume) ainsi que les mouvements d'abduction / adduction sont permis au niveau des articulations métacarpo-phalangiennes (MCP). En revanche, pour le pouce, ces derniers sont faibles et seule la flexion est prise en considération pour l'articulation MCP1. Pour les autres doigts en revanche, l'abduction écarte le doigt de l'axe de la main, et l'adduction l'en rapproche. L'axe de la main est constitué par le troisième métacarpien et le majeur. Cet axe a un rôle particulier car, pendant les mouvements d'écartement et de rapprochement des doigts, le majeur reste pratiquement immobile. Ceci est également renforcé par le fait que lorsqu'on écarte volontairement les doigts, l'axe de chacun des cinq doigts vient converger vers la base de l'éminence thénar, en un point qui correspond approximativement au tubercule du scaphoïde (Savescu, 2006).

Au niveau des articulations inter-phalangiennes (IP), seuls les mouvements de flexion / extension sont possibles. La flexion permet à la main de se fermer et d'assurer une prise. L'axe de ces rotations est transversal et légèrement oblique, par conséquent lorsqu'on ferme le poing, les doigts convergent vers le point scaphoïdien.

On peut synthétiser ces observations d'anatomie fonctionnelle en modélisant chaque doigt comme étant un segment à trois articulations. Celles en contact avec la paume possèdent deux degrés de liberté, une en flexion / extension et une en abduction / adduction. Les deux articulations distales de chaque doigt sont, quant à elles, vues comme des liaisons de type rotation pure, en flexion / extension.

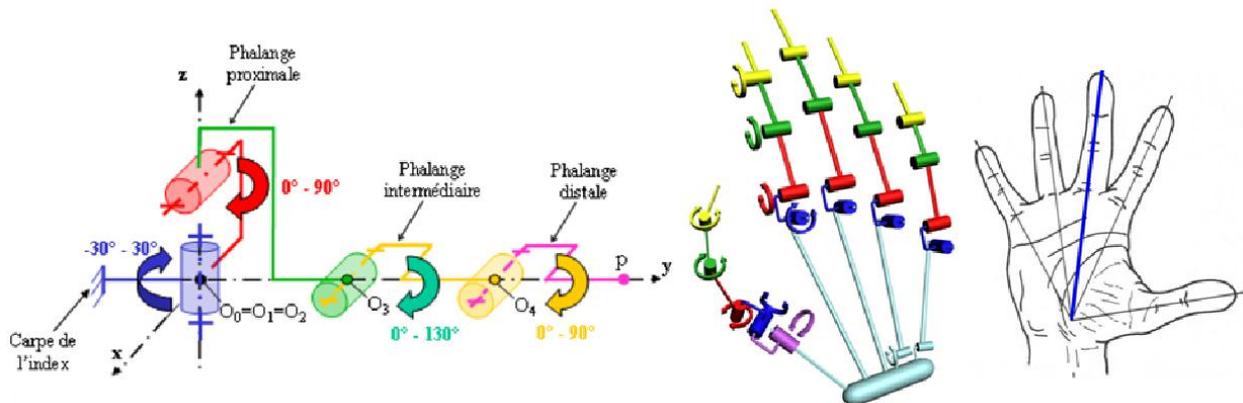


Figure 7 : A gauche : Modélisation mécanique de l'index, tirée de (Pouliquen, 2006), au milieu : modélisation mécanique de la main, à droite : axe de la main (Savescu, 2006)

Le tableau suivant (tiré de (Pouliquen, 2006)) révèle les limites de chaque articulation des doigts de la main. Dans une case, [x,y] signifie la limite minimale x et la limite maximale y par rapport à l'axe de la main (le majeur), f/e signifie : flexion / extension et ab/ad : adduction / abduction.

Tableau 3 : Limites articulaires des doigts de la main

	CMC f/e	CMC ad/ab	MCP (PP) f/e	MCP (PP) ad/ab	IPP (MP) f/e	IPD (DP) f/e
Auriculaire	[-2° ; 2°]	[0° ; 5°-10°]	[0° ; 90°]	[-15° ; 15°]	[0° ; 110° - 130°]	[0° ; 65° - 90°]
Annulaire	[-1° ; 1°]	[0° ; 2°]	[0° ; 90°]	[-15° ; 15°]	[0° ; 110° - 130°]	[0° ; 65° - 90°]
Majeur	0°	0°	[0° ; 90°]	0° (axe)	[0° ; 110° - 130°]	[0° ; 65° - 90°]
Index	0°	0°	[0° ; 90°]	[-30° ; 30°]	[0° ; 110° - 130°]	[0° ; 65° - 90°]
Pouce	[-25° ; 120°]	[0° ; 35°-40°]	[10° ; 90°]	0°	-	[0° ; 65° - 90°]

Ce tableau réalisé par (Pouliquen, 2006) permet de déduire facilement le degré de liberté de chaque articulation. Ici, il est considéré que l'articulation métacarpo-phalangienne (MCP) du pouce n'a qu'un seul degré de liberté et l'articulation carpo-métacarpienne (CMC) n'en a que deux. De plus, les flexions / extensions et les adductions / abductions de l'auriculaire et de l'annulaire sont très faibles et peuvent être considérées comme négligeables. Ces valeurs varient d'un auteur à l'autre (e.g., (Kapandji, 1980; Marieb, 1993; Pouliquen, 2006)) à cause de la différence de définition des mouvements. Marieb considère par exemple les mouvements d'abduction / adduction par rapport à l'axe du majeur alors que Kapandji considère les mouvements absolus.

Les mains nous permettent d'effectuer un grand nombre d'actions, aussi bien celles nécessitant de la précision que celles nécessitant de la force. Notre intérêt se focalise sur la reconnaissance des mouvements de la main, nous allons donc nous attarder sur les actions nécessitant de la précision et demandant plus ou moins de finesse lors de la manipulation d'objets virtuels. La principale fonction de

la main est la préhension basée sur l'opposition, c'est-à-dire la faculté qu'a le pouce de se porter vers les autres doigts pour former des pinces et des prises. Le nombre de prises différentes est impressionnant, depuis la simple pince bidigitale jusqu'aux « prises plus action » (Napier, 1956). Nous en dressons ici les grandes catégories en un rapide panorama :

### Les prises élémentaires

- Les pinces : seule l'extrémité des doigts interviennent sans participation de la paume ;
- Les pinces bipolaires (bidigitales) : seuls deux points de contact maintiennent l'objet ;
  - o La pince pouce-index est la plus précise et la plus utilisée ;
  - o La pince pouce-médius, moins utilisée, est aussi excellente, mais sert surtout en suppléance ;
  - o Les pinces pouce-annulaire et pouce-auriculaire sont peu utilisées, les deux derniers doigts servant surtout pour le verrouillage des prises à « pleine paume » ;
- Les pinces multipolaires (polydigitales) : au moins trois points de contact maintiennent l'objet ;
  - o tridigitale : pouce-index-médius, la plus utilisées (crayon, alimentation) ;
  - o tétradigitale : pouce-index-médius-annulaire, (utilisée aussi pour l'alimentation).

### Les prises simples

Cette dénomination permet de les distinguer des « prises plus action », que nous envisagerons plus loin. Les deux premières font intervenir la paume et/ou la première commissure :

- Les prises sphériques : prises digito-palmaires entourant l'objet de toutes parts, suivant son volume (e.g., lorsqu'on saisit un fruit à pleine main) ;
- Les prises cylindriques : prises digito-palmaires verrouillées par le pouce. Elles peuvent être de plus ou moins gros calibre, suivant le diamètre du cylindre, mais elles sont d'autant plus efficaces que les doigts sont plus « enroulés » dans la paume ;
- Les prises « en crochet » : Elles permettent la suspension d'un objet (sac ou valise) vers le bas, ou la suspension du corps grâce aux membres supérieurs : alpinisme (crochet simple) ou barre fixe et trapèze (crochet verrouillé par le pouce). Elles peuvent concerner un seul doigt, l'index ou plusieurs des quatre doigts longs ;
- La prise « de soutien » ou de « poussée » : prise de « soutien » contre la pesanteur (e.g., une ou deux mains, à plat et en supination, poignet en extension, peuvent soutenir un plateau ou une main « en cuillère » peut prendre de l'eau ou du sable). La prise de « poussée » correspond à une ou deux mains à plat, en pronation, poignet en extension à angle droit, pour par exemple pousser horizontalement un objet lourd ;
- Les prises de « suppléance ». Lors de la perte fonctionnelle du pouce, deux types de prises peuvent être utilisées:
  - o la prise digito-palmaire : les quatre doigts longs sont enroulés dans la paume autour de l'objet, mais la prise manque de fermeté en raison de l'absence de verrouillage ;
  - o la prise interdigitale : l'objet de petit volume, une cigarette par exemple, est saisie entre deux doigts longs, prise de suppléance parfois très utile malgré son manque de fermeté et de précision.

### Les prises complexes

- Les prises centrées : par exemple la tenue d'une fourchette qui nécessite en plus de la prise centrée, la stabilisation en flexion du poignet ou encore l'utilisation d'un tournevis, où la prise centrée s'associe à la prono-supination ;

- Les « prises plus action » : une partie de la main, le pouce en général, agit sur un objet tenu par le reste de celle-ci, c'est-à-dire que la main agit sur elle-même. Ces prises plus action sont innombrables dans la vie quotidienne, au point qu'on ne les remarque même plus :
  - o allumer un briquet ;
  - o faire tourner une toupie ;
  - o lancer une bille avec la détente du pouce ;
  - o tenir et utiliser des baguettes pour saisir la nourriture.

Ces quelques exemples, partiellement représentés dans les Figure 8 et Figure 9, illustrent la richesse des actions que nous réalisons quotidiennement et qui nous semblent naturelles. La capture et la reconnaissance de ces mouvements est pourtant loin d'être un problème trivial. Par exemple, pour un système basé sur des caméras, certains problèmes viennent s'ajouter à la complexité des mouvements, des problèmes d'occlusions, de couleurs, de décors, de lumières, de segmentation, etc. Une reproduction respectant la réalité le plus fidèlement possible nécessite donc un dispositif adapté et la prise en compte des problèmes inhérents à ce dispositif.

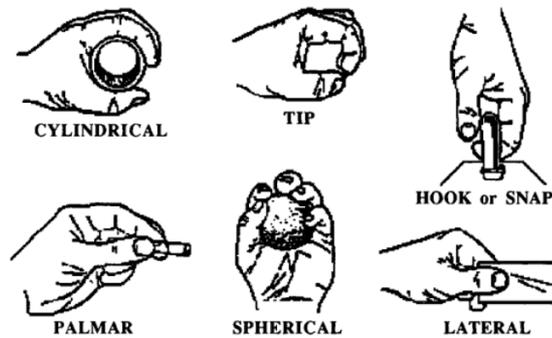


Figure 8 : Exemples de classification de postures de la main, tirés de Schlesinger (MacKenzie & Iberall, 1994)

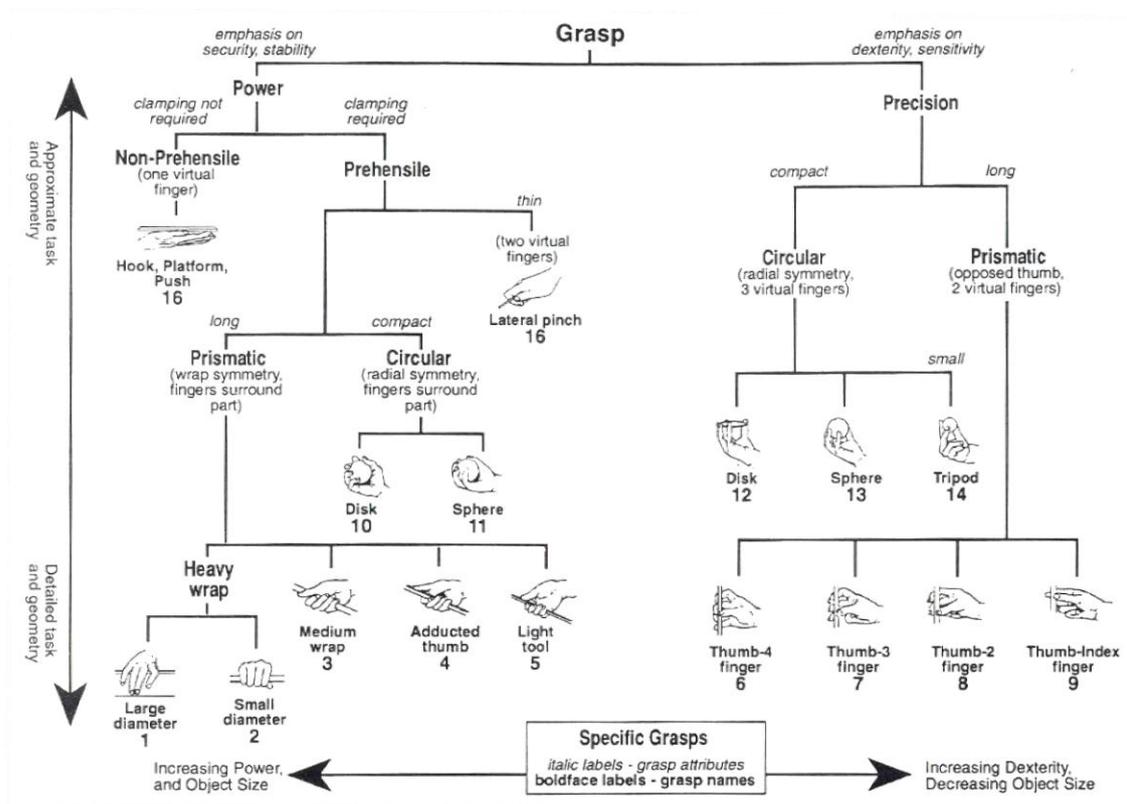


Figure 9 : Classification des posture de la main d'après Cutkosky (MacKenzie & Iberall, 1994)

### 3 Une taxonomie des interfaces actuelles de capture des mouvements

La complexité de la main humaine est telle qu'à l'heure actuelle, il est très difficile (voire impossible) de la reproduire mécaniquement à l'identique. Par ailleurs, comme l'a montré Bernstein, pour accomplir la plupart des actes moteurs, il n'existe pas qu'une seule façon de faire mais une infinité. Cela revient à dire que le système moteur humain est redondant et qu'il dispose en général de plus de degrés de liberté qu'il n'est vraiment nécessaire.

La capture de mouvements décrit l'activité d'analyser et d'exprimer le mouvement humain en termes mathématiques (Bray, 2006). La tâche de capture de mouvements peut être divisée en plusieurs phases distinctes : l'initialisation, le *tracking*, l'estimation de pose et la reconnaissance de gestes. La capture de mouvements a été réalisée pour la première fois par Eadweard Muybridge en 1887 dans ses travaux intitulés : « Animal Locomotion ». D'après (Menache, 1999) : « *Motion Capture is the process of recording a live motion event and translating it into usable mathematical terms by tracking a number of key points in space over time and combining them to obtain a single 3D representation of the performance* ».

La capture des mouvements de la main peut s'appuyer sur des techniques que nous qualifions de « logicielles » et de « matérielles ». Pour les premières, la capture ne se fait pas uniquement de façon logicielle, il reste bien entendu une interface physique qui permet de capturer un flux vidéo ou de données 3D. Cependant, cette dernière reste simple et la majorité du travail repose sur le traitement d'images et plus généralement sur la vision par ordinateur. Pour les secondes, il s'agit de systèmes de capture qui peuvent être de natures diverses : électromagnétiques, mécaniques, optiques ultrasoniques, etc. Nous ne nous intéresserons pas, dans cet état de l'art, aux surfaces tactiles multi-points, dans la mesure où elles ne permettent pas, tout comme les souris classiques, de travailler en trois dimensions de façon directe. Nous ne traiterons donc que des systèmes permettant de travailler (en même temps) sur 3 degrés de liberté, au minimum. Ces systèmes peuvent être catégorisés selon de nombreux critères, cependant nous avons décidé de les organiser selon 3 principaux : les degrés de libertés qu'ils offrent, la présence ou non de retour d'effort et leur transparence. Nous allons dans un premier temps définir ces notions, puis nous ferons une taxonomie détaillée de ces systèmes.

#### 3.1 Critères de catégorisation

D'après (Hayward & Astley, 1996), les périphériques peuvent être classés en 3 groupes, selon leur nombre de degrés de liberté :

- les interfaces possédant peu de degrés de liberté contrôlés (2 à 3). On les appelle alors LDOF (Low Degree of Freedom Device) ;
- les interfaces pour lesquelles un grand nombre de degrés de liberté contrôlés est possible (4 à 6) aussi appelées (HDOF : High Degree of Freedom Device) ;
- les interfaces à très grand nombre de degrés de liberté contrôlés (plus de 6), on les appelle (VHDOF : Very High Degree of Freedom Device).

Le second critère a trait à la présence ou non de retour d'efforts, également appelé retour kinesthésique. D'après le traité de la réalité virtuelle (Meseure & Kheddar, 2006), cela « correspond aux forces et aux déplacements qui s'opposent aux mouvements d'un objet quelconque que l'on manipule interactivement. Ces forces sont principalement ressenties par l'opérateur au niveau de ses articulations (poignée, coude, épaule, etc.) »

Enfin, le dernier critère repose sur le port ou non de « matériel » par l'utilisateur, comme des capteurs, des pastilles, des gants, un exosquelette... Les systèmes sont alors qualifiés de transparents (ou non) par certains auteurs, et d'intrusifs ou de non-intrusifs par d'autres comme par exemple (Bray, 2006). D'après (Fuchs & Mathieu, 2003), « au niveau physique, l'interface est transparente si elle ne gêne pas, ou peu, physiquement l'utilisateur, en particulier dans ses mouvements ». D'après (Collet, 1999), « tout système qui par son fonctionnement oblige à porter de l'équipement ou génère des perturbations visuelles ou sonores est appelé système intrusif. » Par soucis de cohérence, dans ce mémoire, nous adopterons uniquement le terme utilisé par Fuchs et Mathieu.

## 3.2 Solutions à base de capteurs

### 3.2.1 Type de capteurs

Il existe plusieurs types de capteurs. D'une part, ceux dits « incrémentaux », utilisant le même principe de fonctionnement que le joystick. C'est-à-dire que le déplacement de l'objet contrôlé se fait dans le sens où l'on pousse le senseur et ce de façon incrémentale. Autrement dit, lorsque l'objet (par exemple une souris 3D) est en butée, le déplacement continue dans la même direction. La vitesse est proportionnelle à l'amplitude du mouvement ou à la force appliquée. La rotation se fait également à une vitesse proportionnelle à celle générée par l'utilisateur (Flückiger, 1998). Les capteurs mesurant les déplacements sont généralement basés sur un système optique (photorécepteurs et diodes).

D'autre part, il existe d'autres senseurs qui sont dits « absolus ». Ces derniers sont composés d'un élément mobile qui se déplace dans l'espace et dont la position est calculée par rapport à un élément de base fournissant une référence fixe. Les moyens de positionnement peuvent être dans la majorité des cas mécanique, magnétique, optique ou à ultrasons (Flückiger, 1998).

L'ensemble des capteurs détaillés dans les parties suivantes est représenté dans la Figure 10. Des illustrations de chaque périphérique présenté dans cette partie sont regroupées sur la Figure 12 page 32 (pour les capteurs VHDOF) et sur la Figure 23 page 61 (pour les capteurs HDOF)

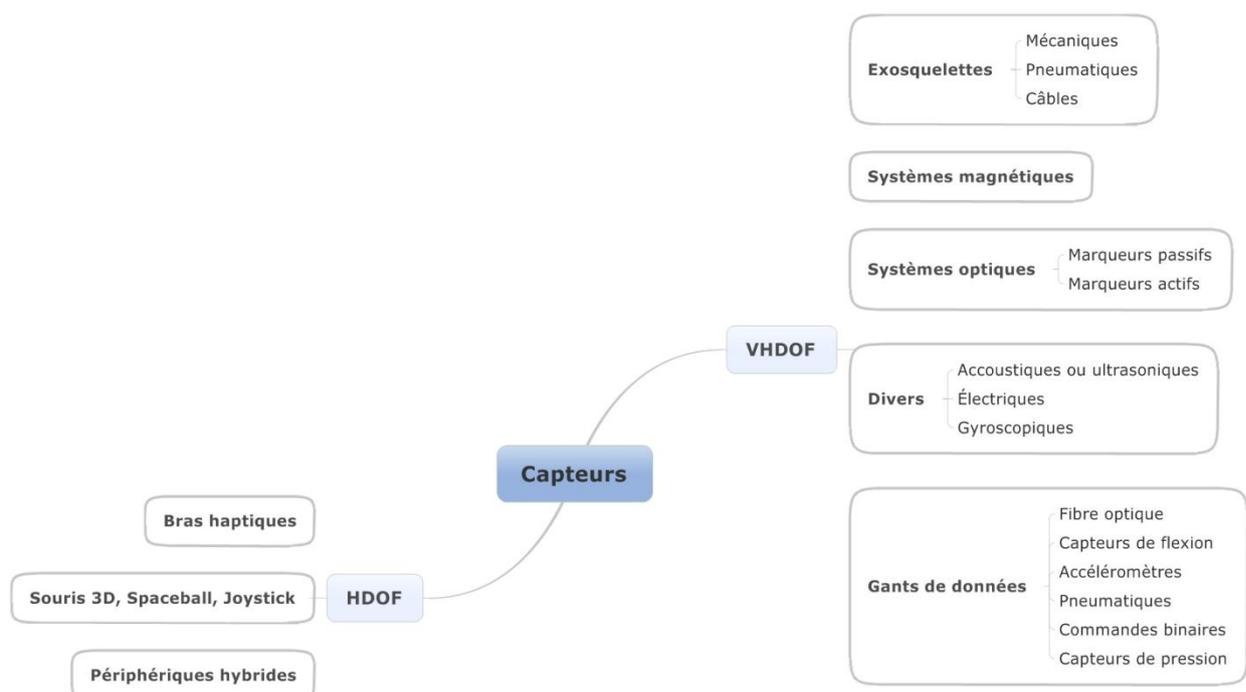


Figure 10 : Capteurs regroupés par catégories

## 3.2.2 Les capteurs « *Very High Degree of Freedom* »

### 3.2.2.1 Exosquelettes

Un des moyens les plus connus pour capturer les mouvements des mains consiste à utiliser des exosquelettes (par exemple ceux utilisés pour les gants à Effet Hall), des armatures et des articulations solidaires d'interrupteurs, de potentiomètres ou d'autres systèmes mécaniques complexes comme des fils retenus par des moteurs ou encore des vérins. L'avantage de ces systèmes est la précision des mesures qui en résultent (les données rotationnelles fournies en temps réel sont d'une très grande fiabilité, un codeur optique pouvant comporter plusieurs centaines de milliers de traits par tour). Cependant, cette précision est obtenue au détriment du poids, de l'autonomie et de la liberté de mouvements des utilisateurs, gênés par des liaisons filaires la plupart du temps. Les données sont cependant plus faciles à manipuler que celles des mouvements isolés nécessitant un appel à la cinématique inverse (comme dans le cas des systèmes optiques par exemple). Les systèmes utilisant des exosquelettes ont également l'avantage de ne pas induire de problèmes d'occlusion et fournissent par conséquent, à tout moment, des données sans pertes de signal. Enfin, un autre avantage important est la possibilité d'appliquer un retour d'effort. D'autres périphériques existent mais ils n'apportent cependant pas d'innovation particulière par rapport à ceux listés ici. Ce sont la plupart du temps des périphériques similaires, qui utilisent les mêmes principes ou technologies tout en essayant d'en améliorer les caractéristiques.

#### Périphériques à système mécanique

Parmi les périphériques les plus utilisés on trouve par exemple le DHM (Dextrous Hand Master), qui utilise des capteurs à effet Hall. Il a été originellement créé par le Utah/MIT en 1987. Il possède 20 ddl avec une précision allant jusqu'à 0,5 degré et pèse environ 300 grammes. Contrairement aux gants de type « DataGlove » (5DT) et « PowerGlove » (Nintendo), qui ne mesurent que la flexion des doigts, le DHM est capable de mesurer les flexions de chaque phalanges et se révèle à la fois plus précis et moins sensible à la taille de la main de l'utilisateur. Il est en revanche relativement lourd et nécessite une liaison filaire. De plus, il ne permet pas de connaître la position de la main de l'utilisateur dans l'espace ; il faut donc l'associer à un capteur supplémentaire pour déterminer sa position et son orientation. De nombreux travaux et études ont été réalisés sur ce type de périphérique. Une évaluation de ce périphérique a été réalisée par (Sturman & Zeltzer, 1994). Une seconde version plus précise (Makower et al., 1990), développée et vendue par (Sturman, 1992) utilisait un total de 20 capteurs à effet Hall comme potentiomètres pour déterminer la flexion des 3 articulations de chaque doigt, l'abduction/adduction pour chacun d'eux et 4ddl pour le pouce. Ce périphérique pouvait donc être utilisé pour la reconnaissance de gestes et de postures complexes, en revanche un temps conséquent était nécessaire pour l'installation et la désinstallation. D'autres versions améliorées ont depuis été créées par le laboratoire de Biorobotique de l'université McGill au Québec (Rohling & Hollerbach, 1993) ou encore par le laboratoire de robotique de Paris (Turki & Coiffet, 1995; Tzafestas & Coiffet, 1997). Enfin, des versions simplifiées et allégées (possédant par exemple 10 ddl et pesant 130g (Noaman et al., 2008)) et/ou sans fils ont vu le jour. Le prix d'un DHM oscillait entre 15000\$ et 18000\$ dans les années 1990.

Parmi les systèmes mécaniques, on trouve également le CyberGrasp et le CyberForce d'Immersion (Immersion, 2011) qui permettent un retour d'effort perpendiculaire à l'extrémité de chacun des 5 doigts. Le retour d'effort dans le contexte des téléopérations apporte généralement une amélioration significative dans le temps mis pour l'accomplissement des tâches (Buttolo et al., 1995; Hill, 1977; Massimino & Sheridan, 1994). En revanche leur poids est supérieur, environ 450g pour un CyberGrasp.

Ce dernier permettant le retour d'effort fut l'un des premiers exosquelettes commercialisé. Il nécessite, cependant, l'usage d'un CyberGlove pour capturer les mouvements. Ce type de gants est par exemple utilisé dans des opérations de télémanipulation de main motorisée (Turner et al., 2000).

#### Périphériques à système pneumatique

On trouve par ailleurs des systèmes pneumatiques permettant un retour d'effort, mais ils sont moins fréquents. Le gant Rutgers Master II créé par l'université Rutgers (New Jersey) ne pèse que 80g et permet un retour d'effort grâce à des vérins pneumatiques. Seul le petit doigt n'a pas de retour d'effort mais ce type de gant est moins fatiguant à utiliser que les autres en grande partie grâce à son poids réduit (Bouzit et al., 2002). Les gants de données pneumatiques convertissent la pression générée par la flexion du doigt en un signal électrique. La résolution permise est de l'ordre de 0,1° en angle et de 0,3mm en longueur, l'imprécision de position étant due au fait que les bases des pistons sur la paume ne sont pas fixes. De plus, la bande passante étant relativement faible pour des actionneurs pneumatiques (moins de 20Hz d'après (Hollerbach et al., 1992)), cela pose donc des problèmes pour simuler des surfaces rigides. Ce type de périphérique requiert souvent par ailleurs l'usage des gants de données de type 5DT.

#### Périphériques à câbles

Enfin on trouve les interfaces à câbles comme le SPIDAR (SPace Interface Device for Artificial Reality) développé au Tokyo Institute of Technology. (Sato, 2002) l'a décliné dans plusieurs versions. Toutes celles-ci ont en commun d'être constituées d'un certain nombre de blocs moteurs placés au sommet d'une pyramide et reliés au préhenseur par des câbles tendus qui s'enroulent sur des poulies entraînées par des moteurs et qui permettent d'assurer à la fois un tracking et un retour d'effort. La première version (le SPIDAR I) ne fournit qu'un retour d'effort 3D pour un seul point (Hirata & Sato, 1992). La résolution de ce système est de l'ordre de 0,025 mm pour un espace de travail de 900 x 600 x 600 mm. Le SPIDAR-8 est sorti dernièrement et contient 24 blocs moteurs permettant d'appliquer un retour d'effort sur 4 doigts de chaque main (3 blocs moteurs par doigt). L'inconvénient de ce système est son espace de travail très limité à cause du nombre de câbles (Yamada et al., 2002). D'autres travaux portent sur ce type d'interfaces à câbles et tentent de mettre au point des systèmes similaires comme ceux de l'université de l'Ohio par exemple (Williams II, 1998). L'intérêt principal de ce type de périphériques est la très haute précision qu'ils offrent. En revanche, l'espace de travail est petit et très restreint par les câbles. Il existe néanmoins un modèle, le SPIDAR H, qui a été développé pour de grands environnements comme un CAVE et qui permet donc de travailler à l'échelle humaine. En revanche, ce modèle ne fournit que 3 ddl pour un seul doigt de chaque main (Hasegawa et al., 2006).

### **3.2.2.2 Systèmes magnétiques**

La deuxième catégorie très fréquemment utilisée est la capture de mouvement par systèmes magnétiques. Les systèmes de capture de mouvements magnétiques utilisent un ensemble de capteurs positionnés sur la main de l'utilisateur (sur chaque phalange) pour mesurer le champ magnétique basse fréquence généré par un émetteur source. Ces capteurs mesurent leur relation spatiale (position et orientation) par rapport à ce transmetteur central qui émet un champ électromagnétique en un point réel. Ces capteurs sont des bobines de fils électriques perturbant le champ magnétique émis par le transmetteur central aussi appelé antenne. Les capteurs et les sources sont câblés à une unité de contrôle électronique qui détermine leurs emplacements dans le champ. Dans chaque capteur, se trouvent 3 bobines de cuivre disposées selon les 3 axes comme le montre la Figure 11. Le calcul des différences de potentiel de chaque bobine des capteurs avec l'antenne donne par conséquent à la fois la position et l'orientation du capteur.

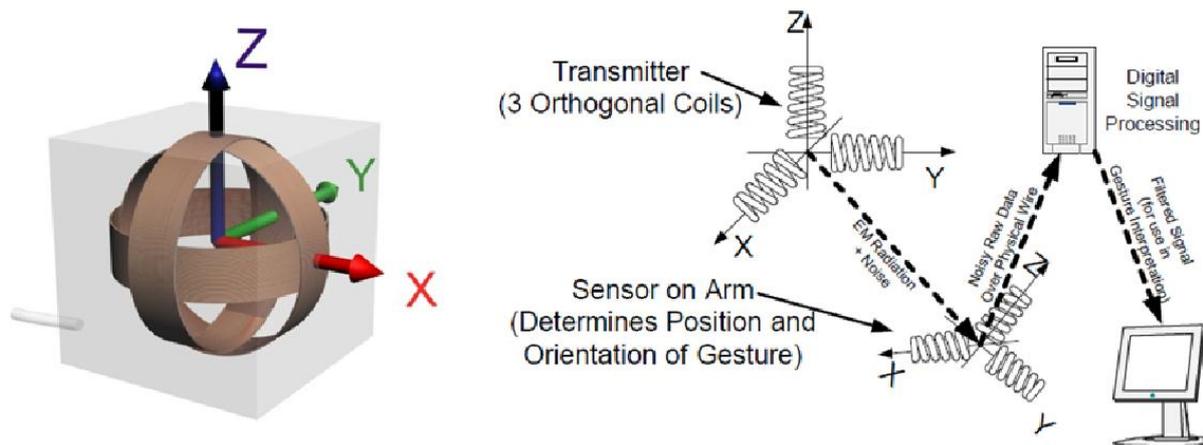


Figure 11 : à gauche, 3 bobines de cuivres disposées de façon orthogonales dans un capteur magnétique ; à droite, composants d'un système magnétique de tracking des mains (Miners, 2001)

Cette technologie a malheureusement de très nombreux inconvénients, ce qui ne l'empêche pourtant pas d'être très souvent utilisée pour la capture de mouvements du corps entier ou plus précisément des mains (Ma et al., 2011).

L'inconvénient majeur est que le traitement des données doit être suivi d'un calcul de cinématique inverse pour connaître la position la plus probable des doigts de la main de l'utilisateur. Cette technique est très souvent utilisée dans l'animation en temps réel, mais demande, du fait de la cinématique inverse, une certaine puissance de calcul. Cette approche produit, à partir de 6 capteurs, de bons résultats la plupart du temps, mais elle induit généralement une charge de calcul importante pour l'ordinateur ce qui peut conduire à une certaine latence entre les mouvements réels effectués et ceux retranscrits à l'écran. Elle s'ajoute à la latence due au système utilisé : 22ms pour un *Ascension Flock of Birds*, 4ms pour un *Polhemus FasTrack* et 20ms pour un *Polhemus IsoTrack II* (Miners, 2001). De plus, les capteurs ont tendance à bouger pendant les sessions de capture s'ils ne sont pas fixés directement sur la peau comme c'est le cas lors de l'utilisation de gants. Les technologies des marques *Ascension* et *Polhemus* diffèrent principalement par la fréquence d'échantillonnage, la précision, le volume de travail et la sensibilité aux perturbations magnétiques.

De même, ces systèmes nécessitent un passage obligatoire par une phase d'initialisation puis une synchronisation de tous les récepteurs, ce qui peut demander un certain temps avant de commencer l'utilisation effective. L'installation des différents capteurs doit se faire à chaque changement d'utilisateur et demande par conséquent un nouveau calibrage puisque la taille des mains varie d'une personne à l'autre.

Un autre inconvénient majeur est que chaque capteur nécessite son propre câble (qui est blindé pour limiter les perturbations et est donc généralement assez épais). Cela peut rapidement rendre le dispositif lourd à porter. Si chaque phalange, le dos de la main et le poignet possèdent chacun un capteur, on arrive à un total de 16 ou 17 capteurs (selon la pose de 2 ou 3 capteurs sur le pouce). Cela représente au minimum 16 câbles pour chaque main, ce qui est de nature à limiter la liberté de mouvements de l'utilisateur (Mitobe et al., 2006). Enfin, plus le nombre de capteurs utilisés est grand, plus cela peut mener à des erreurs de positionnement dues aux interférences engendrées par les capteurs les uns sur les autres.

Enfin, les systèmes de capture magnétiques ont des réactions très négatives face à des perturbations dues à des champs magnétiques extérieurs (moteurs à proximité, téléphones portables, sources d'alimentation, etc.) ou à des pièces métalliques dans la zone de capture. L'utilisation d'un tel système doit donc se faire dans un environnement contrôlé dans lequel on aura réduit au maximum les structures métalliques, notamment dans les sols, les murs et les plafonds. De plus, le taux d'échantillonnage de ces systèmes est relativement faible en raison du filtrage effectué pour tenter d'éliminer le bruit : 144Hz pour l'*Ascension Flock of Birds*, 120Hz pour le *Polhemus FasTrack* et 60Hz pour le *Polhemus IsoTrack II*. Enfin, la zone de capture est très petite par rapport à celle que permet un système optique.

En revanche, l'avantage principal de ce type de système réside dans l'absence d'occlusion comme cela peut se produire avec les systèmes optiques. La résolution est assez bonne avec une erreur possible de l'ordre de 0,5 mm et 0,1 degré pour une distance de 30,5 cm entre l'émetteur et le récepteur pour un *Ascension Flock of Birds* par exemple (Miners, 2001). Différentes recherches ont été menées pour tenter de rendre ce type de système plus léger et portable (par connexion *Bluetooth*) comme par exemple les travaux de (Bezdicsek & Caldwell, 2006).

### 3.2.2.3 Systèmes optiques

Il existe principalement deux technologies utilisées dans la capture de mouvement optique : celle dite réflexive, aussi appelée passive, et celle à lumière pulsée (la plupart du temps à base de diodes électroluminescentes), aussi appelée active. Les systèmes optiques de capture de mouvements utilisent très fréquemment des caméras vidéo propriétaires pour suivre les mouvements des marqueurs réfléchissants (ou des LEDs dans le cas des systèmes à lumière pulsée) attachés à des endroits particuliers du corps ou de la main dans le cas qui nous intéresse ici. Des systèmes à une ou deux caméras sont adaptés pour capturer les mouvements du visage, tandis que des systèmes de 8 à 16 caméras (voire plus) sont nécessaires lorsqu'on veut capturer les mouvements d'un corps entier. La capture optique est considérée comme le « haut de gamme » dans le domaine de la capture de mouvement. Les systèmes passifs peuvent aller de 15 000\$ (pour un système simple et peu performant) à 250 000\$ (voire plus) et nécessitent souvent un grand espace dédié. Ces systèmes ont l'avantage d'être très configurables (il est possible de placer les marqueurs où l'on veut) et produisent des données avec 3 degrés de liberté pour chaque marqueur et les informations de rotation doivent être déduites de l'orientation relative d'au moins trois marqueurs et ce grâce à la cinématique inverse.

Pour les deux technologies, les positions 3D de chaque marqueur sont obtenues par la triangulation des données de prise de vue d'au moins 2 caméras synchronisées : la synthèse des coordonnées (x,y) d'un même objet vu sous différents angles permet de déduire les coordonnées (x,y,z) de cet objet. La valeur en niveau de gris de chaque pixel peut être utilisé pour fournir une précision du niveau du demi-pixel en trouvant le centre de la gaussienne ainsi formée. Ce système engendre des problèmes optiques complexes (comme la parallaxe, la distorsion des lentilles utilisées, etc.) qu'il faut bien évidemment prendre en compte. Une bonne calibration de tous ces paramètres permet une grande précision des données récoltées. Un objet avec des marqueurs attachés à des positions connues (*rigid body*) est utilisé pour calibrer les caméras et obtenir leurs positions ainsi que pour calculer la distorsion de chaque lentille. Si au moins deux caméras calibrées voient un marqueur, un repère en 3 dimensions peut être obtenu.

La capture optique offre une grande liberté de mouvements mais a l'inconvénient majeur de l'occlusion possible de certains marqueurs qui engendre des pertes de données, même si des algorithmes de plus

en plus performants permettent de déterminer par le calcul les positions manquantes (Boulic et al., 1998; Herda et al., 2000, 2001). La multiplication des caméras minimise également ce problème mais augmente d'autant le prix et la puissance de calcul nécessaire. Le second inconvénient est le temps important demandé par l'installation de tous les marqueurs sur la main. Là aussi, il faut avoir recours à la cinématique inverse pour déterminer la position de chacun des doigts. Enfin, cette méthode, tout comme celle utilisant l'électromagnétisme, ne permet pas d'avoir du retour d'effort sans l'usage de matériel supplémentaire.

### Les marqueurs passifs

Les systèmes optiques passifs reposent sur l'utilisation de petits capteurs réfléchissants, appelés marqueurs, fixés sur la main. Les caméras émettent un rayonnement (en général rouge et/ou infrarouge), qui est réfléchi par les marqueurs puis renvoyé à ces mêmes caméras. Le seuil de la caméra peut être ajusté de sorte que seuls les marqueurs réfléchissants qui brillent soient échantillonnés en ignorant donc la peau et les tissus. Contrairement aux systèmes actifs ou de type magnétique, les systèmes passifs ne nécessitent pas que l'utilisateur porte un équipement électronique et donc de câbles. A la place, des balles en caoutchouc avec des bandes réfléchissantes sont attachés sur les vêtements ou directement sur la peau. Ce type de système peut capturer un grand nombre de marqueurs à la fois et à des cadences aussi élevées que 2000 Hz. Le taux de rafraîchissement pour un système donné est souvent équilibré entre la résolution et la vitesse de capture : un système à 4 mégapixels fonctionne normalement à 370Hz, mais il est possible de réduire la résolution pour atteindre un rafraîchissement de 2000Hz. En revanche un traitement manuel ou informatique est nécessaire pour pouvoir exploiter les données de positions des marqueurs. Ce type de système est très utilisé pour la capture de mouvements précise (par exemple pour la robotique) et rapide comme dans l'analyse d'un geste sportif ou d'un musicien virtuose (Goebl & Palmer, 2009).

### Les marqueurs actifs

Les systèmes dits actifs utilisent des LED à la place de marqueurs passifs. Les caméras, sensibles à un seul type de longueur d'onde, permettent de visualiser les marqueurs sous forme de tâches vidéo blanches ou en niveau de gris. Ce type de système triangule les positions en illuminant très rapidement toutes les LEDs alternativement de sorte à n'en allumer qu'une seule à la fois. Il est cependant possible d'allumer plusieurs LEDs en même temps à condition qu'un logiciel se charge de les identifier grâce à leurs positions relatives les unes par rapport aux autres, ce qui demande une puissance de calcul supérieure. L'alimentation de chaque marqueur peut être fournie de façon séquentielle en phase avec le système de capture, fournissant ainsi une identification unique de chaque marqueur pour une image donnée. Cela réduit toutefois le taux de rafraîchissement résultant. La capacité d'identifier chacun des marqueurs de cette manière est très utile dans le cadre d'application temps réel car on sait alors la position d'un marqueur directement et sans risque de confusion. En effet, la technologie active permet d'éviter plusieurs problèmes qui sont, souvent, inhérents aux systèmes passifs comme par exemple la confusion de marqueurs, les données bruitées ou manquantes et les réflexions parasites. Contrairement au système passif, les LEDs sont alimentées et émettent elles-mêmes leur propre lumière. Puisque la loi des carrés inverses stipule qu'une énergie lumineuse est inversement proportionnelle au carré de la distance de l'origine de cette source lumineuse, (autrement dit qu'un sujet à une distance donnée ne reçoit qu'un quart de la lumière que reçoit un autre sujet à une distance 2 fois plus petite), il suffit donc d'augmenter la puissance d'émission des LEDs pour augmenter les distances ainsi que le volume de capture.

Des systèmes comme *PhaseSpace*<sup>1</sup> utilisent la technologie active et peuvent atteindre des précisions très importantes grâce à des caméras 12 mégapixels, possédant des taux de rafraîchissement allant jusqu'à 480 images par seconde. Ce système permet de réaliser une capture en temps réel et d'obtenir une précision allant jusqu'au demi millimètre. De plus, la latence est inférieure à 10ms et il est possible de mettre jusqu'à 256 marqueurs actifs uniques sans aucun soucis de confusion entre eux. Cette société a également décliné un modèle uniquement pour la capture des mouvements de la main : les *PhaseSpace Gloves*. Ce sont des gants sur lesquels sont fixées 8 LEDs et qui coûtent environ 1000\$ pièce auxquels il faut rajouter un système de capture de type IMPULSE à 8 caméras d'un montant d'environ 56000\$<sup>2</sup> (53000€<sup>3</sup>).

#### 3.2.2.4 Autres technologies

D'autres technologies, moins répandues, existent également. Par exemple les systèmes acoustiques ou ultrasoniques, qui ont un fonctionnement semblable aux systèmes optiques : un émetteur produit un son qui est capté par 3 récepteurs positionnés de façon orthogonale. La triangulation permet de positionner dans l'espace un récepteur selon le temps que met le son à lui parvenir. Si le problème d'occultation est absent, le système est sensible aux réflexions sonores parasites (Sturman & Zeltzer, 1994). L'inconvénient majeur réside dans l'impossibilité de superposer plusieurs sons : chaque capteur doit donc émettre séquentiellement. Le gant *PowerGlove* de *Mattel* (Sturman & Zeltzer, 1994), créé en 1989 et initialement vendu 75\$ comme contrôleur pour la console de jeu Nintendo, se servait de cette technologie pour déterminer la position, le roulis et le lacet de la main dans l'espace. En revanche, les flexions des doigts (hormis le petit doigt) étaient mesurées grâce à un système à encre conductrice, ce qui permettait d'avoir 4ddl, les valeurs n'étaient pas très précises (les flexions des doigts étaient uniquement codées sur 2 bits) mais son faible prix était très attractif. Des développeurs ont détourné son utilisation afin de pouvoir l'utiliser sur ordinateur, ce qui constituait une alternative peu chère pour capturer les mouvements de la main (Kadous, 1995; Pausch, 1991). Le *Z-Glove* (Zimmerman et al., 1986) développé en 1986 par *VPL* est également un gant qui utilisait des capteurs ultrasoniques. Ce dernier a été développé en premier lieu pour des applications qui nécessitaient des manipulations d'objets avec la main et utilisait également de 5 à 15 capteurs (à fibre optique) pour mesurer la flexion des doigts. Deux capteurs ultrasoniques servaient à déterminer la position, le roulis et le lacet de la main dans l'espace. Le manque de précision de ses capteurs en a limité son utilisation, des tests ayant révélé que la précision ne dépassait pas les 5 à 10 degrés en rotation (Wise et al., 1990).

Parmi les autres technologies peu utilisées, citons également les traqueurs électriques, qui permettent de mesurer les signaux électriques générés par l'activité musculaire (électromyogramme) grâce à des électrodes placés à la surface de la peau (Saponas et al., 2009; Saponas et al., 2010; Wheeler & Jorgensen, 2003) ou encore des traqueurs gyroscopiques qui mesurent des moments angulaires et des accélérations par des capteurs inertiels placés sur les doigts.

#### 3.2.2.5 Gants de données

Il existe de nombreuses sortes de gants qui reposent sur différentes technologies, nous allons citer ici les principales. Les gants de données (encore appelés gants électroniques, gants numériques ou gants sensitifs) sont les interfaces les plus courantes pour la détection partielle ou totale des mouvements relatifs des doigts par rapport au poignet. Ils utilisent différentes technologies de capteurs généralement

---

<sup>1</sup> <http://www.phasespace.com/productsMain.html>

<sup>2</sup> <http://www.worldviz.com/purchase/pricelist.html>

<sup>3</sup> <http://www.est-kl.com/prices.html>

placés dans le gant ou dessus et sont à fibre optique, à lamelles (ou à encre conductrice), à capteurs de pression, à accéléromètres ou encore pneumatiques.

### Les gants à fibre optique

Un des premiers gants instrumenté décrit dans la littérature est le *Sayre Glove* développé par Thomas Defanti, Daniel Sandin et Rich Sayre en 1977 (Sturman & Zeltzer, 1994). Ce gant utilisait des capteurs basés sur la lumière avec des tubes flexibles. Une lumière est envoyée à l'entrée du tube et un photorécepteur est placé à l'autre extrémité. Lorsque le doigt se plie, la quantité de lumière qui traverse le tube varie ce qui permet de déterminer le degré de pliure du doigt. Ce gant possédait 7dof, cependant il a été créé pour le contrôle multidimensionnel de *sliders* et autres *widgets* 2D et n'avait pas la sophistication ni la précision requise pour déterminer les postures de la main ou pour permettre la reconnaissance des gestes.

Les gants plus récents reprennent ce principe. Chaque doigt du gant comporte une fibre optique qui épouse sa forme. Une lumière est envoyée d'un côté de la fibre et un capteur est positionné à l'arrivée de la fibre, lorsque le doigt est plié, cela provoque donc une atténuation de la lumière, ce qui permet alors de déterminer le degré de fermeture du doigt.

Les gants les plus connus utilisant cette technologie sont les *DataGloves* développés par *Fifth Dimension Technologies* (5DT). On trouve par exemple le *DataGlove Ultra 5* ou *14*. Ce sont des gants de données avec 5 ou 14 capteurs qui mesurent la flexion des doigts à l'aide de fibres optiques sur 8bits (256 valeurs de flexion de doigts), chaque gant possède également deux capteurs d'inclinaison pour le tangage et le roulis de la main. La version la plus classique, à 5 capteurs, qui permet de mesurer 7dof, fonctionne grâce à une connexion USB, un câble de 3m et est vendue 950€. Une version plus précise (à 10 bits soit 1024 valeurs de flexion possibles) à 14 capteurs (2 par doigt) est vendue 4500€. Il existe également un kit sans fil, fonctionnant en *Bluetooth*, qui peut se rajouter au *DataGlove* et qui coûte 1200€, à ajouter au prix du gant. Enfin, on peut trouver les versions *DataGlove 5* et *14 MRI* avec une transmission des données effectuée par fibre optique (et non par câble électrique) ce qui permet de pouvoir les utiliser dans l'imagerie à résonance magnétique (IRM) sans risque de perturbation du signal. Pour ces modèles il faut compter environ 4600\$ pour la version 5 capteurs et 9100\$ pour celle avec les 14 capteurs. Le taux de rafraîchissement des données est d'environ 75Hz et le poids d'un gant sans fil est de 300g. Puisque ces gants ne mesurent qu'une approximation de la flexion des doigts (malgré les 1024 valeurs différentes qui peuvent être détectées en théorie), la précision globale n'est pas suffisante pour faire de la reconnaissance de gestes très évoluée. En revanche, c'est suffisant pour détecter si un doigt est levé (pointage) ou si le poing est fermé par exemple.

### Les gants à capteurs de flexion

Les gants de données à capteurs de flexion (aussi appelé gants à lamelles) mesurent la variation de résistance d'une lamelle de silicium, de plastique ou de métal recouverte d'une fine couche d'encre conductrice, lorsque le doigt sur lequel elle est fixée se déforme.

Parmi les gants les plus connus utilisant cette technologie, on trouve les *P5* (qui utilisent également un système optique) ou encore les *CyberGloves*, développés initialement par Kramer dans ses travaux sur le « talking Glove » (Kramer & Leifer, 1987) et qui a déposé un brevet sur cette technologie (Kramer, 1991). D'après (Kessler et al., 1995), le *CyberGlove* permet de détecter au degré près les flexions des doigts. Le *CyberGlove II* est un gant sans fil avec 18 ou 22 capteurs haute résolution qui mesurent les angles produits par les flexions. La version sans fil à 18 capteurs est disponible pour 12000\$ et la version câblée à 22 capteurs pour 18000\$. Le taux de rafraîchissement est de 90 enregistrements par seconde

en moyenne. La version 3 du *CyberGlove* code les informations sur 12 bits et un capteur supplémentaire a été ajouté sur la paume de la main. Les données sont plus précises et peuvent être envoyées par *Wifi*. Il est enfin possible d'ajouter le *CyberTouch* afin d'améliorer le *CyberGlove II*. Ce système est une option qui permet le retour d'informations tactile. Ainsi, le gant équipé disposera de six petits stimulateurs vibrotactiles sur chaque doigt et sur la paume, chaque stimulateur pouvant être individuellement programmé pour varier la force de sensation de contact. Cette version enrichie coûte 22000\$. Une caractéristique intéressante du *CyberGlove* est qu'il numérise la tension de sortie de chaque capteur et ensuite modifie la valeur en utilisant une fonction de calibration linéaire. Cette fonction utilise des valeurs de gain et d'offset pour représenter la pente (le coefficient directeur) et l'ordonnée à l'origine de l'équation linéaire. Cette équation permet une calibration logicielle du gant et le rend ainsi plus robuste pour une grande variété de taille de main (LaViola, 1999).

Un autre modèle de gant qu'on peut trouver est le *DataGlove* de *X-IST*. Ces gants détectent les pressions exercées ainsi que les inclinaisons de la main (sur 2 axes). Deux modèles de gants sont disponibles, soit avec 5 capteurs de pression, soit avec 5 capteurs d'angle de flexion. Il existe une version plus précise : le *DataGlove HR System* qui possède 14 capteurs d'angle de flexion (2 sur le pouce et 3 sur les autres doigts). Les prix varient selon le nombre de capteurs par gant : 1300€ pour une main avec 5 capteurs de pression ou de flexion, 3950€ pour 2 mains avec 5 capteurs de pression, 5 capteurs de flexion et 2 capteurs d'inclinaison et enfin, 5500€ pour 2 mains avec 14 capteurs de flexion par main et 2 capteurs d'inclinaison. Ces gants nécessitent une liaison filaire et pèsent 150g. Le taux de rafraîchissement de ces gants varie de 100 à 200Hz par capteur selon les réglages de filtrage des données choisis.

Le *DG5-VHand* vendu par *DGTech* fonctionne avec le même principe et permet une précision théorique de 10 bits (soit 1024 valeurs différentes possibles de flexion). Il possède 5 capteurs (un sur chaque doigt) et il existe une version sans fil qui fonctionne en *Bluetooth*. La précision de l'orientation de la main est de 0,5 degrés d'après le constructeur et le taux de rafraîchissement est de 25Hz. Un accéléromètre 3 axes permet d'avoir les accélérations de la main et d'en déduire le tangage ainsi que le roulis. Les capteurs peuvent s'enlever et le gant est lavable.

Enfin, le *ShapeHand* développé par *Measurand* utilise également cette technologie. Avec ses 40 capteurs, il est le gant doté du plus grand nombre de capteurs. Ce système a le très grand avantage d'être réversible et donc de pouvoir s'adapter aussi bien sur un gant gauche que sur un droit. De plus, les gants peuvent être changés facilement (en cas d'usure), être lavés (pour des raisons d'hygiène) et avoir 3 tailles différentes (petit, moyen et grand) pour s'adapter facilement à la morphologie de la main de l'utilisateur. Le *ShapeHand* est léger et peut également se décliner en version sans fil.

### Les gants à accéléromètres

Ce type de gant repose sur l'utilisation d'accéléromètres : des capteurs sensitifs répartis le long de deux axes orthogonaux (X et Y), à l'intérieur de l'accéléromètre, fournissent une mesure proportionnelle au déplacement d'un poids, suspendu par des « ficelles » par rapport à sa position au repos. Comme la masse se déplace par rapport au centre, soit à cause d'une accélération soit à cause d'une inclinaison par rapport au vecteur gravitationnel, le capteur peut être utilisé pour mesurer des positions angulaires absolues. L'axe des Y pointe vers l'extrémité du doigt et l'axe X mesure les roulis de la main ou les abductions individuelles de chaque doigt quand ceux-ci sont en position verticale (Hernandez-Rebollar et al., 2002).

L'*AcceleGlove* est constitué d'un ensemble de six accéléromètres à trois axes chacun : un sur chaque doigt et un sur le dos de la main. Ces six accéléromètres déterminent leur orientation par rapport au

vecteur gravitationnel. Les capteurs sont placés sur le dos des phalanges médiales, sur le dos de la phalange distale du pouce et sur le dos de la paume. Les capteurs peuvent être posés directement sur la main ou bien sur un gant. Ce système est très léger (54g) et permet 18 ddl, mais souffre toutefois d'un taux de rafraîchissement des données relativement faible (35Hz). Ce gant peut se décliner en trois tailles (6 à 7 pouces, 7 à 7,5 et 7,5 à 8,5). Cette technologie a également été utilisée dans les travaux de (Lobo et al., 2011) et (Trindade et al., 2011). Il sert principalement pour la reconnaissance de gestes et ne permet pas de déterminer la position de la main dans l'espace, il faut donc lui ajouter un système magnétique, par exemple. Un gant coûte 500\$.

L'*Acceleration Sensing Glove* possède également cinq accéléromètres fixés à l'extrémité de chacun des doigts, plus un placé sur le dos de la main (Perng et al., 1999). Il est possible de reconnaître vingt-huit postures efficacement à partir de ces mesures. Par contre, le fait que ces mesures soient faites en référence à la gravité implique que certaines postures, pourtant très différentes visuellement, ne peuvent pas être différenciées par le gant (Naud, 2008).

#### Les gants pneumatiques

Les gants pneumatiques convertissent la pression générée par la flexion du doigt en un signal électrique proportionnel au degré de pliure du doigt (Sun et al., 2009; Sun et al., 2008). En 1987, lorsque *VPL Research* a sorti son premier gant de données, *Nissho Electronics* est devenu leur distributeur exclusif et leur partenaire technique au Japon. Depuis, *Nissho* a grandement amélioré et développé le concept de ce gant, ce qui l'a conduit à commercialiser le *SuperGlove* en mai 1995 pour environ 4500€ le gant (chaque gant additionnel coûtait 1800€). Ce gant fonctionnait grâce à un système pneumatique qui permettait de connaître la pression exercée par chacun des doigts. Le gant possédait 10 capteurs et une option sans fil était disponible. La précision en flexion était de 0,3 degrés.

Il existe également des gants pneumatiques qui permettent uniquement un retour d'effort sur chaque doigt comme le *PneuGlove* (Connelly et al., 2010). Contrairement au *Rutgers Master II* (cf. 3.2.2.1 p. 20) qui utilise des vérins pneumatiques pour le retour d'effort, ce gant utilise quant à lui des chambres à air. Ce gant est principalement utilisé dans des applications de rééducation de la main (Fischer et al., 2007; Jia et al., 2007) mais il peut servir également à la manipulation d'objets en environnement virtuel (Connelly et al., 2009). Un système sans fil, le *Shadow Monitor* (Gentner & Classen, 2009; Simone et al., 2007), peut lui être ajouté afin de mesurer les degrés de pliure des doigts grâce à des capteurs de flexion. Le *PneuGlove* peut aussi bien être utilisé avec des objets réels que virtuels et permet de réduire les mouvements de prise et relâche d'objets. Son intérêt est assez limité car il ne permet, avec ses cinq chambres à air, que de forcer l'extension des doigts. Il reste également possible de simuler le contact avec des objets en remplissant plus ou moins les chambres à air. La technologie pneumatique sous cette forme est cependant très peu utilisée.

#### Les gants de commandes binaires

Ce type de gant est très différent de ceux que nous avons pu voir jusqu'à présent et relève d'un usage assez limité dans le sens où il n'est pas possible de connaître la position des doigts dans l'espace ni même leur flexion. Ces gants sont capables de déterminer uniquement si les extrémités des doigts entrent en contact les uns avec les autres. Il est possible de détecter les pincements car lorsque les doigts se touchent, ils forment ainsi des boucles conductrices.

Les *PinchGloves* (Fakespace, 1997) dont le premier prototype a été créé par *Mapes* (University of Central Florida) ont été rachetés par *Fakespace Inc.* qui les a ensuite commercialisés. Ces gants sont un des systèmes les plus connus utilisant ce principe. Ils sont livrés par paire et leur prix est de 1750€. Leur

intérêt est limité si l'on souhaite réaliser une capture de mouvement fidèle des doigts. De plus, il n'est pas possible de fournir en temps réel une représentation virtuelle de la posture de la main de l'utilisateur. L'utilisation directe et sans calibration constitue toutefois leur avantage.

Le *Peregrine Glove* est, lui aussi, un gant utilisant des boucles pour connaître les différentes zones qui sont entrées en contacts. Il permet de reconnaître une trentaine de combinaisons différentes et se connecte à l'ordinateur via un câble USB. Il est possible de programmer chacun des contacts afin de lui associer une touche du clavier ou une action à effectuer. Son prix est bien inférieur au *PinchGlove*, puisqu'il coûte 150\$. La fixation qui permet de transmettre les informations de contact à l'ordinateur peut se retirer très facilement grâce à un système aimanté, permettant ainsi de changer facilement de gant. Ce dernier est lavable et peut donc être nettoyé pour des questions d'hygiène, sans que cela ne risque d'endommager les pistes électriques qui sont robustes et prévues pour un usage militaire. Un système de calibrage permet d'adapter le *Peregrine Glove* à la taille de la main de l'utilisateur et les zones de contacts peuvent être redéfinies pour améliorer la précision et faire en sorte que le gant réponde de façon optimale à chaque utilisateur.

#### Les gants à capteurs de pression

Ce type de gant donne à peu près les mêmes résultats que les *PinchGloves*, c'est-à-dire qu'il est possible de dire quels doigts sont entrés en contact avec un objet. L'avantage de ce gant est qu'il n'est pas nécessaire, contrairement *aux PinchGloves*, de faire toucher deux doigts entre eux, car il n'y a pas de boucle électrique, le fonctionnement repose sur des capteurs de pression. Lorsqu'un doigt touche quelque chose (un autre doigt, un objet, ou encore une surface quelconque) la pression exercée sur le capteur varie, ce qui permet de déterminer quel doigt est entré en contact.

Le *TouchGlove (I-CubeX)* possède 6 capteurs de pression par gant (un sur chaque extrémité des doigts et un sur la paume) et coûte 350\$. La zone de détection de pression est de 13mm et l'envoi des données se fait de façon filaire. Son prix est bien inférieur aux *PinchGloves* mais les possibilités sont supérieures puisqu'il n'est pas limité aux contacts entre les doigts uniquement.

#### **3.2.2.6 Synthèse**

Les performances de tous ces gants sont inégales et varient également d'une technologie à l'autre. De plus, la majorité des gants présentés ici impose un calibrage avant chaque utilisation. De même, pour la plus grande partie d'entre eux, les tailles sont uniques et par conséquent mal adaptées à la morphologie des mains de chaque utilisateur, ce qui conduit à des données imprécises voire erronées dans de nombreux cas. La taille de la main est alors considérée comme une source de bruit (Hernandez-Rebollar et al., 2002). Enfin, les *CyberGlove*, *DataGlove*, *Dexterous Hand Master (DHM)*, et le *PowerGlove*, donnent uniquement une approximation globale du degré de pliure d'un doigt et donc de la forme générale de la main (Sturman, 1992). Il n'est malheureusement pas possible de connaître la flexion exacte de chacun des doigts, ce qui peut être un problème non négligeable si l'on veut une mesure précise pour enregistrer ou pour faire de la reconnaissance de posture ou de gestes.

Si tous les périphériques présentés précédemment permettent généralement plus de 6 degrés de libertés (en admettant l'utilisation d'un nombre suffisant de capteurs optiques, mécaniques ou électromagnétiques), ils nécessitent cependant tous l'usage de matériel à fixer sur la main de l'utilisateur. On peut donc les qualifier de périphériques « non-transparents » car ils demandent d'une manière ou d'une autre, le port de capteurs afin de déterminer les mouvements effectués. La section suivante décrit succinctement des périphériques de réalité virtuelle qui permettent de capturer les mouvements de la main. Même si ces derniers suivent les mouvements d'un seul point à la fois et

autorisent au maximum 6ddl, ils sont cependant couramment utilisés dans les tâches de manipulation d'objets en environnement virtuel.

### 3.2.3 Les capteurs « *High Degree of Freedom* »

#### 3.2.3.1 Bras haptiques

Les bras haptiques permettent de suivre les mouvements de la main selon 3 à 6 degrés de liberté, mais également de lui appliquer un retour d'effort selon 3 degrés de liberté (en translation). Les différents modèles diffèrent par la zone de travail, le couple du retour d'effort, les degrés de libertés et leur prix qui est fonction des paramètres précédents.

Les bras à retour d'effort ont un fonctionnement isotonique c'est-à-dire que le déplacement de l'objet virtuel est égal au déplacement de l'extrémité de l'interface. D'après la définition du dictionnaire Larousse, l'adjectif isotonique désigne «une contraction musculaire telle que la force développée reste constante alors que la longueur du muscle diminue». Ainsi, un périphérique isotonique doit avoir une résistance constante (le plus souvent nulle).

Le principal inconvénient des bras à retour d'effort est d'une part le nombre limité de degrés de liberté qu'ils permettent (le mouvement d'un seul point est suivi), d'autre part la zone de tracking est relativement faible pour les petits modèles. Les grands modèles peuvent également induire un phénomène d'inertie. Plus on souhaite que la zone de travail et la force de la rétroaction soient grandes et plus le prix du matériel augmente ainsi que son encombrement. En revanche, l'avantage de ce type de périphérique est leur précision.

#### 3.2.3.2 Souris « 3D », Spaceball et joystick

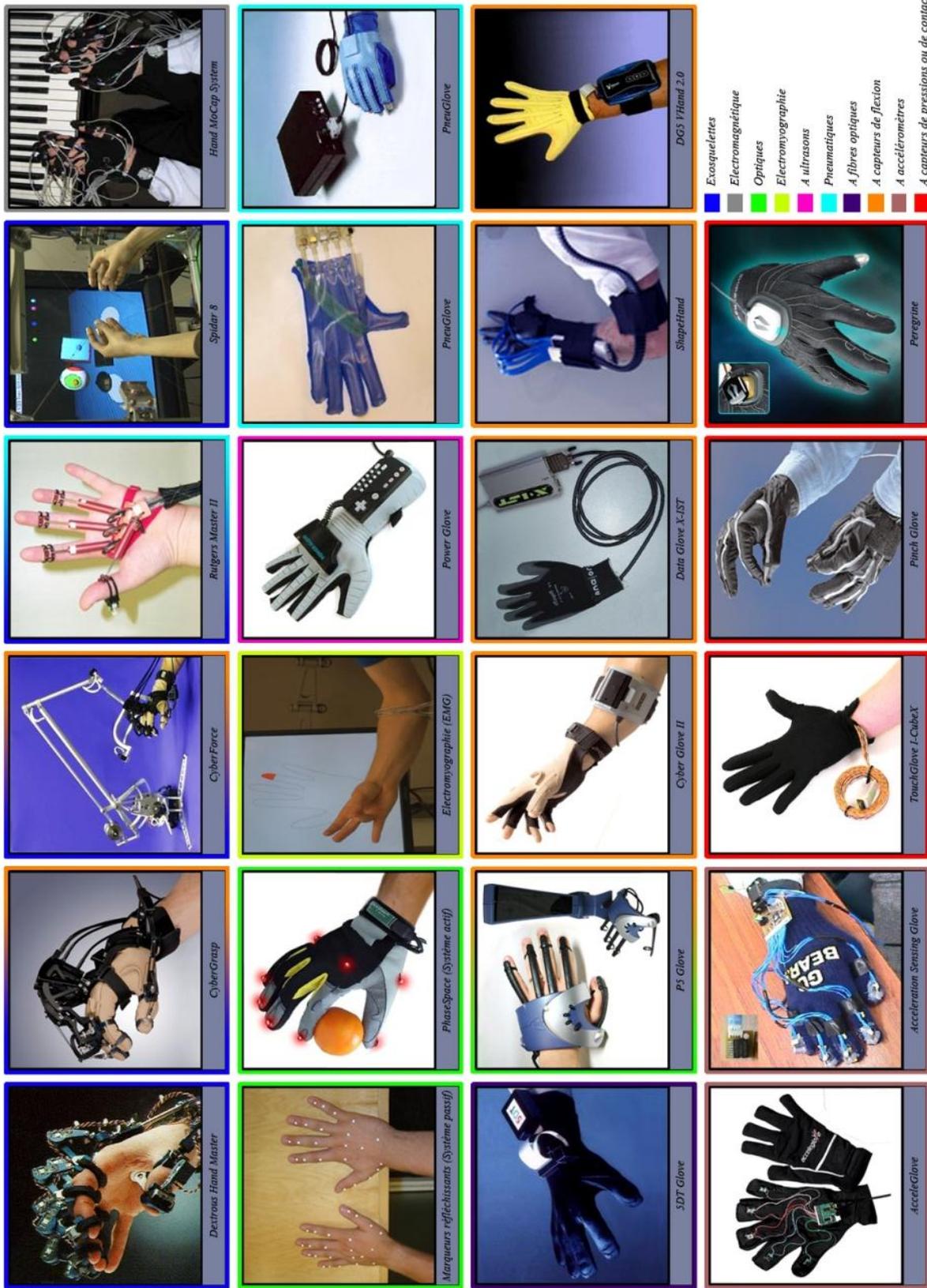
Les souris 3D, les *Spaceball* ainsi que joysticks les plus récents permettent d'avoir au minimum 3 degrés de liberté. Ces périphériques sont d'un coût relativement faible (\$40 à \$400) mais ont l'inconvénient d'avoir des fonctionnements isométriques ou élastiques (Gosselin et al., 2003), contrairement aux traqueurs magnétiques, mécaniques ou optiques, appelés senseurs « absolus ». L'utilisation de senseurs incrémentaux induit que le déplacement de l'objet contrôlé se fait dans le sens où l'on pousse le senseur, la vitesse étant proportionnelle à la force appliquée. La rotation se fait également à une vitesse proportionnelle à celle générée par l'utilisateur (Mizuuchi, 2006). Le geste de l'utilisateur ne se fait donc pas à l'échelle 1. Ainsi, un périphérique isométrique mesure des forces ou des couples mais ne bouge pas. Les joysticks, la *SpaceMouse* ou encore la *Spaceball* sont des périphériques isométriques (Casiez, 2004). Cette dernière permet de mesurer les forces et les couples appliqués par l'utilisateur suivant six degrés de liberté.

Les périphériques isométriques ont une résistance infinie et les périphériques isotoniques ont une résistance nulle ou constante. Entre ces deux modes se trouvent les périphériques élastiques qui ont une résistance variable. Pour ce type d'interface, la force de résistance augmente avec le déplacement de l'effecteur qui est replacé au centre (dans une position neutre) grâce à des ressorts la plupart du temps. Les joysticks ou la *SpaceMouse* sont des périphériques élastiques (Casiez, 2004).

#### 3.2.3.3 Périphériques hybrides

Certains périphériques, hybrides, combinent plusieurs des techniques citées précédemment et permettent ainsi de suivre les mouvements de la main de l'utilisateur selon 6 degrés de liberté. La *WiiMote* de *Nintendo* en est certainement l'exemple le plus connu et le plus répandu. Ce dispositif combine une manette, équipée de plusieurs capteurs (caméra infrarouge, accéléromètres, ...), et un

élément fixe (*Sensor Bar*) comprenant des LED infrarouges. Un calcul de triangulation permet à la manette de calculer sa position par rapport aux LED IR. Si l'orientation peut-être améliorée avec un module supplémentaire (un gyroscope intégré au *Motion Plus*), la capture de la position impose d'avoir toujours la « sensor bar » (et donc les LEDs IR) dans le « champ de vision » de la manette. De plus, ce type de périphérique n'offre le suivi à 6 ddl dans l'espace que d'un unique point (la main). Ce suivi est loin d'être précis.



Systèmes de capture de mouvements des mains « matériels »

Figure 12 : Systèmes de capture de mouvements des mains « matériels »

### 3.3 Solutions basées sur le traitement d'images

Toutes les solutions vues précédemment reposent sur l'utilisation de matériels et de divers capteurs. Cependant, de nombreuses recherches se consacrent à la reconnaissance et au suivi des mouvements de la main sans avoir recours à l'utilisation de ceux-ci, et s'appuient sur une ou plusieurs caméras. Ces solutions sont alors basées sur le traitement d'images à partir d'un ou plusieurs flux vidéo. Ce domaine s'appelle la vision par ordinateur (aussi parfois appelée vision artificielle, vision numérique ou plus récemment vision cognitive) et traite aussi bien des informations en 2D qu'en 3D. La vision par ordinateur requiert généralement une combinaison de traitements bas niveau (pour améliorer la qualité de l'image, supprimer le bruit, améliorer les contrastes, etc.) et haut niveau pour « comprendre » l'image, en tirer les informations utiles, reconnaître les formes et les « interpréter ». Dans ce chapitre, nous allons décrire les principales techniques utilisées pour la capture et la reconnaissance de mouvements. Une liste exhaustive est impossible à réaliser, au regard de la quantité phénoménale d'applications, d'algorithmes et de traitements qui existent. Nous n'avons donc pas la prétention de toutes les lister et encore moins de toutes les connaître.

Le traitement d'image réalisé sur ordinateur demande une puissance de calcul conséquente et il est important de prendre en compte de nombreux paramètres selon le résultat désiré. Ainsi, l'éclairage est très important car les données extraites de l'image sont directement dépendantes de la qualité de cette dernière. Si l'image est très sombre et n'offre pas de contraste, il peut être difficile d'en tirer des informations pertinentes par exemple. De même, un traitement informatique peut donner d'excellents résultats si on prend le temps d'appliquer à l'image de nombreuses opérations (qui peuvent durer de plusieurs secondes à plusieurs heures), mais rend le résultat totalement inutile si on souhaite que le système permette un suivi des mouvements en temps réel. Ce critère correspond donc au temps de calcul maximum alloué aux différents traitements afin d'obtenir un résultat temps réel. Un problème fréquent dans la vision par ordinateur est le fond, c'est-à-dire l'environnement derrière la personne à capturer, qui peut être fixe et connu (fond uni de couleur verte par exemple), ou bien mobile et changeant (e.g., passage de personnes derrière l'utilisateur principal). Il faut également prendre en compte des problèmes complexes comme la distorsion des lentilles, des problèmes de parallaxe ou d'occlusion qui sont inhérents à tout système basé sur des caméras. De même, l'utilisation ou non de marqueurs ou repères visuels sur la main peut considérablement simplifier ou compliquer les traitements à réaliser. Enfin, il faut bien séparer le suivi de mouvements de la main qui demandent des traitements informatiques à part entière et la reconnaissance de ces mouvements qui reposent sur d'autres traitements et algorithmes. On peut distinguer deux types de reconnaissance : la reconnaissance statique et dynamique. La première se base uniquement sur l'image en cours et utilise les données spatiales, la deuxième sur une suite d'images et étudie l'évolution au fil du temps.

(Moeslund & Granum, 2001; Moeslund et al., 2006) ont listé les principales hypothèses relatives aux systèmes de capture de mouvements utilisant la vision par ordinateur :

- Le sujet reste à l'intérieur d'une zone déterminée (workspace) ;
- La caméra ne bouge pas ou alors de manière constante ;
- Un seul individu est présent à la fois dans le workspace ;
- Le sujet fait toujours face à la caméra ;
- Il n'y a pas d'occlusion ;
- Les mouvements sont lents et continus ;
- La lumière est constante ;

- L'arrière plan est statique et uniforme ;
- La posture de départ est connue ;
- Des marqueurs sont placés sur le sujet ;
- Les vêtements du sujet ont une couleur spécifique.

Ces quelques hypothèses sont des contraintes à prendre en compte dans de très nombreux systèmes afin qu'ils fonctionnent correctement. Cependant, chacun d'eux ne demande pas forcément de les respecter toutes à la fois. Certains n'ont pas besoin par exemple de marqueurs sur la personne et d'autres ne demandent pas que les vêtements du sujet aient une couleur spécifique. Cependant, de tous les systèmes basés sur la vision par ordinateur, aucun de ceux que nous allons décrire dans la suite de ce chapitre ne fonctionne sans avoir à émettre aucune de ces hypothèses. Nous verrons par la suite que certains systèmes très évolués et robustes permettent cependant de se dispenser d'une très grande partie de ces contraintes (cf. 4.3 p. 50).

Nous allons ici nous consacrer uniquement à la capture et au suivi des mouvements de la main en ne prenant en compte que les techniques dites temps réel. Les techniques qui ne sont pas temps réel ne nous intéressent pas ici car elles ne permettent pas d'interagir en direct avec l'environnement virtuel et ne sont donc pas adaptées aux tâches que nous souhaitons réaliser. Les systèmes que nous allons voir reposent dans la majorité des cas sur plusieurs techniques combinées et non sur une seule en particulier. Aucune technique ne suffit à elle seule à assurer une capture fidèle des mouvements de la main et des doigts. Il faut pour cela utiliser plusieurs méthodes afin de combler les lacunes de certains algorithmes et les renforcer avec les avantages d'autres. On trouve deux grandes catégories : les systèmes qui utilisent la vision monoscopique et ceux qui utilisent la vision stéréoscopique (ou multi-vues).

Les algorithmes proposés ont plusieurs objectifs. Le premier est une certaine efficacité algorithmique, c'est-à-dire que les applications doivent pouvoir suivre les mouvements à la cadence d'acquisition de la vidéo (entre 15 et 30 images / seconde ; 25 au minimum pour avoir du temps réel et 30 ou plus idéalement). Le deuxième objectif est la robustesse : les données visuelles extraites des images et utilisées pour le suivi sont souvent imprécises, bruitées voire parfois aberrantes, la luminosité peut varier, ou le fond changer et il faut que le système continue à fonctionner même dans ces cas là. Le dernier objectif est d'obtenir une bonne précision : deux types de problèmes peuvent être rencontrés : l'accumulation d'erreurs au cours du suivi qui peut conduire à une dérive progressive du résultat par rapport à la réalité et l'instabilité des paramètres qui peuvent induire une cohérence temporelle insuffisante (latence, ralentissement progressif).

En premier lieu, le tracking nécessite d'extraire l'objet que l'on souhaite suivre du fond (qu'il soit fixe ou mobile). Si le fond est fixe, il est relativement facile de soustraire la main : on peut par exemple prendre une image du fond sans la main et ensuite soustraire sur chaque image les pixels qui n'ont pas changé. On ne garde alors dans l'image que les pixels correspondant à la main. En vision par ordinateur, le terme utilisé pour l'opération qui consiste à identifier et séparer les différentes parties d'une image s'appelle la segmentation. Cette étape permet de séparer les objets pertinents du premier plan (la main par exemple) des informations inutiles (comme le fond) ; elle est souvent suivie de traitements pour réduire le bruit, améliorer la netteté, etc.

L'ensemble des techniques détaillées dans les parties suivantes est représenté dans la Figure 13. La « hiérarchie » représentée correspond au plan de cette partie. L'ensemble des illustrations de chacune des techniques est regroupé sur la Figure 15 page 46.

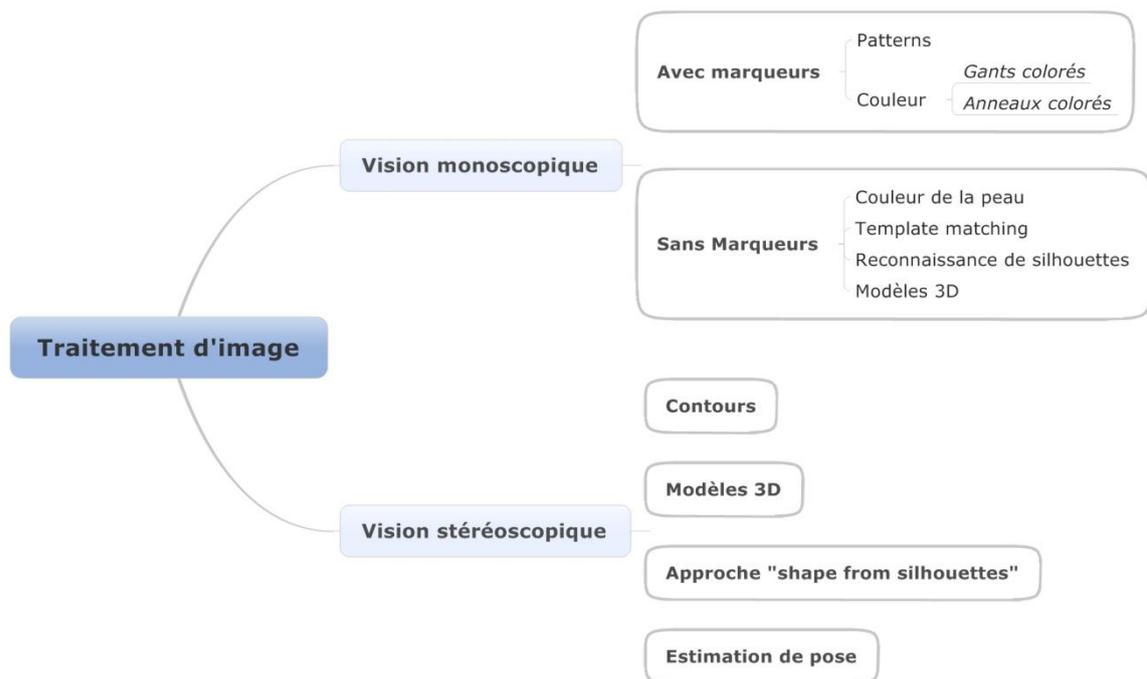


Figure 13 : Techniques de traitement d'image abordées dans notre travail

### 3.3.1 Vision monoscopique

La vision monoscopique ne permet pas à elle seule d'assurer un suivi précis des mouvements de la main en 3 dimensions, elle est donc souvent associée à d'autres techniques. Par exemple, la reconstruction de scènes 3D est possible en s'appuyant sur des indices visuels présents dans l'image. Cela nécessite cependant un apprentissage car il faut de nombreuses images de référence pour faire fonctionner ce genre de système qui ne permet pas d'obtenir une carte de profondeur exacte mais permet d'apporter des informations supplémentaires lorsqu'il est couplé à un autre système (Saxena et al., 2007). Parmi les systèmes monoscopiques, différentes approches sont possibles.

#### 3.3.1.1 Approches avec marqueurs

Dans cette catégorie, on trouve tous les systèmes qui utilisent des repères visuels pour aider à différencier les différentes parties de la main et faciliter les traitements informatiques. Ces repères peuvent être des bagues de couleurs qu'on fixe sur les phalanges des doigts, des gants colorés, des pastilles de différentes grosseurs ou couleurs, etc. La différence avec un système optique à marqueurs passifs est que l'image doit être traitée pour n'obtenir que les marqueurs colorés qui nous intéressent et supprimer tout le reste. Il n'est pas nécessaire de faire ces traitements avec un système optique à marqueurs passifs puisque seules des tâches blanches sont présentes dans l'image.

#### Systèmes reposants sur des « patterns »

La première technique repose sur l'utilisation de marqueurs spécifiques, qui se composent d'un cadre carré de couleur noire entourant une image binaire représentant un identifiant unique. Ces marqueurs sont très souvent utilisés pour la réalité augmentée. Une caméra détecte le pattern utilisé et en déduit sa position ainsi que son orientation en fonction des déformations de l'image. En réalité augmentée, ce type de marqueurs permet de plaquer des informations (2D ou 3D) en surimpression avec une orientation et une position correctes.

Ce procédé a été repris dans de nombreuses applications et un système a été mis au point pour détecter les mouvements des doigts de la main (Pamplona et al., 2008). Des cubes avec des marqueurs sur

chaque face sont disposés sur 4 doigts (tous sauf le pouce) et une caméra est placée au niveau de la paume de l'utilisateur. La caméra envoie le flux vidéo à un programme qui se charge, grâce à des techniques de vision par ordinateur (un algorithme de « *template matching* » (cf. 3.3.1.2 *template matching*) et des bibliothèques spécifiques (en l'occurrence *ARToolKitPlus* et *V-Art*), de retrouver les positions relatives des extrémités des doigts ainsi que leur orientation. Une fois que les extrémités sont détectées, il faut appliquer une technique de cinématique inverse pour estimer la position de chacune des articulations des doigts et ainsi recréer les mouvements de l'utilisateur dans l'environnement virtuel. Un capteur de mouvements a été ajouté afin de déterminer la position (X, Y, Z), le tangage, le roulis ainsi que le lacet de la main. Comme les positions et les orientations des extrémités des doigts sont relatives par rapport à la position et à l'orientation de la caméra et que ces dernières sont connues grâce au capteur de position, il est possible de connaître la position et l'orientation de chacun des doigts afin de les retranscrire en environnement virtuel. Le poids de ce système dépend du poids de la caméra utilisée mais pèse environ 200g (avec une caméra *PointGrey Research* qui pèse 115g). La résolution de la caméra est de 1024x768pixel pour un *framerate* de 30fps permettant ainsi des traitements temps réel. Cette technique a l'inconvénient d'utiliser des marqueurs et nécessite une calibration importante sans laquelle les données obtenues seraient complètement erronées. L'installation du système sur la main est assez longue et peu aisée et le confort est réduit de par la position et le poids de la caméra. En revanche, les détections des positions des extrémités des doigts sont précises (0,59mm d'erreur moyenne). Cependant, comme cette technologie fait appel à la cinématique inverse, elle donne donc les positions et orientations les plus probables des différentes articulations des doigts. Cela ne reflète pas forcément la réalité et la fidélité des mouvements en est donc réduite. De même, du fait de l'appel à la cinématique inverse, le système a une latence non négligeable de 237ms et fonctionne à 23fps ce qui permet tout de même un suivi des mouvements des doigts en temps réel et ce pour un coût très réduit.

#### Systèmes reposant sur la couleur

Ces systèmes utilisent, quant à eux, la couleur comme information de base pour suivre les mouvements des mains ou des doigts. Ils reposent sur l'utilisation de marqueurs de différentes couleurs (souvent uniques) placés sur les doigts voire sur toute la main ; de cette façon, il est possible de savoir où se situe chaque partie en fonction de la couleur qui lui est associée. Les systèmes se basant sur la couleur pour suivre les mouvements sont sensibles à la luminosité de la scène et travaillent dans différents espaces de couleurs. La majorité des traitements repose sur l'utilisation de filtres, afin de nettoyer les images, réduire le nombre d'informations et ne récupérer que les informations utiles pour le suivi des mouvements. L'usage de marqueurs colorés sur les gants aide à repérer plus facilement les différentes zones de la main sans avoir besoin de faire de segmentation (en principe).

#### Gants colorés

Différentes techniques sont employées comme par exemple la segmentation du gant en différentes parties. Chaque doigt à une couleur spécifique et des marqueurs sont placés sur les phalanges, il suffit alors de déterminer la position de chacun des marqueurs pour connaître la position des phalanges sans risque de confusion comme le fait (Geebelen et al., 2010). Le suivi des doigts se fait en temps réel et le système utilise 2 caméras pour déterminer la position 3D des marqueurs. Une calibration est nécessaire au début pour connaître les paramètres des caméras ainsi que leurs orientations respectives. Pour cela, un damier est utilisé ainsi que la méthode de (Zhang, 2000) pour le calibrage. Ce type de gants a été utilisée dans de nombreux projets et fait toujours l'objet de recherches (Dorner, 1994; Tokatli, 2005). Cependant dans de nombreux cas les marqueurs ne suffisent pas et il faut utiliser des techniques complémentaires pour connaître les positions et les orientations des articulations ou tout du moins les

estimer. Il est alors possible d'avoir recours à la cinématique inverse, à des bases de données et à bien d'autres techniques.

Une des méthodes parmi les plus récentes basée sur la couleur et qui donne d'excellents résultats est celle utilisée par le MIT avec leur gant coloré (Wang, 2008; Wang, 2011; Wang & Popovi, 2009). Le gant utilisé a un patchwork de 20 zones colorées de même taille avec 10 couleurs différentes qui sont placées de façon judicieuse de sorte à ne pas avoir d'ambiguïté quelque soit la position et l'orientation de la main. Pour que le système fonctionne correctement, il faut une lumière contrôlée comme dans la plupart des systèmes basés sur la couleur et également une calibration de la caméra. De plus il est nécessaire de faire une soustraction de fond afin de ne garder que les gants dans l'image. Le fond doit donc être fixe et ne pas changer au cours du temps ou a minima avoir une couleur très différente des 10 présentes sur le gant. La caméra se charge de capturer le gant et des algorithmes sont utilisés pour nettoyer l'image, simplifier les zones colorées et réduire les informations à une image de 40x40 pixels. Lorsque la simplification est effectuée, le résultat est comparé à une base de données contenant de très nombreuses postures possible de la main grâce à une méthode de « *template matching* » (cf. 3.3.1.2 *template matching*). Le résultat le plus proche est alors renvoyé et le modèle virtuel adopte la posture choisie dans la base de données grâce au squelette qui lui est associé. Ce système permet un framerate d'environ 10 à 15fps et accuse environ 160ms de latence, ce qui rend ce système un peu trop lent pour une application temps réel, mais qui peut suffire dans certains cas (Wang & Popovi, 2009). Du fait de l'utilisation d'une seule caméra, la précision est loin d'être bonne en ce qui concerne la profondeur et donc la position de la main dans l'espace en général (une erreur de 15cm est possible entre la position réelle et celle calculée). En revanche, comme le gant est coloré jusqu'au poignet, les orientations et les positions relatives des doigts par rapport au poignet sont assez bonnes. De même, grâce au pattern particulier des gants, certains problèmes d'occlusion peuvent être évités. La base de données est très complète, avec au total un ensemble de 18000 configurations différentes de doigts qui ont été enregistrées au préalable avec un *CyberGlove II* (Wang & Popovi, 2009). La base de données retourne donc un résultat proche de la réalité. Un algorithme de cinématique inverse a été ajouté afin de rendre les résultats plus précis entre le modèle proposé et le modèle réel, mais cela aux dépends du taux de rafraîchissement.

### *Anneaux colorés*

D'autres techniques reposent uniquement sur l'usage de pastilles ou d'anneaux à mettre au bout des doigts. Le principe est le même que précédemment, c'est-à-dire que le but est de repérer dans l'image la position des différentes couleurs et d'en déduire la position des doigts correspondants. L'avantage de ce procédé est qu'il n'y a pas besoin d'utiliser de gants particuliers ce qui présente des intérêts multiples. Premièrement du point de vue hygiénique, lorsqu'on travaille longtemps avec des gants les mains peuvent devenir moites assez rapidement (selon la sudation des personnes) et ainsi produire une sensation de gêne pour l'utilisateur, réduisant par la même occasion le sentiment d'immersion.

Ce type de marqueurs est par exemple utilisé dans le système réalisé par Mistry (Mistry et al., 2009). Un marqueur est placé sur 4 doigts (le pouce et l'index de chaque main) et une caméra placée sur le corps de l'utilisateur (au niveau du sternum) se charge de capturer la scène. Le fond peut changer et l'algorithme segmente les différents couleurs provenant des marqueurs positionnés sur les doigts. L'image est ensuite nettoyée et améliorée grâce à des filtres (des opérateurs morphologiques d'ouverture et de fermeture). Le centre des différentes taches sert alors à déterminer la position des extrémités des doigts dans l'espace. Les bibliothèques *EmguCV* et *OpenCV* sont utilisées pour réaliser les traitements d'images. Enfin, les positions sont analysées pour déterminer les mouvements qui ont été effectués et les actions correspondantes à déclencher. Pour reconnaître les mouvements, un modèle de

Markov caché (en anglais HMM pour *Hidden Markov Models*) est utilisé. Ce système à l'avantage d'être léger et ne gêne aucunement les mouvements de l'utilisateur, en revanche tous les mouvements doivent être exécutés devant soi dans une zone assez limitée qui varie en fonction du recul et de l'angle de vue de la caméra. L'usage d'une seule caméra ne permet pas de travailler en 3 dimensions et les différentes actions possibles avec ce système ne font appel qu'à 2 dimensions. Son intérêt est donc limité si l'on veut faire des interactions 3D. Le principe pourrait cependant être réutilisé mais il faudrait alors l'adapter afin de permettre la 3<sup>ème</sup> dimension, en ajoutant par exemple une caméra supplémentaire et travailler sur le principe de la stéréoscopie. (Chua et al., 2002) utilise également des marqueurs positionnés sur les doigts ainsi qu'une méthode basée sur le repositionnement d'un modèle 3D. Dans son cas les pastilles servent principalement à simplifier la reconnaissance des différents doigts (cf. 3.3.2.2 - *Systèmes basés sur des modèles 3D* p. 43).

### 3.3.1.2 Approches sans marqueurs

#### Système reposant sur la couleur de la peau

Cette technique utilise, quant à elle, la couleur comme information de base pour suivre les mouvements des mains ou des doigts. Parmi les techniques existantes reposant sur la couleur, certaines se basent uniquement sur la couleur de la peau pour extraire les mains des images et ensuite déterminer leurs déplacements au cours du temps (Jones & Rehg, 2002). Comme nous l'avons vu précédemment, les systèmes se basant sur la couleur pour suivre les mouvements sont sensibles à la luminosité de la scène et travaillent dans différents espaces de couleurs. Pour cela les traitements reposent sur l'utilisation de filtres. Le principe de base est d'utiliser un ensemble de plages de seuils pour chaque canal de l'image séparément. L'inconvénient majeur de cette technique est qu'il ne faut pas d'objet de la même couleur que celle dont on veut suivre les mouvements dans l'image sans quoi la détection ne sera pas correcte. De plus, la luminosité doit être constante car les couleurs changent en fonction de l'éclairage (Brethes et al., 2004; Malik & Laszlo, 2004). De même, lorsque l'algorithme se base sur la couleur de la peau, les personnes dont la peau est foncée ne se situent pas dans la même plage de couleur que les personnes avec une peau blanche. Pour remédier à ce problème différentes solutions existent, parmi lesquelles, le changement d'espace de couleur.

La plupart du temps les caméras envoient des images en mode RVB. Dans ce modèle, chaque couleur est exprimée par l'intensité de trois couleurs basiques : le rouge, le vert et le bleu. La forte corrélation entre ces composantes, la luminance et la chromaticité rend ce modèle très sensible aux changements de luminosité (Hassanpour et al., 2008). Il existe plusieurs autres espaces de couleurs qui permettent d'avoir la luminosité des couleurs dans un canal séparé. Les plus connus sont HSV (Hue, Saturation, Value), HSL (Hue, Saturation, Lightness), HSI (Hue, Saturation, Intensity), YUV, YCbCr ou encore RGB normalisé. Pour éviter les problèmes de luminosité et pouvoir travailler quelque soient les conditions d'éclairage, on peut aussi avoir recours à l'espace  $L^*a^*b^*$  où  $L^*$  est la luminance et  $a^*$  et  $b^*$  les coordonnées de chrominance. Cette technique est très souvent utilisée pour déterminer les zones de couleurs correspondant à la peau et donne de très bons résultats (Rautaray & Agrawal, 2011; Yuan et al., 2008). Cependant, si la personne entière est dans le champ de vision de la caméra, la tête sera également détectée. Il faut donc faire une segmentation des données pour bien séparer les différentes zones et les résultats peuvent être erronés si les mains se rapprochent trop ou touchent la tête. Pour palier ce problème des algorithmes comme le Mean-shift ou encore le cam-shift, qui travaillent sur les histogrammes de couleurs des mains, peuvent être utilisés (Kang & Jung, 2008; Smith et al., 2003). Il est également possible d'« entraîner » l'algorithme pour ne détecter que les mains comme le fait (Yuan et al., 2008).

Les exemples cités précédemment fonctionnent avec une seule caméra et ne permettent donc pas de suivre les mouvements de la main en 3D. Il est cependant possible d'utiliser ce principe avec des systèmes stéréoscopiques ou multi-vues, c'est pourquoi nous les avons tout de même présentés.

### Appariement de gabarit (« template matching »)

L'appariement de gabarit plus souvent connu sous son nom anglais « *template matching* » consiste à comparer l'intensité des pixels entre une image de référence (le « *template* ») et plusieurs sous régions de l'image à analyser. En pratique, cela consiste à « superposer » l'image de référence (qui représente en l'occurrence une posture de main quelconque) sur l'image à analyser en la balayant afin de couvrir toute sa superficie. Les endroits les plus propices à la présence de mains sont donc facilement identifiés grâce à des fonctions de coût (distances minimum entre le gabarit et la sous-région). Il existe plusieurs manières de mesurer la similarité entre deux images, la plus simple étant de mesurer l'écart des normes L1 et L2.

Équation 1 : Norme L1

$$\text{Norme L1: } \sum_{(i,j)} |a_{(i,j)} - b_{(i,j)}|$$

Équation 2 : Norme L2

$$\text{Norme L2 : } \sqrt{\sum_{(i,j)} a_{(i,j)}^2 - b_{(i,j)}^2}$$

Où  $a_{(i,j)}$  et  $b_{(i,j)}$  sont les valeurs du pixel  $(i,j)$  dans chacune des images à comparer. Plus ces valeurs sont petites, plus les images sont proches. D'autres manières existent comme le calcul de la cross-corrélation ou bien encore l'opération « OU Exclusif » suivi d'une négation sur chaque pixel correspondants des deux images.

La construction d'un gabarit efficace représente à lui seul un défi car il faut qu'il soit adapté au type d'objet recherché et peut donc nuire à une détection efficace et diminuer la précision des résultats. Il est possible de faire appel à une base de données afin de comparer plusieurs postures de mains différentes dans toute l'image. De cette façon, non seulement les zones où une main est présente seront détectées mais également quel type de posture représente le mieux la main.

Généralement, le calcul d'appariement du *template* sur l'image est informatiquement la partie la plus coûteuse de l'algorithmes de suivi. Cependant, s'il est effectué seulement dans une région de recherche relativement faible, le temps de calcul diminue de façon significative. L'intérêt de la soustraction de fond permet justement d'isoler les mains et donc de connaître les zones sur lesquelles le *template* doit être mis en correspondance. Cela réduit les zones de recherche à des petites parties de l'image uniquement. Cette méthode est très utilisée du fait de sa rapidité. Pour les algorithmes nécessitant un traitement temps réel, cette méthode est par conséquent toute indiquée. De plus, elle est simple à implémenter.

De nombreux projets de recherches utilisent cette méthode pour détecter et suivre les mouvements de la main. Le *template matching* peut reposer sur les contours de la main (Mohr & Zachmann, 2009; Stenger et al., 2007) qui sont très caractéristiques pour les objets articulés. Cependant, une bonne qualité des contours peut être parfois difficile à obtenir et rendue mauvaise à cause de l'éclairage, des paramètres de la caméra, des objets environnants, de la couleur du fond, des ombres, etc. Le *template matching* peut également reposer sur les silhouettes de la main (Mohr & Zachmann, 2010a, 2010b) ou la couleur (Stenger, 2006).

L'inconvénient de cette méthode est qu'elle nécessite de nombreux gabarits pour pouvoir effectuer une seule correspondance. Par exemple dans (Mohr & Zachmann, 2010b), l'algorithmes utilise 1536 *templates* différents pour être en mesure de reconnaître la position « Main ouverte » et 1536 *templates* pour la position « pointage ». Il faut donc au préalable une grande base de données avec de nombreux

exemples de main effectuant l'action recherchée. De plus si la méthode n'est pas optimisée ou si la taille de l'image dans laquelle on travaille augmente, l'algorithme peut devenir très rapidement coûteux en temps de calcul. Afin d'optimiser l'algorithme, les *templates* peuvent être rangés sous forme d'arbres, ce qui permet d'arriver au *template* final (celui considéré comme correct) beaucoup plus rapidement que si tous les *templates* étaient mis en correspondance chacun leur tour.

### Reconnaissance de silhouettes

Cette technique utilise la silhouette de la main de l'utilisateur pour positionner et faire correspondre au mieux un modèle 3D. La silhouette, calculée à partir de la projection du modèle 3D, est mise en correspondance avec la silhouette de la main réelle et la différence entre ces deux silhouettes est donnée sous forme de distance (Pavlovic et al., 1997; Potamias & Athitsos, 2008; Thayanathan, 2005). La position du modèle 3D est considérée comme étant bonne lorsque la distance est minimale. Cette technique peut également se servir de primitive pour aider au repositionnement du modèle 3D. Ces primitives peuvent être géométriques (points, lignes, cercles), le contour de la main 3D ou réelle, etc. Cela ressemble beaucoup à la technique du « *template matching* »

Ce principe est utilisé par le système développé à Oxford par (Prisacariu & Reid, 2011). Dans leur cas, un accéléromètre a été ajouté afin de lever les ambiguïtés lorsque plusieurs solutions sont possibles (plusieurs silhouettes peuvent correspondre). Les résultats du tracking sont efficaces, en revanche une seule posture de main est possible (l'index et le majeur levés, l'annulaire et l'auriculaire pliés et le pouce collé à l'index, cf. Figure 15 page 46). En revanche l'orientation de la main peut être quelconque. Le système fonctionne selon 6 ddl (3 pour la position et 3 pour l'orientation) et est robuste face aux changements de luminosité, au flou de mouvement (lors de mouvements rapides) et aux occlusions.

Le projet de recherche *Virtual Touch Screen (VTS)* de (Prisacariu & Reid, 2011; Tosas, 2006; Tosas & Bai, 2007) permet de suivre les mouvements des doigts de la main via un algorithme qui détecte la couleur de la peau et les contours de la main. La main est seulement détectée en 2 dimensions, mais il est possible de simuler la troisième dans une certaine mesure (par rapport à la proportion de la taille de la main dans l'image). Le suivi ne se fait donc pas de manière précise en 3D et les seules interactions permises sont en 2D. Une version améliorée (VTSS) par (Zhao & Dai, 2009) permet de suivre les mouvements de la main entière et non plus juste des extrémités des doigts. Pour cela, il utilise un modèle simplifié de main et une méthode de cinématique inverse. Étant donné que la version précédente pouvait donner les positions des extrémités des doigts, les positions et orientations des articulations peuvent être calculées si la position du poignet est connue. Le système fonctionne toujours sur le même principe et utilise une seule caméra, ce qui ne résout malheureusement pas le problème de positionnement 3D. Il est possible de connaître plus ou moins bien la posture de la main, mais en aucun cas de connaître sa position dans l'espace.

### Systèmes basés sur des modèles 3D

Les systèmes basés sur les modèles 3D (en anglais « *model based* ») ont pour objectif de recalibrer un modèle articulé 3D de la main en adaptant la posture la plus vraisemblable correspondant à l'image perçue par la caméra (Ouhaddi & Horain, 1998). C'est l'une des techniques les plus utilisées pour estimer les postures de la main à partir d'un seul flux vidéo. Si la majorité des systèmes possèdent un modèle générique (qui peut être la moyenne de plusieurs personnes), certaines systèmes permettent cependant, lors d'une initialisation, de créer un modèle sur mesure et adapté à la morphologie de la main de chaque utilisateur. Ce modèle prend donc la forme, l'apparence et possède une structure cinématique et une position initiale proche de celle de l'utilisateur. Ainsi, différents modèles existent

comme par exemple les modèles squelettiques (des segments de droite)(Dorner, 1994), des modèles volumiques (des cylindres par exemple) (Chua et al., 2002; Gorce & Paragio, 2005; Gorce et al., 2008; Rehg & Kanade, 1995; Romero et al., 2010) (Chua utilise également des marqueurs placés sur les doigts afin de faciliter la reconnaissance), ou encore des modèles surfaciques déformables ( $\beta$ -Splines ou maillage) (Heap & Hogg, 1996; Kuch & Huang, 1995). La plupart des systèmes supposent que la position initiale est connue comme position spéciale de départ ou est spécifiée manuellement. La projection du modèle de la main 3D est ensuite mise en correspondance avec l'image provenant de la caméra grâce à la minimisation d'une fonction de coût généralement basée sur les contours (Heap & Hogg, 1996; Rehg & Kanade, 1994; Rehg & Kanade, 1995; Stenger et al., 2003; Stenger et al., 2006), les silhouettes (obtenue grâce à la segmentation de couleur) (Ouhaddi, 1999; Ouhaddi & Horain, 1998; Wu et al., 2001) ou encore le flux optique (Lu et al., 2002; Lu et al., 2003). Tous ces exemples de travaux permettent donc de recalibrer un modèle 3D à partir d'une image 2D mais ne permettent malheureusement pas un positionnement dans l'espace du modèle numérique selon les 3 dimensions, puisque la profondeur est toujours manquante, il faut par conséquent ajouter un système de tracking 3D. A la place d'un système de tracking, une technique de stéréoscopie pourrait être utilisée ce qui permettrait une estimation de la position de la main dans l'espace grâce à la carte de profondeur.

L'inconvénient de toutes les techniques reposant sur la vision monoscopique est qu'il est impossible d'obtenir des informations 3D précises et donc un suivi des mouvements de la main et des doigts en 3 dimensions. Pour ce qui est des applications 2D, ces systèmes ont largement fait leurs preuves, mais ils ne peuvent malheureusement plus servir tels quels lorsque l'on veut faire de l'interaction 3D en environnement virtuel. Il faut alors les coupler avec d'autre technologie (des capteurs de positions par exemple) ou encore les adapter sur des systèmes stéréoscopiques. Grâce aux algorithmes de traitement d'images, il est possible de « transformer » un objet 3D en 2D (pour l'afficher sur un écran d'ordinateur par exemple) grâce aux projections, mais il n'est malheureusement pas possible de faire l'inverse car la projection ne conserve pas les distances, les angles ni le parallélisme. Pour conserver ces informations indispensables il faut faire appel à la stéréoscopie.

### *3.3.2 Vision stéréoscopique / multi-vues*

Pour remédier à l'absence de données de profondeur, il est possible d'utiliser au minimum 2 caméras qui grâce à la stéréoscopie peuvent apporter cette information. Une corrélation est effectuée entre les 2 caméras et la troisième dimension est ainsi calculée, ce qui permet en plus des avantages cités auparavant, de positionner la main dans l'espace et effectuer un réel suivi 3D. Il est à noter que la stéréoscopie n'est pas le seul moyen de récupérer des informations de profondeur en vision par ordinateur, mais c'est une des techniques les plus utilisées à l'heure actuelle. Tous les systèmes vus précédemment peuvent par conséquent être réutilisés dans les systèmes stéréoscopiques comme c'est le cas par exemple des systèmes basés sur la couleur de la peau (Elmezain et al., 2008) ou bien avec l'usage de gants colorés (Theobalt et al., 2004). Dans le cas de travaux de (Theobalt et al., 2004) les mouvements de la main sont suivis grâce à un stroboscope et des marqueurs colorés positionnés sur un gant. Le tout est capturé par un système à 4 caméras. Dans les travaux de (Dankers et al., 2007), la stéréoscopie est utilisée quant à elle pour suivre en 3D les mouvements de la main. Une segmentation à partir de la couleur de la peau permet d'isoler la main du fond. La segmentation est robuste et peu sensible aux changements de luminosité grâce au système stéréoscopique. Le système créé permet un traitement temps réel (27 fps) et des occlusions partielles ou complètes peuvent être gérées.

Notre cerveau détermine les informations de profondeur à partir de notre environnement de nombreuses façons : à partir d'indices de profondeur dans l'image, de mouvements de parallaxe, des

ombres, des textures, des occlusions, des indices oculomoteurs (la convergence et l'accommodation), de la perspective linéaire, des hauteurs relatives, des tailles relatives et grâce à la stéréoscopie.

La vision stéréoscopique par ordinateur est très utilisée pour la reconstruction de profondeur et se base sur les caractéristiques de la vision binoculaire humaine. En vision par ordinateur, c'est un ensemble de techniques mises en place pour reproduire la perception du relief à partir de deux images planes. Le principe est d'utiliser deux caméras en les espaçant de quelques centimètres (la distance interoculaire moyenne est de 6,3 cm) : les différences entre les 2 images capturées, dues aux angles de vue différents d'un même objet, permettent de calculer l'écart de position de chaque point composant les images et ainsi de reconstruire la carte de profondeur (« *depth map* ») de la scène (Scharstein & Szeliski, 2002). Il faut pour cela mettre en correspondance sur les deux images des points d'intérêts et les interpoler. Il est très souvent nécessaire de rectifier les deux images avant de les mettre en correspondance, à cause de la distorsion des lentilles des caméras et de leurs orientations respectives. Cette mise en correspondance peut être également faite avec des contours. Cependant, on peut s'interroger sur la précision de l'estimation de ces disparités en raison de l'absence de caractéristiques fortes sur une main (pas d'angles obtus, de changements de couleur nets, etc.).

De nombreux algorithmes existent pour obtenir cette reconstruction. Bien qu'ils soient de plus en plus performants (Stefano et al., 2004), ils doivent cependant encore être améliorés pour permettre un traitement temps réel avec des résolutions intéressantes. De plus, si le traitement pour reconstruire la profondeur peut se faire en temps réel, il demande tout de même un temps conséquent. Les traitements supplémentaires qui viennent alors s'ajouter pour faire la capture et le suivi des mouvements rendent, s'ils sont compliqués, le système final trop lent pour être considéré comme temps réel. Il est alors nécessaire d'utiliser un cluster de PCs. De même, la stéréoscopie nécessite une initialisation pour connaître les paramètres des 2 caméras (leurs orientations, les caractéristiques des lentilles, etc.) ainsi que leur espacement. Cette étape est obligatoire afin d'avoir une carte de profondeur précise.

### **3.3.2.1 Systèmes basés sur les contours**

Un système binoculaire est parfois utilisé uniquement pour réduire les occlusions ainsi que le nombre de solutions possibles comme par exemple dans le cas des systèmes basés sur les silhouettes (Kato & Xu, 2006) ou les contours (Gumpp et al., 2006; Romero et al., 2008). Cela permet de lever des ambiguïtés et ainsi donner un résultat plus précis (Delamarre & Faugeras, 1998; Dewaele et al., 2004; Kato et al., 2007; Rehg & Kanade, 1995). (Campos & Murray, 2006) utilise dans ses travaux plusieurs caméras au lieu d'un système binoculaire pour adapter un modèle 3D à partir des silhouettes détectées par les différentes vues.

(Gumpp et al., 2006) utilise également la stéréoscopie pour améliorer ses résultats quant au placement de son modèle 3D à partir des contours de la main. Une erreur en profondeur peut être perçue comme négligeable sur une caméra mais être visible sur la seconde, il faut alors repositionner le modèle 3D. Cela conduit, par conséquent, à des résultats plus précis. En premier lieu, une segmentation est réalisée à partir de la couleur de la peau dans l'espace de couleur HSV. Ensuite, le modèle 3D est positionné par rapport au contour de la main sur les images grâce à une fonction de coût. Ce système est temps réel et permet de positionner en 3D et orienter un modèle de main ainsi que ses doigts dans l'espace.

### 3.3.2.2 Systèmes basés sur des modèles 3D

Dans cette catégorie, on retrouve, tout comme en vision monoscopique, le recalage de modèle 3D sur les informations obtenues. Dans ce cas, le modèle n'est pas recalé sur une image 2D, mais à partir du nuage de points provenant de la carte de profondeur obtenue grâce à la stéréoscopie. Cette méthode a été utilisée dans de nombreux projets (Delamarre & Faugeras, 1998, 1999; Dewaele et al., 2004). Le modèle utilisé est constitué de cônes tronqués et de sphères et possède 27ddl. Une phase d'initialisation dans laquelle la position de la main est connue est nécessaire. Le suivi s'effectue ensuite à partir de cette position à l'aide d'un filtre de Kalman<sup>4</sup> pour prédire la position à l'instant suivant. Des forces sont appliquées au modèle 3D afin de le faire correspondre au nuage de points obtenu par la stéréoscopie. Le filtre de Kalman permet d'estimer la position suivante et donc dans quelle direction cette force doit être appliquée. Si la force appliquée ne permet pas une correspondance optimale du modèle 3D et du nuage de point, alors une autre force est appliquée dans une direction différente. Ce système, bien que fonctionnel, ne permettait malheureusement pas un tracking en temps réel de la main et des doigts au moment de l'écriture de la publication en 1998.

Les travaux de (Nirei et al., 1996) et de (Fujiki et al., 2005) reposent quant à eux sur un modèle 3D recalé à partir des formes des silhouettes des mains provenant des images 2D obtenues par les caméras. Le système développé est temps réel et permet d'estimer la pose de la main en 3D à partir d'un système binoculaire. L'usage de 2 caméras permet une meilleure précision et la stéréoscopie est principalement utilisée pour réduire les occlusions et éliminer les multiples solutions qui peuvent exister.

### 3.3.2.3 Approche « shape from silhouettes »

Il existe également des reconstructions multi-vues qui nécessitent plusieurs caméras tout autour de la main. La reconstruction du volume 3D s'appuie sur différentes méthodes. La méthode « *shape from silhouettes* » datant des années 70 consiste à prendre la main sous différents angles à l'aide des caméras et à soustraire l'image de fond sur chaque point de vue. (Laurentini, 1994) appelle « enveloppe visuelle » (« *visual hull* » en anglais), l'enveloppe 3D maximale d'un objet à modéliser dont la projection selon tous les points de vue est à l'intérieur de toutes les silhouettes. L'enveloppe visuelle est par conséquent reconstruite à partir de tous les contours de ces silhouettes. La reconstruction 3D surfacique permet une reconstruction efficace en temps réel, même si l'objet n'est pas vu par une des caméras (Michoud et al., 2006) : seule l'enveloppe visuelle de l'objet est prise en compte. Ainsi selon le positionnement et le nombre des caméras, l'enveloppe visuelle ne sera pas la même. Par exemple sur la Figure 14 (à gauche), l'enveloppe visuelle grise, qui représente l'intersection des différentes images perçues par les 3 caméras, est sensée représenter la boule orange. Si les caméras sont bien positionnées, il est possible d'avoir une enveloppe visuelle assez précise. Cependant, il est impossible d'avoir une enveloppe concave, elle est toujours convexe, ce qui apporte un manque de précision et peut poser problème dans certains cas. Plus le nombre de caméras est grand et meilleure est la représentation, cependant plus il y a de caméras et plus la puissance de calcul nécessaire est grande. Un projet français d'envergure utilise cette technique, il s'agit de la plateforme *GrImage* du laboratoire Gravir à l'INRIA Rhône-Alpes (Rhône-Alpes, 2007), qui permet de faire les reconstructions 3D de personnes entières ou d'objets, de les insérer dans des environnements virtuels, d'appliquer les textures réelles, d'ajouter des ombres, d'appliquer de la physique et de faire des interactions. Le temps réel est possible grâce à un cluster de pc et plusieurs caméras.

---

<sup>4</sup> Le filtre de Kalman est un estimateur récursif. Cela signifie que pour estimer l'état courant, seuls l'état précédent et les mesures actuelles sont nécessaires. L'historique des observations et des estimations n'est ainsi pas requis (Kalman, 1960).

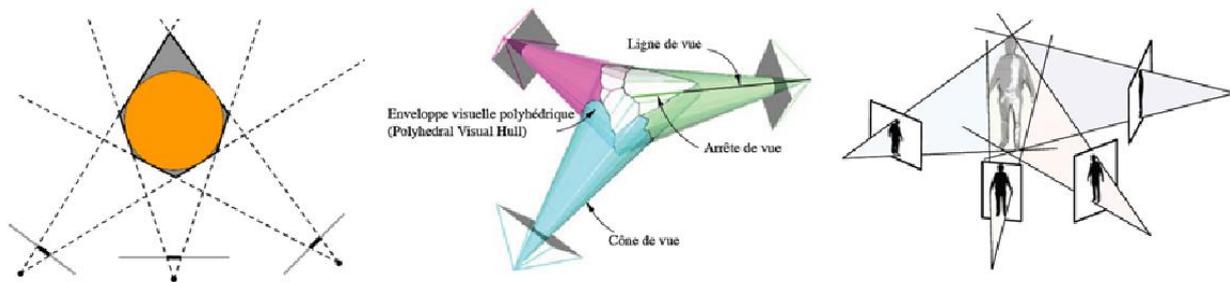


Figure 14 : Principe de la technique Shape-from-silhouettes

Cette technique a été utilisée par (Ueda, 2003) qui utilise une méthode qui estime tous les angles des articulations de la main pour manipuler un objet dans un environnement virtuel. Dans sa méthode, les régions des mains sont extraites des multiples images obtenues par le système multi-caméras. En intégrant les images des silhouettes obtenues par les caméras, une posture de la main est reconstruite sous forme d'un modèle constitué de *voxels* (des petits cubes) dont la précision dépend du pas de la grille utilisée (ce qui influence également le temps de calcul) (Cheung et al., 2000; Szeliski, 1993). Dans ce cas, la reconstruction n'est plus surfacique mais volumique, elle n'utilise pas les contours mais les silhouettes. Ensuite tous les angles des articulations sont estimés grâce à une correspondance en 3 dimensions entre le modèle reconstruit (en *voxels*) et un modèle 3D articulé. Des forces sont appliquées sur le modèle 3D afin de le faire correspondre au mieux au volume recréé à partir des silhouettes. Bien que cette technique soit simple et rapide, elle ne permet pas d'estimer les poses globales et locales de façon simultanée comme le fait Causo. Un état de l'art des méthodes de reconstruction 3D volumétrique d'enveloppe visuelle a été présenté par (Slabaugh et al., 2001).

L'approche « *shape from silhouettes* » a également été utilisée par (Causo et al., 2010). Un ensemble de 4 caméras est utilisé pour récupérer les silhouettes de la main à partir desquelles un volume est reconstruit. Ce volume constitue l'enveloppe de la main et est aussi composé de *voxels*. Lorsque le volume est récréé à partir des différentes vues, le modèle 3D est appliqué dessus et une fonction de coût est calculée. Lorsque la valeur est minimale, le modèle 3D est considéré comme correctement positionné et la posture est adoptée. Le fait d'utiliser plusieurs caméras permet de réduire les risques d'occlusions et améliore la précision lors de l'application du modèle 3D. Dans l'approche de (Causo et al., 2010), les postures globales et locales sont estimées simultanément en utilisant un filtre de Kalman (*Unscented Kalman Filter*, UKF) et non pas à partir de l'application de forces sur le modèle 3D. Au lieu d'être limité aux mouvements des doigts, leur système permet également d'orienter la paume de la main et de la positionner dans l'espace en même temps contrairement à (Ueda, 2003).

#### 3.3.2.4 Estimation de pose

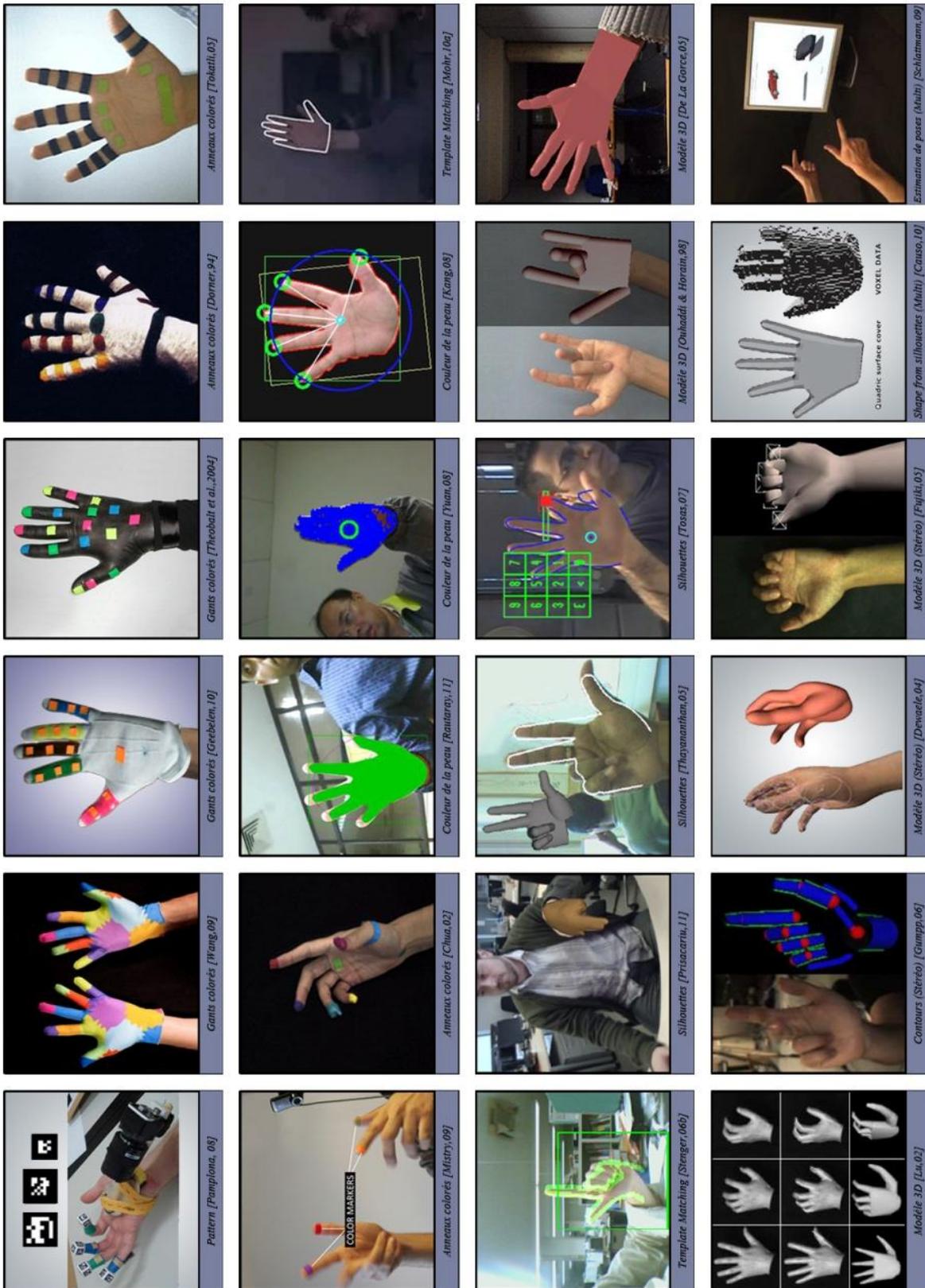
Enfin, il existe également la technique dite « d'estimation de pose ». Markus Schlattmann propose un système de manipulation de données 3D qui s'appuie sur trois caméras et n'utilise aucun capteur. Les 3 caméras se chargent de reconstruire un modèle grossier des mains de l'utilisateur (c'est un volume convexe réalisé à partir des zones d'intersection des 3 vues des caméras, comme celui obtenu par la méthode « *shape from silhouettes* »). Après cette reconstruction 3D des mains de l'utilisateur, le système détecte la pose des mains pour effectuer les interactions avec l'objet 3D (Guan & Turk, 2005; Schlattmann & Klein, 2009). Pour cela, les silhouettes sont comparées à une base de données et la position la plus proche est renvoyée. Si la pose est liée à une action précise, cette dernière est lancée (par exemple l'action de prendre un objet, ou « *picking* »). Ce système est peu coûteux, temps réel et permet une interaction bi-manuelle, mais l'utilisateur doit conserver une pose particulière de la main

pendant quelques instants afin que le système détecte quelle action il souhaite exécuter. De plus, cette technique impose à l'utilisateur de connaître les poses reconnues par le système, c'est-à-dire qu'il doit apprendre un langage particulier peu intuitif.

Le même principe a été repris par (Romero, 2007) avec deux caméras. Les poses sont estimées à partir de la position des extrémités des doigts et grâce à une technique proche de la cinématique inverse.

Des systèmes plus simples permettent, à partir de 2 caméras, de suivre les mouvements des mains en 3D. Sur chaque image provenant des deux caméras, une segmentation est effectuée afin de ne garder que les mains à l'image (Argyros & Lourakis, 2006). Ce traitement est fait de façon indépendante et sans mise en correspondance entre les deux images. Une fois que les différentes mains (s'il y en a plusieurs) sont identifiées, le centre de chacune d'elles est récupéré et est ensuite mis en correspondance entre les 2 caméras. Cela permet, par une méthode classique de triangulation, de connaître la position 3D des différentes mains de façon rapide et simple. Ce système autorise un suivi de la main selon 3ddl mais ne permet pas de connaître l'orientation de la main et encore moins des différents doigts. De plus, aucune posture ne peut être reconnue, seule la position de la main est étudiée.

Figure 15 : Systèmes de capture de mouvements des mains « logiciels »



Systèmes de capture de mouvements des mains « logiciels »

## 4 Discussion

### 4.1 Synthèse des avantages et désavantages des solutions présentées

Nous allons maintenant synthétiser les avantages et désavantages des solutions matérielles et logicielles présentées dans la partie précédente.

#### **Coût**

Dans le cas des systèmes matériels, le coût est souvent très élevé si l'on souhaite un suivi de qualité ; les exosquelettes, les systèmes magnétiques et les systèmes optiques sont les solutions qui donnent les meilleurs résultats mais qui sont également les plus chères. Les solutions basées sur les gants procurent une moins bonne précision mais sont financièrement plus accessibles. Du matériel grand public (comme les bras à retour d'effort, les souris 3D, etc.) a été développé, mais il ne permet malheureusement de suivre les mouvements dans l'espace que d'un seul point à la fois. De plus, l'espace de travail est très réduit pour les solutions à faible coût. Les systèmes logiciels, bien que moins précis, permettent quant à eux une plus grande zone de tracking pour un coût très faible (le prix d'une ou de deux caméras) ce qui rend ces solutions accessibles à tous.

#### **Confort d'utilisation**

Les solutions matérielles requièrent toutes le port de capteurs, marqueurs réfléchissants ou exosquelettes. Dans de très nombreux cas, l'usage de ces capteurs nécessite des connexions filaires, ce qui peut gêner l'utilisateur dans ses mouvements. Des solutions sans fil existent mais il faut dans ce cas utiliser des blocs qui se chargent d'envoyer les informations par ondes radio ce qui rend le système plus lourd à porter. Certains matériels comme les exosquelettes peuvent être volumineux, lourds et impressionnants, ce qui peut gêner l'utilisateur aussi bien physiquement que psychologiquement (certaines personnes sont réticentes face à des systèmes complexes à positionner sur la main, d'autant plus s'ils sont à retour d'effort). Les gants sont la plupart du temps à taille unique et donc peu adaptés à la morphologie des utilisateurs, ce qui peut également gêner le confort. De même, le port de gant peut faciliter la transpiration, ce qui induit des problèmes d'hygiène dans le cas d'utilisateurs multiples. Les solutions logicielles (sans marqueurs ou avec anneaux colorés) n'ont pas les problèmes du poids, de câbles ou encore d'hygiène, ce qui améliore d'autant le confort d'utilisation.

#### **Puissance de calcul**

Pour les solutions matérielles comme les exosquelettes, les données sont très simples à traiter, il suffit dans un premier temps de les nettoyer et ensuite de les utiliser. Les positions des doigts et les orientations sont obtenues directement. En revanche, pour les systèmes optiques ou magnétiques, il est nécessaire, en plus du nettoyage des données, de faire appel à la cinématique inverse pour en déduire les positions et les orientations des doigts. Pour les systèmes optiques, un algorithme supplémentaire doit être utilisé pour déduire par triangulation la position 3D des marqueurs à partir des différentes vues des caméras. De plus, il est possible d'utiliser des algorithmes pour prédire les positions en cas d'occlusions. En ce qui concerne les gants de données, les informations de flexion des doigts sont données directement, tout comme les périphériques de type bras à retour d'effort ou les souris 3D. Aucun temps de calcul supplémentaire n'est nécessaire, les informations peuvent être lues directement. En revanche, pour tous les systèmes logiciels, tous les traitements d'images, segmentation, analyse, suivis et reconnaissances de gestes sont faits de façon algorithmique, ce qui veut dire qu'une puissance de calcul conséquente est nécessaire pour parvenir à du temps réel. Cela peut nécessiter des ordinateurs très puissants voire des clusters de pc dans certains cas.

## Taille de la main de l'utilisateur

Dans la plupart des solutions matérielles, la taille de l'appareillage ou des gants est unique, elle ne peut s'adapter à la morphologie de chaque utilisateur. Cette taille est la taille moyenne qui a été définie sur un échantillon de plusieurs personnes, ce qui fait que dans la majorité des cas, les données sont relativement fiables, en revanche lorsque les personnes ont des mains très grandes ou très petites par rapport à la moyenne, les données peuvent être imprécises voire totalement erronées. Certains systèmes peuvent néanmoins s'adapter, par exemple le gant « *ShapeHand* » qui possède 3 tailles différentes ou encore certains exosquelettes qui permettent de régler la position des articulations afin de correspondre au mieux à l'utilisateur. Les systèmes optiques, s'ils fonctionnent avec des marqueurs et non des gants, sont adaptés à la morphologie de l'utilisateur car les marqueurs utilisés sont fixés où l'on veut sur la main. En revanche une calibration est nécessaire et le modèle virtuel de la main doit être adapté pour que la cinématique inverse soit correctement calculée. En ce qui concerne les solutions logicielles, la plupart du temps la taille de la main n'a pas énormément d'importance et ne pose pas de problèmes particuliers pour les résultats. En revanche, la couleur de la peau peut être un obstacle dans certains cas, alors que le problème ne se pose pas avec les systèmes matériels.

## Calibration

La majorité des solutions précédemment citées ont besoin d'une phase de calibrage, aussi bien pour les solutions matérielles que logicielles. Pour les systèmes matériels (exosquelettes et gants par exemple), la calibration est nécessaire pour connaître la position au repos et voir les minima et maxima possibles dans les amplitudes des mouvements. Pour les systèmes optiques, la calibration est nécessaire pour connaître les paramètres des lentilles des caméras, les orientations et positions de chacune d'elles ; pour cela, un « *rigid body* » est utilisé. En ce qui concerne les solutions logicielles, une phase de calibration est souvent nécessaire. Cela sert à la fois à enregistrer l'histogramme de la couleur de la peau, la taille et les caractéristiques de la main, mais aussi à obtenir les différents paramètres comme l'éclairage ambiant ou encore les données relatives aux caméras pour la distorsion ou la triangulation dans le cas des systèmes stéréoscopiques.

## Le bruit

De très nombreux systèmes sont sujets à des perturbations dues à l'environnement d'utilisation. Il faut très souvent, d'une part un environnement contraint et d'autre part nettoyer les données si l'on souhaite obtenir de bons résultats. Les exosquelettes, les gants et les périphériques de type « bras à retour d'efforts », souris 3D, etc. sont peu sujets aux bruits. En revanche, les systèmes magnétiques, acoustiques et optiques sont particulièrement sensibles respectivement au métal, aux fréquences parasites ou aux parois, et à la lumière infrarouge. Les solutions logicielles sont, quant à elles, toutes sujettes au problème de changement de luminosité (certaines arrivent cependant à en réduire les effets si les variations sont faibles). De plus, les couleurs des habits de l'utilisateur et du fond peuvent apporter des imprécisions et donc du bruit. Enfin, dans le cas des solutions utilisant des caméras, les problèmes de distorsion dus aux lentilles, même si ce n'est pas du bruit à proprement parler, déforment l'image et il faut éventuellement prendre en compte ce paramètre dans les algorithmes.

## Précision

Les exosquelettes offrent une excellente précision en position et en rotation, tout comme les systèmes magnétiques et les systèmes optiques (la résolution des caméras étant en générale assez grande) dans des conditions d'utilisation normale (i.e. il n'y a pas de perturbations particulières dans la zone de tracking). Les systèmes de type bras à retour d'effort et souris 3D sont également des interfaces très

précises. En ce qui concerne les gants, la précision est un peu moindre, bien que théoriquement du même ordre, les gants ont le désavantage d'être moins précis si la taille de la main de l'utilisateur n'est pas dans la moyenne. Toutes les solutions matérielles hormis les systèmes optiques donnent des informations codées en bits, plus le nombre de bits est grand et plus la précision théorique augmente. Pour les systèmes basés sur l'usage de caméras (aussi bien pour les solutions matérielles que logicielles), la précision dépend de la résolution des caméras et donc les informations sont données en pixels. Les solutions professionnelles utilisant des marqueurs réfléchissants atteignent des précisions importantes qui sont du même ordre que les exosquelettes ou les systèmes magnétiques. En revanche pour les solutions logicielles, les résolutions sont en règle générale beaucoup moins grandes et la précision du système l'est également. De plus, la précision dépend énormément des algorithmes utilisés pour le suivi des mouvements.

### **Occlusion**

Les problèmes d'occlusion ne sont présents que pour les systèmes basés sur l'usage de caméras. Tous les autres systèmes donnent des valeurs de positionnement quelque soit la position et l'orientation de la main. En revanche, puisque les solutions optiques reposent sur l'usage de caméras, si la main de l'utilisateur n'est pas en face ou mal orientée par rapport à celles-ci, il peut se produire des occlusions partielles ou totales ce qui conduit à une perte du suivi des mouvements. Des algorithmes peuvent essayer de prédire (dans une certaine mesure) les positions pendant l'occlusion, mais cela ne peut se faire qu'au prix d'un temps de calcul supplémentaire en utilisant des filtres de Kalman par exemple. L'occlusion est l'inconvénient majeur dans les solutions optiques. Pour tenter d'y remédier, il faut ajouter des caméras ce qui coûte cher et demande une puissance de calcul supplémentaire pour prendre en compte tous les flux vidéo et ensuite les trianguler. Ainsi, si une caméra ne voit plus la main de l'utilisateur, une autre caméra peut la voir et le suivi peut continuer.

Comme nous l'avons expliqué auparavant, toutes les solutions proposées sont temps réel ou sont légèrement en dessous du seuil : environ 25 à 30 rafraîchissements par seconde, voire beaucoup plus pour les systèmes haptiques où le rafraîchissement doit être très supérieur pour avoir un bon rendu. Les solutions non temps réel ne permettant pas d'interagir avec l'environnement virtuel de façon directe.

## **4.2 Faiblesse des solutions présentées**

Aucun des périphériques présentés ici n'est directement adapté à l'ensemble des applications possibles en réalité virtuelle. Le choix d'utiliser une technologie plutôt qu'une autre répond à un ensemble de critères et de contraintes liés aux tâches et à l'application à réaliser. Ces contraintes sont nombreuses : capture temps réel ou non, temps de réponse minimum, nature des mouvements à traquer, absence de liaison physique entre l'utilisateur et le système, volume de travail désiré, précision, résolution, fréquence d'échantillonnage, sensibilité à la température, à la pression, à l'humidité, à l'environnement (lumière, objets métalliques, objets réfléchissants, parois, etc.) ou encore le prix. En tout état de cause, aucun des systèmes présentés précédemment n'est parfait.

Les systèmes matériels sont ceux qui procurent une très grande fiabilité et un positionnement dans l'espace très précis. Cependant, ils nécessitent la plupart du temps des connexions filaires et surtout l'usage de matériel souvent contraignant pour l'utilisateur, que ce soit au niveau de l'encombrement, du poids, de la liberté de mouvements, de l'installation fastidieuse du matériel, etc. De plus, ils sont également sensibles à différents facteurs (lumière, bruit, métal,...) qui peuvent briser les résultats et les rendre totalement erronés. D'autre part, dans de très nombreux cas, la taille et la configuration du matériel sont uniques et ne peuvent pas être adaptées exactement à la morphologie de l'utilisateur. Sur

certains systèmes, il est tout de même possible d'avoir une adaptation propre à l'utilisateur, mais cela au prix de plusieurs réglages et ajustements relativement longs. Leur utilisation est restreinte et nécessite la présence d'un expert que ce soit pour l'installation, la calibration, ou le traitement des données. Aucun des périphériques matériels ne permet de faire un suivi et une capture des mouvements totalement transparents pour l'utilisateur : il est toujours nécessaire d'avoir recours à du matériel à porter (capteurs, marqueurs réfléchissants, exosquelettes, ...) ou à l'utilisation d'un objet à tenir en main (manette, souris 3D, trackball, ...). Enfin, le dernier critère de taille en défaveur de ces systèmes est leur prix. Les appareils de capture de mouvements très précis sont onéreux et ne sont donc pas à la portée du grand public, ce qui en limite la diffusion aux grandes industries ou aux laboratoires de recherche.

Les solutions logicielles permettent d'outrepasser la majorité de ces contraintes. L'absence (ou la très faible présence dans le cas de port d'anneaux colorés) de matériel rend le système transparent (ou quasiment) pour l'utilisateur ce qui rapproche l'interaction en environnement virtuel des environnements réels. De plus, leur prix est accessible, même aux particuliers, car ils ne nécessitent l'usage que d'une ou deux caméras dans la majorité des cas. Pour autant, les solutions logicielles sont également sujettes à différents inconvénients qui peuvent apparaître rédhibitoires. En premier lieu il est nécessaire d'avoir recours à des algorithmes de traitement d'images souvent complexes et gourmands en ressources. Comme tous les systèmes optiques, des occlusions partielles ou totales sont possibles et les changements de luminosité ou de fond sont des facteurs importants dont dépendent fortement les résultats. De même, de nombreux paramètres sont à prendre en compte, comme les lentilles des caméras, les distorsions possibles, les problèmes de parallaxe, la couleur de la peau ou des habits de l'utilisateur ; un environnement contrôlé est de fait très souvent nécessaire pour que le système fonctionne correctement. Les algorithmes de traitement d'images, s'ils ne sont pas optimisés ou trop complexes, peuvent rendre le système lent et donc peu ou pas adapté aux applications temps réel.

Contrairement aux applications 2D où la souris et le clavier ont fait leurs preuves depuis très longtemps, il n'existe pas de système de capture de mouvement équivalent en 3D qui fasse l'unanimité. Un système en particulier peut être très bien adapté à une tâche mais présenter un défaut d'utilité ou d'utilisabilité pour une autre et cela à cause de la diversité des actions que l'on peut réaliser en 3 dimensions.

Pour obtenir un périphérique qui puisse apporter le maximum d'avantages tout en minimisant les contraintes, il faut donc faire un compromis entre précision, transparence, contraintes environnementales, temps réel et adaptabilité.

### 4.3 Émergences et promesses des caméras 3D

Comme nous l'avons expliqué précédemment, le périphérique idéal devrait permettre d'avoir à la fois un suivi des mouvements des mains en temps réel, une bonne précision, être transparent pour l'utilisateur (et donc avoir un poids nul), peu sensible à l'environnement (qui peut changer) et aux perturbations dans celui-ci. De l'état de l'art réalisé, il apparaît que le système qui pourrait s'approcher le plus de cet idéal serait un système basé sur la vision par ordinateur. En effet, une solution basée sur cette technique permet d'enlever tous les principaux inconvénients des périphériques matériels comme : le poids (qui est nul puisqu'il n'y a rien à mettre sur les mains), les liaisons filaires qui gênent les mouvements, les perturbations de l'environnement (métal, son, lumière, etc.), le prix important tout en permettant de s'adapter relativement facilement à la morphologie de l'utilisateur sans risque d'obtenir des données erronées. Un périphérique basé sur la vision par ordinateur peut ainsi réaliser une capture des mouvements sans aucun capteurs ou marqueurs et donc autorise le périphérique à être

totallement transparent pour l'utilisateur. D'un autre côté, le système idéal devrait être en mesure de s'affranchir des inconvénients inhérents aux systèmes basés sur les caméras comme : les perturbations liées aux changements de fond ou de lumière, les problèmes d'occlusion, le nombre d'individu dans la zone de travail, la couleur de la peau ou des habits et surtout permettre une capture des mouvements en 3 dimensions et en temps réel.

Comme nous l'avons vu dans l'état de l'art présenté précédemment les solutions proposées ne permettent malheureusement pas de remplir toutes ses conditions. Les seuls systèmes permettant une capture en 3D précise sont ceux qui reposent sur l'utilisation d'au moins 2 caméras afin de faire, à partir de la stéréoscopie et d'une triangulation, une carte de profondeur. Le problème avec les systèmes stéréoscopiques est que la majorité de la puissance et du temps de calcul de l'ordinateur sert à recréer cette carte de profondeur, ce qui laisse peu de ressources et de temps pour le reste des traitements et de l'utilisation (dans le cas du temps réel). De plus, les images doivent absolument contenir des points d'intérêts remarquables pour pouvoir reconstruire par triangulation une carte de profondeur précise. Dans certains cas, un traitement peut être effectué au préalable afin d'éliminer, d'une part, les données inutiles et d'autre part, de faciliter le calcul de la stéréoscopie. Dans ce cas, les images ne contiennent plus que les mains de l'utilisateur et il peut être difficile de réaliser la triangulation faute de points d'intérêt et donc d'avoir des distances précises.

Une solution intéressante serait donc de faire appel à un système qui se charge de calculer la profondeur sans déporter le calcul sur l'ordinateur. De tels systèmes existent et sont appelés caméras à temps de vol, en anglais « *time-of-flight cameras* » (TOF), RGB-D cameras (*RGB-Depth*) ou encore Z-cam. Une évaluation de l'utilisation de ces caméras dans le domaine de l'informatique graphique a été effectuée par (Kolb et al., 2009). La technologie des caméras 3D est assez récente et la première caméra 3D fut créée par Lange (Lange, 2000; Lange & Seitz, 2000). Elle permettait d'obtenir une carte de profondeur de quelques centimètres de côté sur des distances allant jusqu'à 10 à 20 mètres avec un taux de rafraîchissement de 10 fps. La source d'illumination était composée de 160 LEDs modulée à 20Mhz avec une puissance optique totale d'environ 800mW. Ces caméras se sont beaucoup améliorées depuis les années 2000. Des illustrations de plusieurs caméras 3D présentées dans cette partie sont regroupées sur la Figure 23 page 61.

Les caméras 3D peuvent reposer sur différents types de capteurs qui dépendent du principe utilisé pour calculer les distances. Les 2 principes les plus communs utilisés pour les systèmes techniques sont la triangulation et le temps de vol. Ces 2 principes peuvent eux-mêmes être séparés en 2 sous-catégories qui sont respectivement : triangulation passive ou active et temps de vol pulsé ou à décalage de phase. Les caméras 3D fournissent en réalité une représentation en 2,5D et en temps réel car une partie seulement de la surface peut être vue par la caméra, cependant on les appelle caméras 3D par abus de langage.

#### **4.3.1 Triangulation**

Cette technique est appelée triangulation puisque l'objet dont on veut mesurer la distance forme un triangle avec 2 parties du capteur. Si le capteur est constitué d'une partie réceptrice et d'une partie émettrice active (une source de lumière par exemple), le principe de mesure est appelé triangulation active. Si le capteur consiste seulement en l'utilisation de 2 récepteurs passifs, le principe est appelé triangulation passive. Ce dernier correspond en d'autres termes à la stéréoscopie, déjà expliquée auparavant (cf. 3.3.2 p. 41).

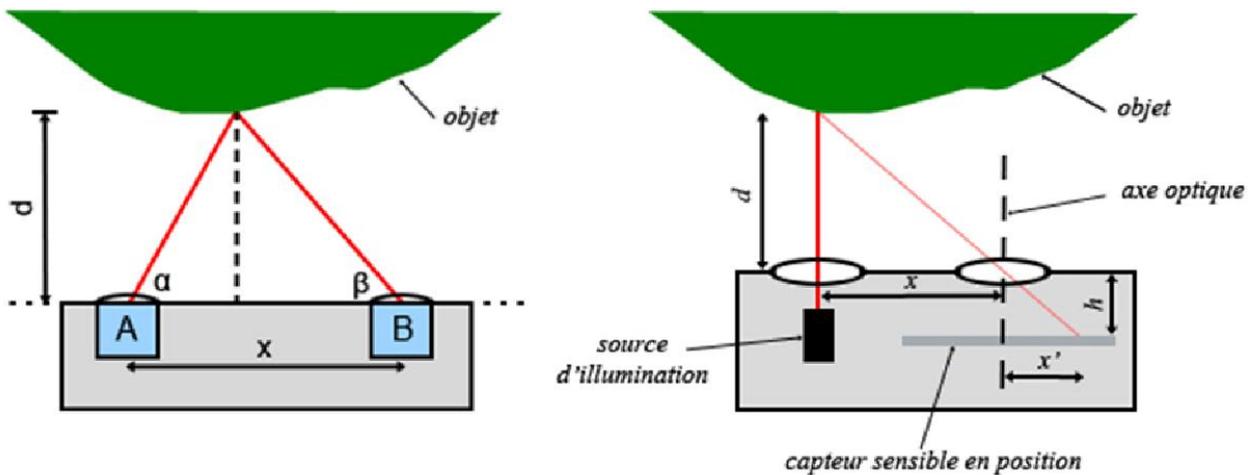


Figure 16 : Principe de fonctionnement d'une triangulation passive (à gauche) et d'une triangulation active (à droite)

#### 4.3.1.1 Triangulation passive

Elle repose sur le principe binoculaire de la vision humaine et se compose de 2 caméras qui observent une même zone de chevauchement (cf Figure 16 - gauche). Si un point  $p$  spécifique est dans le champ des deux récepteurs, il est possible de déterminer sa distance  $d$  par rapport au capteur. Par conséquent, chaque récepteur calcule l'angle entre la ligne formée par le centre du capteur et le point  $p$  et l'axe optique du récepteur. En combinant les angles avec la distance  $x$  entre les deux récepteurs, la distance au point  $p$  est calculée par :

Équation 3 : Calcul de distance par triangulation passive

$$d = \frac{x}{\frac{1}{\tan \alpha} + \frac{1}{\tan \beta}}$$

Où  $\alpha$  est l'angle que forme le récepteur A avec le point  $p$  et l'horizontale et  $\beta$  est l'angle que forme le récepteur B avec le point  $p$  et l'horizontale. Cette formule part du principe que les axes optiques des deux récepteurs sont parallèles. La partie la plus importante pour la triangulation passive est de trouver d'une part des points d'intérêts et d'autre part de trouver la correspondance dans les deux images à la fois pour pouvoir réaliser la triangulation. Chaque point de l'image du récepteur A doit être correctement identifié dans l'image du récepteur B. Si de mauvaises correspondances sont réalisées, cela conduit à des distances erronées (May et al., 2007). Comme nous l'avons expliqué précédemment, la stéréoscopie peut être réalisée en temps réel, cependant les cartes de distances obtenues ne sont pas très précises (peu de niveaux possibles) et la triangulation est très difficile si les images sont faiblement texturées ou comportent peu de points d'intérêts. De plus, la rectification des images, la recherche des points d'intérêt, la mise en correspondance entre ces deux images et le calcul de la triangulation est très gourmand en temps et en ressources.

#### 4.3.1.2 Triangulation active

La Figure 16 (droite) représente la configuration d'un capteur utilisé pour la triangulation active. Une source lumineuse projette un point sur l'objet et la réflexion du point lumineux est mesurée par la partie réceptrice du capteur. Ce récepteur est un capteur qui peut déterminer l'endroit exact où la lumière réfléchi l'a frappé. En connaissant la position de la lentille du capteur, la distance entre l'émetteur et le récepteur  $x$ , leur distance à l'optique  $h$  et le point d'impact de la lumière réfléchi  $x'$ , il est possible de calculer la distance  $d$  de l'objet par la formule (May et al., 2007) :

#### Équation 4 : Calcul de distance par triangulation active

$$d = h \cdot \frac{x}{x'}$$

Une configuration comme celle-ci avec un capteur aussi simple ne permet de mesurer la distance que d'un seul point. Pour déterminer la forme d'un objet, il est alors nécessaire de prendre plusieurs mesures et de faire tourner soit l'objet soit le dispositif autour de celui-ci. Des systèmes de triangulation plus sophistiqués utilisent des sources de lumière ainsi que des récepteurs, tous deux, à 2 dimensions. Ils projettent un motif lumineux sur l'objet qui est reçu par un capteur de caméra 2D par exemple. Un tel système fournit directement des informations de la forme 3D des objets mesurés. On appelle ça des systèmes à lumière structurée. Un système à lumière structurée peut être de n'importe quelle forme du moment que le motif projeté est connu. Dans de nombreux cas, des successions de lignes noires et blanches dans le spectre visible sont utilisées. La déformation des lignes permet de connaître la forme de l'objet et d'en déduire les distances pour chaque pixel. Cependant, même si les résultats sont très bons et que la forme peut être obtenue en temps réel (Zhang et al., 2006), la projection d'un motif sur la personne peut être gênant en particulier si la source de lumière est dirigée vers les yeux. L'inconvénient majeur de l'utilisation de ces lignes noires et blanches est qu'il est nécessaire d'avoir un projecteur puissant et la plupart du temps le système fonctionne dans le noir et non en plein jour. Pour remédier à ce problème, il est possible de projeter sur la personne un motif infrarouge (Guan et al., 2007).

Cette méthode est justement utilisée dans la caméra 3D développée par la société *PrimeSense* (Freedman et al., 2010) ou encore dans la *Kinect* de *Microsoft* sortie en novembre 2010 (qui utilise la technologie *PrimeSense*). Une matrice de points infrarouges (dont la répartition est connue) est projetée sur la scène et le capteur récupère une image dans laquelle tous les points apparaissent (grâce à un filtre qui ne laisse passer que l'infrarouge). Puisque les objets dans la scène déforment la matrice de référence et que cette matrice est connue par la *Kinect*, les décalages de position de chaque point permettent de déduire, par triangulation, la distance qui les sépare de la *Kinect*. Celle-ci permet d'obtenir des informations de distance à peu près aussi précises que les caméras 3D à temps de vol pulsé ou à décalage de phase. De plus la résolution du capteur de profondeur utilisé dans la *Kinect* (640x480) est très supérieure à celles de ces caméras et le taux de rafraîchissement est de 30 fps, ce qui rend ce périphérique très intéressant pour réaliser de la capture de mouvements des mains en temps réel.

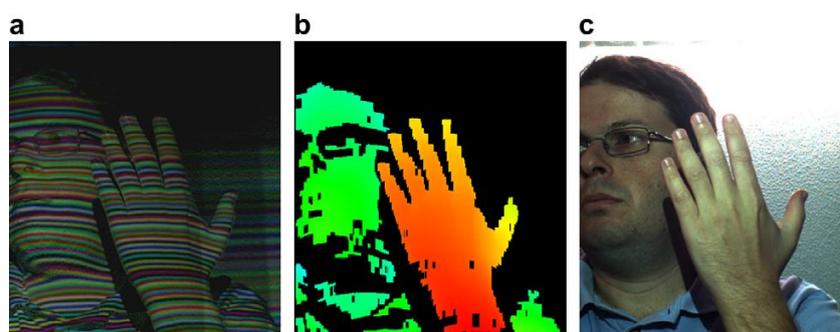


Figure 17 : Triangulation active à lumière structurée (visible)

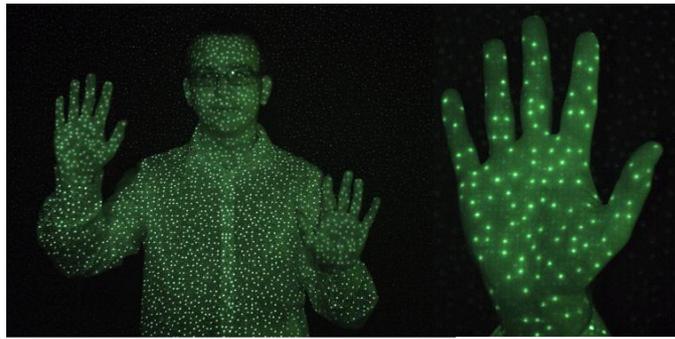


Figure 18 : Triangulation active à lumière structurée (invisible - infrarouge)

### 4.3.2 Temps de vol

Les caméras à temps de vol (TOF) sont des systèmes qui créent des cartes de distance à partir d'un seul objectif (contrairement au principe de la stéréoscopie). Le principe du temps de vol est le même que celui utilisé pour les scanners LIDAR (Cameron et al., 1991) ou encore celui utilisé par les chauves-souris avec l'ultrason, avec l'avantage de pouvoir capturer une scène complète en une seule fois.

#### 4.3.2.1 Temps de vol pulsé

La version la plus simple des caméras à temps de vol est celle qui utilise la lumière pulsée. La source d'illumination est allumée pendant un instant très court et la lumière émise par cette impulsion illumine la scène et est réfléchiée par les objets. L'objectif de la caméra se charge de recueillir la lumière réfléchiée et de l'imprimer sur le capteur. Selon la distance, la lumière entrante subit un retard plus ou moins long. Comme la vitesse de la lumière ( $c$ ) est de 300 000 km/s, ce retard est très court : un objet situé à une distance ( $D$ ) de 2,5m va retarder la lumière de 16,66 nanosecondes (Büttgen et al., 2005) :

Équation 5 : Calcul du temps de retard de la lumière dans une caméra à temps de vol pulsé

$$t_D = 2 \cdot \frac{D}{c} = 2 \cdot \frac{2.5 \text{ m}}{3 \cdot 10^8 \text{ m/s}} = 16.66 \cdot 10^{-9} \text{ s} = 16.66 \text{ ns}$$

La longueur de l'impulsion détermine la portée maximale que l'appareil peut gérer. Avec une longueur d'impulsion de 50ns par exemple, la portée est limitée à 7,5m :

Équation 6 : Calcul de la portée maximale d'une caméra à temps de vol pour une longueur d'impulsion de 50ns

$$D_{max} = \frac{1}{2} \cdot c \cdot t_0 = \frac{1}{2} \cdot 3 \cdot 10^8 \text{ m/s} \cdot 0.5 \cdot 10^{-7} \text{ s} = 7.5 \text{ m}$$

Chaque pixel est constitué d'un élément photo sensible (e.g. une photodiode) qui convertit la lumière réfléchiée en un courant. Dans les caméras à minuterie numérique, un compteur fonctionnant à plusieurs gigahertz est relié à chaque pixel photorécepteur. Il est activé lors de l'envoi de la pulsation et est arrêté quand la lumière frappe le photorécepteur. Dans le cas d'une minuterie analogique, le pixel utilise un double commutateur (G1 et G2) et 2 éléments de mémoire (S1 et S2). Les commutateurs sont commandés par une impulsion ayant la même durée que la lumière pulsée. Le commutateur G1 est activé en même temps que la pulsation de lumière et G2 est activé quand l'impulsion s'arrête. Le signal de commande du commutateur G2 est donc retardé d'exactement la durée d'impulsion. En fonction du délai occasionné par la distance, seule une partie de la lumière pulsée est prélevée en S1 au travers de G1 et l'autre partie est stockée en S2 lorsque G2 est activé. Selon la distance, le rapport entre S1 et S2 change comme le montre la Figure 19 (Haut) (Gokturk et al., 2004). Puisque seulement de petites quantités de lumière frappent le pixel en 50ns, non pas une seule mais plusieurs milliers d'impulsions

sont envoyées et recueillies, augmentant ainsi le ratio signal/bruit. Après l'exposition, les mesures des signaux S1 et S2 sont récupérées. Comme la longueur de l'impulsion lumineuse est définie, la distance peut être calculée à partir de la formule de Medina:

Équation 7 : Calcul de la distance avec la formule de Medina

$$D = \frac{1}{2} \cdot c \cdot t_0 \cdot \frac{S2}{S1 + S2}$$

Dans l'exemple précédent, les signaux sont S1=0.66 et S2=0.33, la distance est donc :

Équation 8 : Calcul de la distance avec S1=0,66 et S2=0,33

$$D = 7.5 \text{ m} \cdot \frac{0.33}{0.66 + 0.33} = 2.5 \text{ m}$$

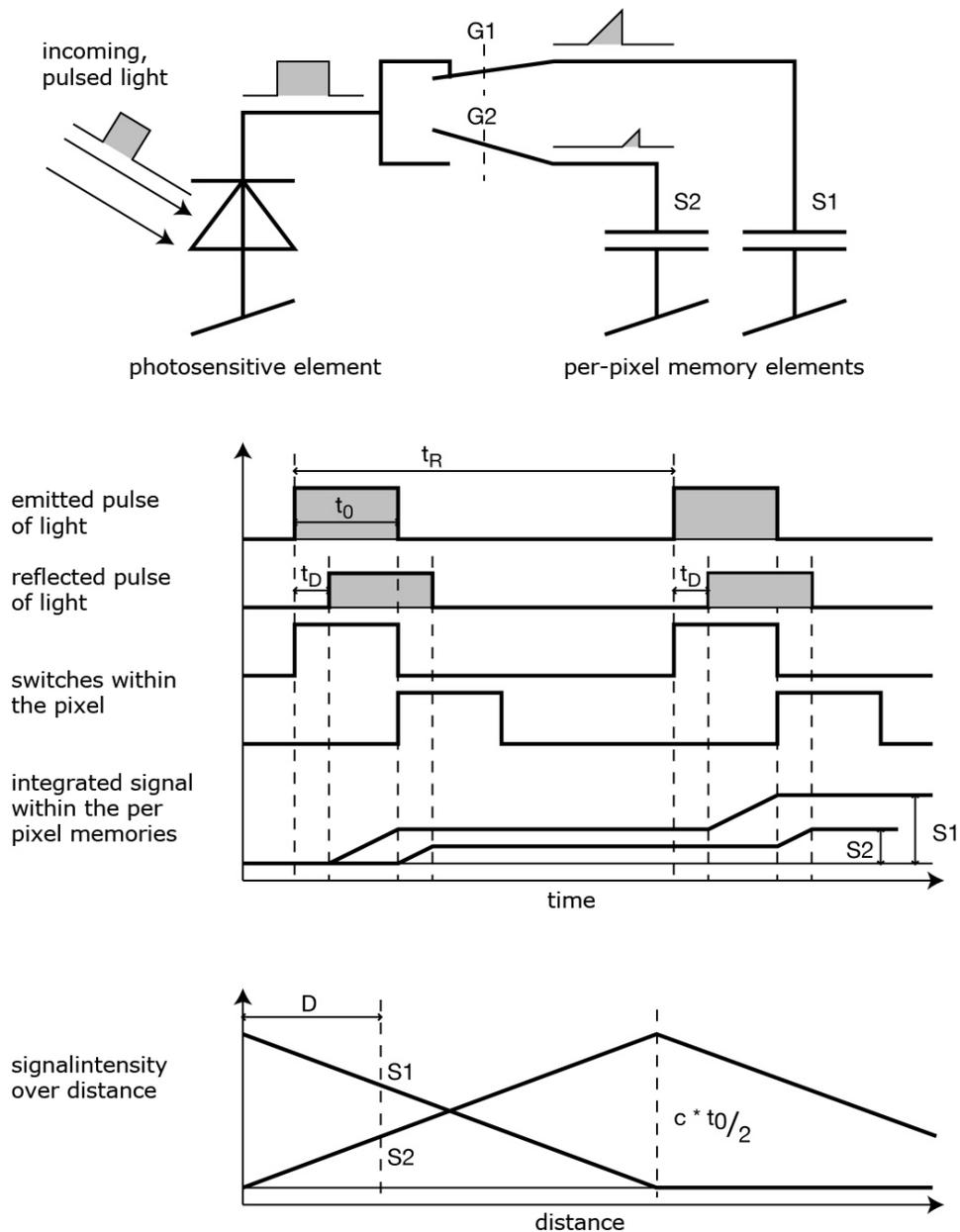


Figure 19 : Principe de fonctionnement d'une caméra à temps de vol pulsé

En cas de présence de lumière parasite, les éléments de mémoire reçoivent une partie supplémentaire du signal, ce qui perturbe la mesure de distance. Pour éliminer du signal la partie parasite, toutes les mesures peuvent être effectuées une deuxième fois sans éclairage de la caméra. De même, si les objets sont plus loin que la plage de distance possible, le résultat est également faux.

#### 4.3.2.2 Temps de vol à décalage de phase

Avec ce principe, les caméras possèdent un projecteur infrarouge qui illumine la scène et les objets avec une lumière dont la phase est connue. Les caméras mesurent ensuite le décalage de phase entre la lumière envoyée et celle reçue, autrement dit le temps que prend l'impulsion pour effectuer le trajet entre l'objet et la caméra. Il est ensuite possible de connaître la distance des objets à la caméra puisque le temps de vol est directement proportionnel à cette distance. Cette mesure de temps de vol est effectuée de façon indépendante sur chacun des pixels du capteur, permettant ainsi d'obtenir une image complète en 3 dimensions de la scène. Une caméra TOF est composée d'un système d'illumination constitué de LEDs ou de diodes lasers. Cette lumière est généralement infrarouge ou très proche pour ne pas interférer avec la lumière ambiante. Le second composant de ce type de caméra est un système optique qui permet de capter la lumière réfléchie par les objets et de la focaliser sur le capteur à temps de vol. Il est également très souvent constitué d'un filtre à bande passante dans l'infrarouge, permettant d'enlever la lumière ambiante qui peut bruyé les données. Le capteur TOF est bien sûr la partie principale de la caméra et permet de transformer la lumière incidente en signal électrique et ce pour chaque pixel qui effectue également le calcul de la distance. Le capteur est similaire à un capteur de caméra standard (2D), cependant puisque chaque pixel doit effectuer un calcul supplémentaire (réalisé par des composants électroniques placés juste à côté) par rapport à ces derniers, cette complexité nécessite une surface plus grande pour chaque pixel, ce qui diminue d'autant le nombre de pixels total pouvant être mis sur le capteur et donc sa résolution globale. L'électronique de ces caméras est une partie très importante du fait qu'elle doit être capable de synchroniser précisément la lumière émise avec le fonctionnement du capteur TOF. Si un décalage se produit entre les 2, les données de profondeurs sont complètement fausses. Pour donner une idée de la précision requise, une erreur de 100 picosecondes de déphasage représente une erreur de distance de 15mm environ, les GPUs actuels atteignent des fréquences de 3GHz (ce qui correspond à des cycles d'horloges de 300 picosecondes) ce qui permettrait une « résolution » de 45mm seulement. D'un autre côté, les fréquences de modulation utilisées pour moduler la lumière sont de l'ordre de plusieurs dizaines de MHz. Cela permet ainsi des précisions dans les distances de l'ordre du millimètre (CSEM, 2009). En fonctionnement normal, l'amplitude du signal reçu est réduite par un facteur  $k$  qui dépend de toutes les pertes optiques. La partie électronique se charge également de convertir les données générées par le capteur TOF en signal numérique et donc les distances en coordonnées sphériques ou cartésiennes. Cette conversion doit évidemment tenir compte des paramètres internes de la caméra comme par exemple les facteurs de distorsion de l'optique. Ces caméras permettent de couvrir des distances allant d'environ 40 cm à plus de 60 m pour certains modèles. Le paragraphe suivant explique plus en détail le fonctionnement technique d'une caméra TOF à décalage de phase.

Un signal lumineux modulé est divisé en deux signaux (un de référence et un mesuré) grâce à un miroir semi perméable aussi appelé « séparateur de faisceaux ». Le signal de référence est envoyé directement au récepteur interne et le signal mesuré est envoyé sur l'objet qui le réfléchit sur le miroir puis sur le récepteur (cf. Figure 20). Le décalage de phase entre ces deux signaux permet d'estimer le temps de vol et donc la distance (May et al., 2007). L'objet est éclairé par une lumière modulée en intensité à l'aide d'un signal sinusoïdal de fréquence  $f_{mod}$  qui voyage à la vitesse constante de la lumière dans le milieu

environnant. En estimant le décalage de phase  $\phi$  (en radian) entre le signal émis et le signal réfléchi, la distance  $d$  peut être calculée comme suit :

Équation 9 : Calcul de la distance pour une caméra à temps de vol à décalage de phase

$$d = \frac{c}{2f_{mod}} \cdot \frac{\phi}{2\pi}$$

Basé sur la périodicité du signal modulé de façon sinusoïdale, cette équation n'est valable que pour des distances inférieures à  $\frac{c}{2f}$ . Les caméras TOF actuellement disponibles fonctionnent par exemple à une fréquence de modulation d'environ 20 MHz. Ainsi, la limite supérieure pour les distances observables de ces caméras TOF est d'environ 7,5m, si la distance de l'objet est supérieure alors la valeur est donnée modulo 7,5. Une caméra 3D renverra par exemple la valeur 1,2m si l'objet se situe à 8,7m. D'autres distances sont également possibles en adaptant la fréquence de modulation. En plus des valeurs de profondeur, les caméras TOF fournissent des valeurs d'intensité qui représentent la quantité de lumière renvoyée par un point précis (Schaller, 2010). Cela permet de réaliser des images en niveaux de gris puisque les données de couleurs sont absentes dans la majorité de ces caméras.

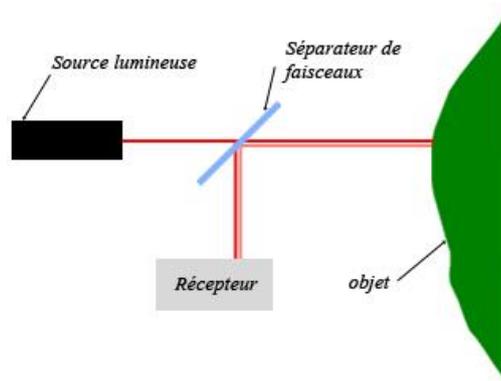


Figure 20 : Principe de fonctionnement d'une caméra temps de vol à décalage de phase (May et al., 2007).

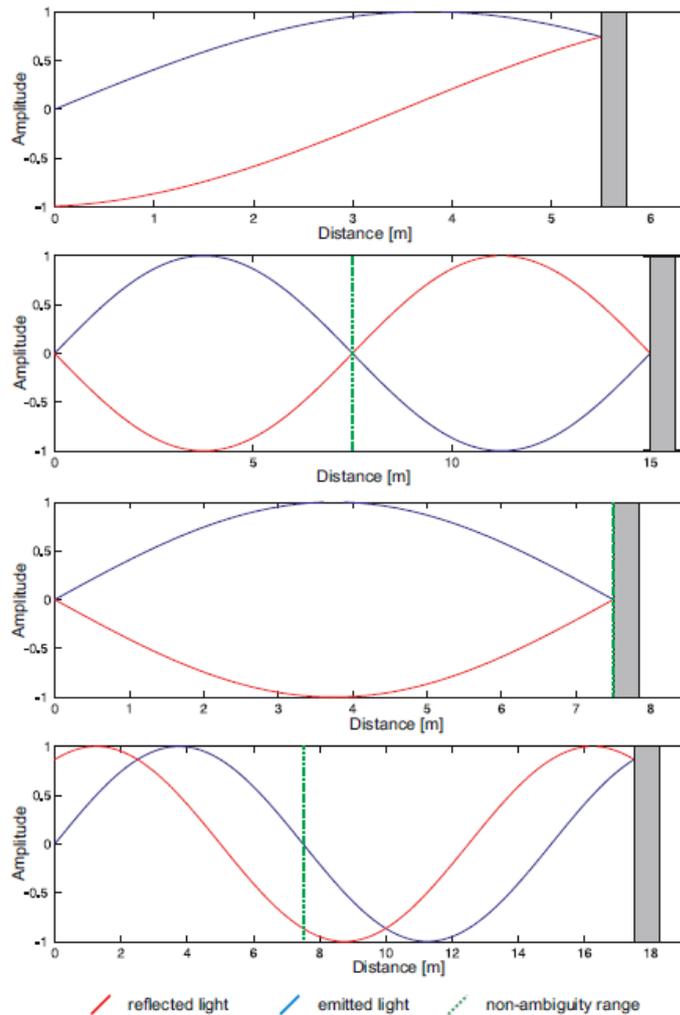


Figure 21 : Décalage de phase d'une onde lumineuse

En raison du principe de mesure, les caméras TOF ont une plage de distance  $d_n$  sans ambiguïté possible. Cela définit tout simplement une plage, où les distances peuvent être calculées de manière unique. La plage dépend de la fréquence de modulation de la caméra, car cette fréquence définit la longueur d'onde du signal émis. La Figure 22 montre la relation entre une distance  $d$  et un décalage de phase  $\phi_d$ . Par ailleurs, cette relation peut aussi être vue dans l'Équation 10, où  $c$ [m/s] correspond à la vitesse de la lumière,  $d_n$  [m] est la distance parcourue par la lumière,  $t$ [s] le temps nécessaire à la lumière pour parcourir la distance  $d_n$ ,  $f_{mod}$ [MHz] la fréquence de modulation,  $d_{nb}$  [m] la distance entre la source de la lumière et l'objet et  $\phi_d$  le déphasage.

Équation 10 : Détail du calcul de la distance pour une caméra à temps de vol à décalage de phase

$$c = \frac{d_n}{t}, \text{ où } t = \frac{1}{f_{mod}} \Rightarrow c = d_n \cdot f_{mod} \rightarrow d_n = \frac{c}{f_{mod}} \Rightarrow d_{nb} = \frac{c}{2f_{mod}}$$

$$d = \frac{c}{2f_{mod}} \cdot \frac{\phi_d}{2\pi}$$

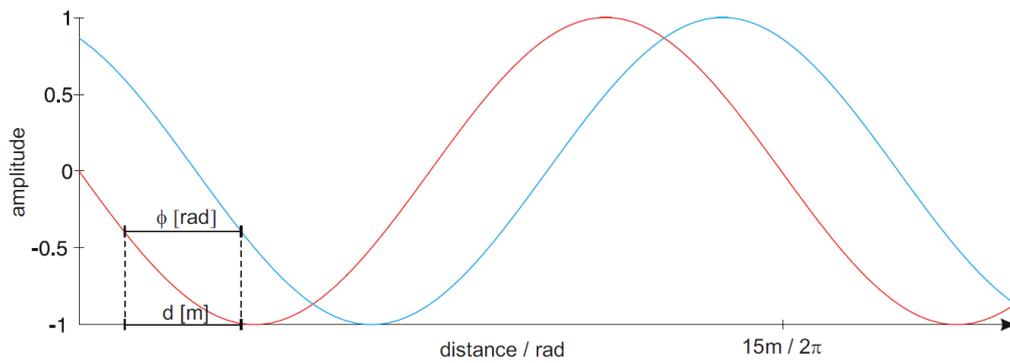


Figure 22 : relation entre une distance  $d$  et un décalage de phase  $\phi$

La Figure 22 montre la relation entre le déphasage et la distance pour une modulation de fréquence fixée  $f_{\text{mod}}=20\text{MHz}$ . La période (une ondulation) est, dans ce cas, de longueur  $2\pi$  ou  $15\text{m}$ . Étant donné un déphasage  $\phi_d$ , le déphasage est proportionnel à la distance  $d$  (Schaller, 2010).

### 4.3.3 Avantages

L'intérêt de ces caméras 3D pour la capture de mouvements des mains est que cette dernière est géolocalisée, en fonction de la position de la caméra 3D (la position de la caméra étant la référence  $(0, 0, 0$  en  $x, y, z$ ). Il n'y a donc pas besoin de capteurs supplémentaires pour connaître la position 3D de la main, comme l'exigent les gants de données qui nécessitent un capteur électromagnétique par exemple. Ce type de caméras possède d'autres avantages comme sa relative insensibilité aux perturbations magnétiques ou dues à la lumière ambiante (le système repose sur de l'infrarouge et permet donc de travailler de jour comme de nuit). De plus, grâce à la carte de profondeur, les caméras 3D permettent d'isoler la personne face à l'écran et de se concentrer éventuellement sur ses mains, en supprimant les personnes et les objets en arrière plan, en se basant uniquement sur les distances et non plus sur les couleurs. Puisqu'elles ne permettent pas de travailler sur les couleurs et que la majorité des algorithmes existants se basent sur ces informations, il est alors nécessaire de trouver de nouveaux algorithmes reposant sur les distances. Les problèmes tels que la couleur du fond, des habits ou encore de la peau ne se posent par conséquent plus avec ces caméras, cependant de nouveaux problèmes apparaissent, en particulier pour la segmentation des données dans les nuages de points obtenus.

D'un autre côté, ces caméras restent très sensibles aux surfaces réfléchissantes et à la lumière directe du soleil dans l'objectif. De plus, l'inconvénient principal de ces caméras est leur faible résolution dans la majorité des cas :  $160 \times 120$  pour la *D-Imager*,  $176 \times 144$  pixels pour la *SwissRanger*,  $204 \times 204$  pixels pour une *PMD* (cf : Figure 23). Si elles ne permettent pas d'obtenir une image couleur, il est cependant possible d'obtenir une image en niveaux de gris grâce à l'intensité lumineuse que chaque pixel reçoit.

Par ailleurs, le calcul des distances est réalisé dans la caméra elle-même et ne demande donc aucune ressource à l'ordinateur qui peut alors utiliser l'ensemble des ressources CPU et/ou GPU pour l'analyse et le suivi des mains. De même, ces caméras autorisent un taux de rafraîchissement de 25 fps (*PMD*) à 60 fps (*SwissRanger*), offrant ainsi la possibilité d'interagir en temps réel. Pour cela, un traitement important du nuage de points provenant du flux de données de la caméra est nécessaire. Si de nombreux algorithmes existent pour l'analyse d'images 2D en couleurs, il en existe beaucoup moins qui soient dédiés au traitement de nuages de points tridimensionnels. Si ce système n'impose aucun marqueur, il reste enfin sensible aux problèmes d'occlusions éventuels, inhérents à tout système basé sur l'utilisation d'une caméra. Par exemple, si l'utilisateur expose le dos de sa main à la caméra, il ne sera pas possible de connaître la position exacte des doigts si ces derniers sont repliés.

Ces caméras étaient peu utilisées avant 2011 pour réaliser des suivis et captures de mouvements en 3D. La compagnie *3DV* avait l'intention d'utiliser leur caméra 3D TOF (la *ZCam*) pour réaliser de la capture de mouvements dans le contexte des jeux vidéo. Cette dernière permettait d'avoir une image couleur tout en y ajoutant les informations de profondeur. La *ZCam* originale sortie en 2000 était un dispositif à ajouter aux caméras couleurs normales afin d'apporter uniquement les informations de profondeur supplémentaires. Elle a depuis été améliorée et avant que la compagnie soit rachetée en mars 2009 par *Microsoft*, elle avait planifié de sortir une caméra tout en un avec pour cible des applications de jeux vidéo. Cette caméra autorisait un taux de rafraîchissement de 60 fps. La caméra vidéo avait une résolution de 1,3 Mpixels et le capteur de profondeur, une résolution de 320x240 pixels. Les informations étaient codées sur 8bits (256 valeurs possibles de profondeur) et la plage de distance s'étendait de 0,5m à 2,5m pour une précision de l'ordre de 1 à 2 cm. Microsoft a également acheté en octobre 2009 la compagnie *Canesta* qui fabriquait ses propres caméras 3D et qui souhaitait aussi les utiliser dans le domaine des jeux vidéo.

L'arrivée de la Kinect sur le marché a suscité un énorme engouement pour cette technologie. Le périphérique de Microsoft a de nombreux avantages. En effet, la technologie utilisée est relativement simple mais permet d'obtenir des résultats assez précis pour faire de l'interaction en environnement virtuel (de l'ordre du centimètre). De plus, ce périphérique possède 2 capteurs, un pour réaliser la triangulation et obtenir la carte de profondeurs, et l'autre pour obtenir une image couleur. La résolution des capteurs est bien supérieure aux autres caméras 3D (640x480 pixels) et le taux de rafraîchissement est de 30 fps. Ce périphérique possède en outre, de nombreux capteurs supplémentaires comme des inclinomètres, des accéléromètres, des micros, etc.

High degree of freedom



Caméras 3D



■ Caméra à temps de vol   
 ■ Caméra à triangulation   
 ■ Possédant une caméra couleur

Figure 23 : Caméras 3D

#### 4.3.4 Comparatif des principales caméras 3D du marché

Grâce aux caméras TOF, le système peut apporter les avantages supplémentaires suivants par rapport aux classiques caméras couleurs : le fond n'a plus besoin d'être fixe et connu, le port de gants ou de marqueurs est inutile (puisque tous les algorithmes doivent se baser sur les distances et non plus sur les couleurs), la lumière ambiante peut varier (dans une certaine mesure) puisque la caméra fonctionne grâce à son propre projecteur infrarouge, les mains de (ou des) utilisateur(s) peuvent être suivies et ce quelque soient leur couleur, leur taille ou leur morphologie. Il n'y a plus besoin de faire attention à avoir des habits de couleur différente du fond et de la couleur de la peau. En revanche, l'inconvénient principal qui reste est l'occlusion possible d'une partie ou de la totalité des mains.

Avant 2011, quelques travaux ont porté sur l'utilisation de ces caméras 3D pour faire de la capture et de la reconnaissance de mouvements des mains (Breuer et al., 2007; Lahamy & Lichti, 2010; Li & Jarvis, 2009; Malassiotis & Srinivasan, 2008; Mo & Neumann, 2006; Suryanarayan et al., 2010) ou encore du corps entier (Holte & Moeslund, 2007; Kollarz et al., 2008). Ces travaux portent surtout sur la reconnaissance des postures de la main et ne permettent pas d'interagir avec l'environnement virtuel en temps réel. Hormis quelques laboratoires qui, au fil des années, se sont intéressés à ces périphériques, la majorité des techniques sans marqueurs portaient sur l'usage d'une ou de plusieurs caméras. Le prix très élevé des caméras 3D, du fait de leur faible commercialisation, ainsi que la faible résolution qu'elles offraient, constituaient très certainement les freins principaux. Une caméra 3D *SwissRanger* coûte environ 7000€ pour une résolution de 176x144 pixels. De même, puisque le nombre d'algorithmes permettant de traiter des nuages de points en temps réel est très faible comparé à ceux traitant les informations de couleur, peu de laboratoires se sont penchés sur la capture des mouvements de la main en temps réel à partir de caméras 3D. L'arrivée de la *Kinect* en novembre 2010 pour un prix de 150€ a permis de populariser largement ce type de périphérique et a, depuis, suscité un réel intérêt pour cette technologie.

## 5 Objectifs, Motivations et Hypothèses

L'objectif du travail présenté ici est la conception et l'évaluation d'un système permettant une interaction 3D transparente et temps réel en environnement virtuel, pour des tâches de sélection, de manipulation et de navigation. Le système est ici à comprendre comme l'association d'une technologie, de traitements algorithmiques et de modalités d'interactions. Nous souhaitons en évaluer de manière détaillée les performances d'une part, et la mesure de l'acceptabilité et des préférences des participants d'autre part. La finalité de notre travail est la mise à disposition d'un système efficace, efficient, confortable, performant et améliorant le sentiment d'immersion. Notre objectif général, et la finalité de notre travail, est motivée par la revue des matériels et techniques existantes, ainsi que par la quasi absence d'études démontrant l'intérêt de ce type de matériel pour l'interaction 3D temps réel. Nous allons dans un premier temps rappeler brièvement les raisons qui ont motivé le choix de la technologie de caméra 3D, puis nous présenterons le modèle d'interaction qui en découle. Nous présenterons ensuite nos hypothèses générales, définirons les critères d'évaluation et finirons cette partie par la structuration de notre travail.

### 5.1 Le choix d'une technologie

Une technologie ayant pour objectif de faire « disparaître » l'interface afin de permettre un mode d'interaction « naturel » aux utilisateurs est souvent présentée comme l'aboutissement final des interfaces de manipulation directe. (Fuchs & Moreau, 2003) parlent à ce propos d'interfaces

comportementales, ce qui implique que les recherches doivent se tourner vers la conception de systèmes « visant à exploiter un comportement humain, naturel et sans acquis préalable ». De même, comme le suggèrent Fuchs et Mathieu, dans le volume 2 du *Traité de la Réalité Virtuelle* (Fuchs & Mathieu, 2003), de manière idéale, les réponses motrices (dans le cadre de la capture de mouvements) doivent être transmises sans support matériel entre l'homme et la machine. Enfin, d'après (Winkler et al., 2007), le fait de n'avoir aucun matériel à porter et de rendre le système transparent pour l'utilisateur est de nature à améliorer le sentiment d'immersion et de présence ce qui est primordial dans le domaine de la réalité virtuelle.

Tout système possède naturellement des inconvénients mais les caméras 3D permettent de s'affranchir de la majorité des contraintes que l'on peut rencontrer avec les autres périphériques (tels que la couleur, la luminosité ambiante, les perturbations liées au son ou au métal, l'utilisation de capteurs, etc.). L'absence de périphérique en contact direct avec l'utilisateur est à même d'apporter une plus-value substantielle par rapport aux périphériques actuels. Ce sont ces raisons qui ont motivé notre choix pour cette technologie. Nous nous sommes orientés vers une caméra *SwissRanger* dans un premier temps, car elle était celle qui avait le meilleur rapport taux de rafraîchissement / résolution au moment du développement (la *Microsoft Kinect* n'était pas encore disponible à la vente et faisait alors l'objet de beaucoup de spéculations) et la précision était quasiment du même ordre entre les différentes caméras 3D. Notre système a ensuite été amélioré par l'utilisation de la *Kinect*, qui a été commercialisée durant la dernière année de notre travail de thèse. Nous avons décidé d'adopter ce nouveau périphérique car il offre une résolution supérieure tout en conservant un taux de rafraîchissement compatible avec la réalisation d'une interaction temps réel.

## 5.2 Notre vision du modèle d'interaction de Norman

Le choix de concevoir un système transparent, et donc de nous baser sur une caméra 3D nous conduit à proposer notre modèle de l'interaction homme-machine inspiré, en partie, du modèle de (Norman, 1988).

Le modèle de Norman (Figure 24) suggère que l'utilisateur envoie des stimuli à un périphérique quelconque (souris, clavier, space mouse, etc...) qui convertit ces informations et les envoie à l'ordinateur ; celui-ci retourne des données visuelles à l'utilisateur. Ce modèle a été enrichi (flèches grises) par (Nedel et al., 2003) par l'ajout des retours entre l'ordinateur et le périphérique, puis entre ce dernier et l'utilisateur lors de l'utilisation de périphériques haptiques.

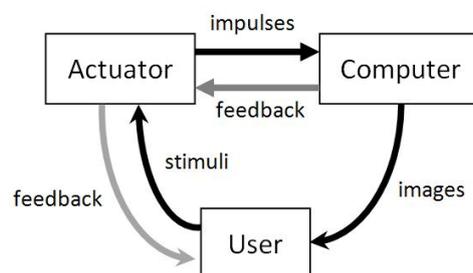


Figure 24 : Modèle de l'interaction homme-machine, tiré de (Norman, 1988)

Dans le système que nous proposons nous n'avons aucun feedback entre l'utilisateur et le périphérique, ni entre l'ordinateur et le périphérique, car le système est totalement transparent pour l'utilisateur. Il n'y a aucun capteurs ni matériel à porter et de ce fait aucun retour possible au niveau du périphérique.

La Figure 25 présente le fonctionnement de notre système, en lien avec le modèle de Norman présenté précédemment.

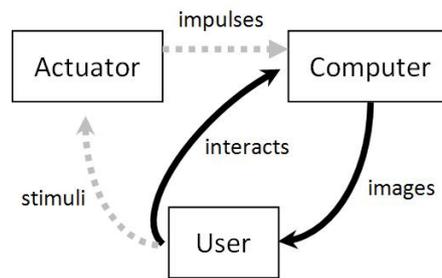


Figure 25 : Notre modèle de l'interaction homme-machine

Le cycle représenté en noir sur le schéma correspond à ce que l'utilisateur perçoit. Quand l'utilisateur bouge sa main, son geste lui est visuellement retransmis par l'ordinateur : ce retour visuel, associé à l'absence de port de matériel rend totalement transparent le périphérique. Il n'en reste pas moins qu'il est présent. La boucle représentée en gris sur notre schéma correspond à ce qui se passe réellement. L'utilisateur bouge son bras dans l'espace, le périphérique (une caméra 3D) se charge d'envoyer les informations à l'ordinateur qui peut ainsi déterminer la position et l'orientation de la main de l'utilisateur, ainsi que les actions qu'il est en train de réaliser. Enfin, un retour visuel est restitué à l'utilisateur. Il est donc très important, pour l'utilisateur, qu'il n'y ait pas de décalage temporel entre l'action réelle qu'il effectue et la restitution visuelle donnée par l'ordinateur. Il faut donc, d'une part que les traitements se fassent en temps réel et d'autre part avec une latence la plus faible possible.

Dans le cas d'une interaction en environnement virtuel, il existe plusieurs tâches (Coquillart et al., 2003) dont les principales sont :

- La sélection, c'est-à-dire positionner sa main (ou plus exactement l'avatar représentant sa main) sur un objet, pour pouvoir par la suite interagir avec lui. Le mouvement réciproque, naturellement, est la désélection ;
- La manipulation, c'est-à-dire la modification de l'état d'un objet supposé ici indéformable. Il s'agit de modifier sa position, son orientation et sa taille ;
- La navigation, c'est-à-dire la possibilité de se déplacer dans l'environnement virtuel et de l'explorer.

Nous avons donc choisi d'implémenter ces 3 tâches au sein de notre système.

### 5.3 Hypothèses générales

Les études expérimentales présentées dans les chapitres suivants visent à évaluer le système que nous proposons. Celui-ci a pour objectif d'obtenir des performances a minima équivalentes aux systèmes couramment utilisés pour les tâches de sélection et de manipulation d'objets 3D ainsi que pour les tâches de navigation. Nous ne visons pas de performances meilleures car nous savons qu'il est très difficile, avec un nouveau système, de dépasser les performances des systèmes avec lesquelles les utilisateurs ont une très grande pratique. L'influence de l'apprentissage étant particulièrement importante dans le cas de l'usage de la souris pour les tâches de sélection. Notre système vise par contre une meilleure acceptabilité, compte tenu des avantages que nous avons mis en avant précédemment. L'évaluation d'un gain se fait comparativement à des systèmes couramment utilisés : la

souris, pour la tâche de sélection d'un objet et les gants de données pour la manipulation et la navigation en environnement virtuel. Nos hypothèses de recherche sont donc les suivantes :

- Hypothèse A : les performances obtenues avec notre système, en termes de temps d'exécution, sont sensiblement équivalentes à celles obtenues avec la souris pour une tâche simple de sélection en environnement virtuel. Ces performances sont indépendantes de la taille de la main des participants, de leur sexe et de leur expertise en réalité virtuelle.
- Hypothèse B : l'acceptabilité est meilleure avec notre système, en termes d'utilisabilité, d'utilité et de sentiments d'immersion, pour une tâche de sélection, en comparaison de la souris.
- Hypothèse C : la précision et le temps de réalisation des tâches de manipulation et de navigation avec notre système sont sensiblement équivalents à ceux obtenus avec des gants de données associés à des capteurs électromagnétiques.
- Hypothèse D : notre système est à même de recueillir une meilleure acceptabilité et de meilleurs jugements subjectifs par les participants, pour des tâches de manipulation et de navigation en environnement virtuel, comparativement à des gants de données associés à des capteurs électromagnétiques. En particulier, du fait de l'absence de port de matériel et d'entrave due aux fils, nous pensons que notre système apporte un meilleur confort, une meilleure précision et une meilleure liberté de mouvements.
- Hypothèse E : les modalités d'interaction que nous proposons sont adaptées à notre système et permettent de réaliser de manière simple, confortable, efficace et efficiente les tâches de sélection, manipulation et navigation.

#### 5.4 Critères d'évaluation

Pour vérifier ces hypothèses, nous avons étudié, d'une part, les performances obtenues et, d'autre part, l'acceptabilité des utilisateurs envers le système que nous avons créé, lors de la réalisation de tâches de sélection, de manipulation et de navigation.

Les performances sont évaluées sous l'angle du temps nécessité par l'exécution d'une tâche, et par la précision dans le cas plus particulier des tâches de déplacement, d'orientation, de redimensionnement et de navigation. Cette précision est évaluée comme une fidélité entre la position, l'orientation et la taille souhaitées d'un objet et ces paramètres effectifs.

D'après (Tricot et al., 2003) « *l'acceptabilité [en anglais : acceptability] désigne la valeur positive ou négative de l'image que se fait un individu (ou un groupe) à propos de l'outil ou du système. En d'autres termes, l'acceptabilité se mesure au niveau des attitudes et opinions des utilisateurs potentiels vis-à-vis du produit lui même, de son utilité et de son utilisabilité.* » Nous avons donc défini l'acceptabilité selon trois sous-catégories : le sentiment d'immersion, l'utilité et l'utilisabilité en comparant à chaque fois notre système à des périphériques considérés comme les plus adaptés aux tâches à effectuer. D'autre part, nous avons décomposé l'utilisabilité en termes d'efficacité, d'efficience et de satisfaction. Ce découpage repose sur les notions d'ergonomie et les définitions tirées de Burkhardt. D'après (Burkhardt & Sperandio, 2004) « *les critères généraux de l'ergonomie concernent l'utilité, l'utilisabilité, l'accessibilité, la dangerosité, la satisfaction et l'acceptabilité. Ces critères délimitent de grandes familles d'objectifs ergonomiques pour la conception et l'évaluation* ». Parmi ces critères certains sont plus importants que d'autres et peuvent être hiérarchisés de façon variable en termes de priorité suivant le projet. Les deux critères principaux d'utilité et d'utilisabilité ressortent cependant plus que les autres. Nous adoptons ici, les définitions d'utilité et d'utilisabilité selon les termes employés par (Burkhardt, 2006).

« En ergonomie, le critère d'utilité [anglais : usefulness] correspond à un avantage ou un agrément significatif pour l'utilisateur dans une activité précise (en termes d'efficacité, de coût, de rapidité, de précision, d'agrément...). Cet avantage est toujours relatif : relatif aux objectifs de l'utilisateur, relatif aux outils existants ou habituellement utilisés, relatif à l'environnement d'utilisation, et relatif aux dépendances avec les autres activités. » (Burkhardt, 2006)

« L'utilisabilité [en anglais : usability] peut se définir comme la mesure (ou une estimation probable) de la performance des utilisateurs, dans le contexte et pour l'utilisation prévue. L'utilisabilité se mesure par divers indicateurs tels que la facilité d'apprentissage, la facilité de mémorisation, l'utilisation sans erreurs, etc. [...] Il en découle que l'utilisabilité n'est pas la propriété absolue d'un dispositif ou de telle technique d'interaction : elle est toujours relative à une tâche donnée, un contexte de réalisation et une population ciblée d'utilisateurs. » (Burkhardt, 2006)

Les critères d'accessibilité et de dangerosité ne sont pas évalués dans nos études. Toujours, d'après Burkhardt, « l'accessibilité concerne l'aménagement de l'accès aux systèmes d'information pour les personnes handicapées » et « la dangerosité regroupe l'ensemble des conséquences négatives possibles associées à l'utilisation du système notamment pour l'utilisateur, sa santé (santé, accident) et son environnement. [...] Au-delà des risques liés à un domaine d'application particulier, la dangerosité des environnements virtuels dépend notamment du type de configuration d'interfaçage exploitée : gants, visiocasque, exosquelette, etc. » (Burkhardt, 2006)

Enfin d'après (Green & Jordan, 2002), « le critère de satisfaction ou de plaisir [en anglais : satisfaction, enjoyment] [...] désigne l'expérience globalement positive qui marque de façon durable les utilisateurs du produit ou de l'outil conçu. Cette expérience s'analyse à travers plusieurs dimensions : satisfaction sur le plan de la pensée, des sensations et des émotions, etc. L'évaluation de ce critère s'appuie sur des échelles d'attitudes ou des questionnaires. »

## 5.5 Structuration du travail

Les études présentées dans les parties suivantes sont au nombre de trois, deux études principales et une étude complémentaire. La première étude est basée sur l'évaluation comparative de notre système et d'une souris pour une tâche simple de sélection 3D en environnement virtuel. Elle s'appuie sur la caméra 3D *SwissRanger SR 4000*. L'objectif premier de cette étude est d'attester l'intérêt de notre système et d'avoir un premier retour des participants sur le système proprement dit ainsi que sur la modalité d'interaction. Cette étude permettra de vérifier les hypothèses A et B.

La seconde étude reprend de nombreux éléments du protocole expérimental de l'étude précédente, et s'appuie sur une version améliorée du système initial, grâce à l'emploi de la *Kinect* de *Microsoft* et l'intégration de nouvelles bibliothèques permettant un temps de calcul réduit tout en offrant une plus grande robustesse. Cette étude permettra de valider les hypothèses C et D et l'hypothèse E dans une moindre mesure.

La troisième étude est complémentaire aux deux premières. Elle vise à prendre en compte les commentaires des participants concernant l'implémentation des modalités d'interaction. Elle opposera donc les modalités d'interaction utilisées pour la manipulation dans l'étude 2 (déplacement, rotation et redimensionnement) à celles implémentées d'après les suggestions et commentaires des participants. Elle apportera ainsi une seconde réponse à l'hypothèse E.

## Chapitre 3

# Souris vs. Cam3D pour la sélection en environnement virtuel : étude 1

---



# 1 Méthodologie

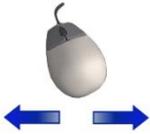
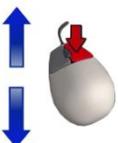
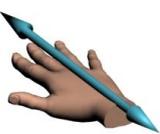
## 1.1 Présentation générale

L'application servant de base à l'expérimentation doit permettre une tâche de sélection d'objets dans un environnement 3D. Nous avons choisi de concevoir un environnement virtuel simple, en nous fixant la contrainte de devoir réaliser des déplacements dans l'environnement virtuel à l'échelle 1, afin de mettre en évidence l'intérêt de la capture de mouvement de la main sans marqueurs, et d'obtenir des résultats qui soient transposables à des environnements immersifs de type « CAVE ». L'application repose donc sur la reconnaissance de la position de la main (coordonnées x, y, z), ainsi que sur la reconnaissance des états « main ouverte » et « main fermée » pour correspondre au geste réel de sélection.

Puisque notre expérimentation vise à comparer l'interaction manuelle sans marqueurs à un périphérique d'interaction courant, ici la souris, un soin particulier a été apporté à rendre équivalents les 2 systèmes en termes de fonctionnalités et de modalités d'interaction. Dans les deux cas l'utilisateur déplace un avatar de main dans l'espace en 3D. Ce déplacement est soit indexé sur le déplacement de la main du participant (la latence est ici négligeable), soit sur les déplacements de la souris. Pour le périphérique Cam3D, nous respectons l'échelle 1 : 1 afin de respecter au mieux les conditions réelles d'utilisation, c'est-à-dire que l'utilisateur doit déplacer sa main de 1 cm pour voir l'avatar de la main se déplacer de 1 cm dans l'EV. En revanche pour la souris, nous avons laissé les paramètres originaux afin de conserver le paradigme classique d'interaction pour lequel un déplacement de la souris induit un déplacement plus important du curseur à l'écran. Le champ de déplacement de la souris se restreint par conséquent à quelques centimètres.

Là où l'interaction avec la main permet les déplacements selon les 3 dimensions, il a été nécessaire d'adapter l'interaction avec la souris. Nous avons choisi la transposition la plus simple de l'interaction courante : déplacement dans le plan de l'écran avec le déplacement normal de la souris, et déplacement en profondeur par un déplacement « avant / arrière » de la souris avec le maintien du bouton droit enfoncé. Cela pour rappeler le fonctionnement du zoom dans certains logiciels. Cette technique a été préférée à l'utilisation de la molette de la souris du fait de la faiblesse de ce système en termes d'ergonomie, de la facilité d'utilisation et du temps requis pour réaliser la tâche. L'interaction avec chaque périphérique est synthétisée dans le Tableau 4.

Tableau 4 : Déplacements et actions de sélection avec la « Cam3D » et le souris

Déplacement / Action	Souris	Cam3D
Haut ↔ Bas		
Gauche ↔ Droite		
Avant ↔ Arrière		
Sélection		

## 1.2 Terminologie

Voici les différents acronymes et termes que nous allons utiliser dans la suite de ce chapitre :

- Environnement virtuel : environnement 3D dans lequel les actions sont effectuées (ici un espace cubique) ;
- Système Cam3D : système utilisant une caméra 3D SwissRanger SR4000 comme interface transparente pour récupérer les mouvements effectués par l'utilisateur ;
- Système Souris : système utilisant une souris classique pour récupérer les mouvements effectués par l'utilisateur ;
- G1 : groupe ayant commencé l'expérimentation en effectuant la tâche avec la souris ;
- G2 : groupe ayant commencé l'expérimentation en effectuant la tâche avec la cam3D ;
- Tâche T : Sélection de 10 cubes par ordre croissant.

Dans la suite du de ce chapitre, afin de simplifier la lecture, nous remplacerons environnement virtuel par EV, système Cam3D par Cam3D, système Souris par Souris.

## 1.3 Hypothèses opérationnelles

Notre objectif scientifique est de montrer qu'en permettant à l'utilisateur d'interagir directement sur l'ordinateur, c'est-à-dire sans intermédiaire notre système de capture de mouvements de la main sera au moins équivalent en termes de performances et meilleur pour ce qui de l'acceptabilité qu'un périphérique basé sur le modèle de Norman (e.g. la souris). Nous souhaitons donc que notre système apporte une plus-value par rapport aux autres périphériques pour les utilisateurs et ce, quelles que soient leurs caractéristiques (sexe, taille de la main, leur expertise en RV,...)

Afin de répondre à cet objectif, nous avons mené une étude visant à évaluer comparativement deux périphériques (i.e. la souris et la Cam 3D) utilisés pour réaliser une tâche de sélection d'objets 3D en environnement virtuel. Nous avons focalisé notre étude sur les performances (rapidité d'exécution de la tâche) et l'acceptabilité (perception de l'utilité, de l'utilisabilité et de l'immersion).

Pour ce faire, nous formulons l'hypothèse expérimentale suivante :

*La capture des mouvements de la main sans marqueurs est aussi performante et mieux acceptée en ce qui concerne la manipulation d'objets en environnement virtuel 3D qu'un périphérique 2D classique de type « souris » et cela quelles que soient les caractéristiques de l'utilisateur.*

Cette étude lie 2 variables dépendantes (les performances (VD1) et l'acceptabilité (VD2)) à 4 variables indépendantes (le périphérique (VI1), le sexe (VI2), la taille de la main (VI3) et l'expertise en RV (VI4)).

Pour éprouver cette hypothèse expérimentale générale, nous allons formuler 7 hypothèses expérimentales plus fines dans le sens où elles isolent chacune une variable dépendante (VD) et une variable indépendante (VI), que nous qualifions d'hypothèses opérationnelles (HO). Elles sont synthétisées dans le Tableau 5.

**Tableau 5 : Hypothèses opérationnelles de l'expérimentation 1**

HO n°	VD en fonction de VI	Intitulé
1	VD1 en fonction de VI1	Les performances (en termes de rapidité) avec la Cam3D sont au moins équivalentes à celles obtenues avec la souris.
2	VD2 en fonction de VI1	L'acceptabilité (en termes de sentiment d'immersion, d'utilité et d'utilisabilité) est meilleure avec la cam3D qu'avec la souris.
3	VD1 en fonction de VI2	Les performances (en termes de rapidité) lors de la réalisation de la tâche avec la cam3D ne sont pas significativement différentes selon le sexe de l'utilisateur
4	VD2 en fonction de VI2	L'acceptabilité (en termes de sentiment d'immersion, d'utilité et d'utilisabilité) de la cam3D n'est pas significativement différente selon le sexe de l'utilisateur.
5	VD1 en fonction de VI3	Les performances (en termes de rapidité), lors de l'exécution de la tâche avec la cam3D, ne sont pas significativement différentes selon la taille de la main de l'utilisateur.
6	VD1 en fonction de VI4	Les performances (en termes de rapidité) sont meilleures pour les personnes avec une bonne expertise en réalité virtuelle.
7	VD2 en fonction de VI4	L'acceptabilité (en termes de sentiment d'immersion, d'utilité et d'utilisabilité) est la même quelque soit l'expertise en réalité virtuelle.

#### 1.4 Tâche de sélection

On entend par tâche de sélection la tâche qui consiste à sélectionner, par ordre croissant, 10 cubes numérotés de 1 à 10 dans l'espace de jeu. Chaque boîte est numérotée et a une couleur unique afin d'aider l'utilisateur à les identifier plus facilement. Pour sélectionner un cube, l'utilisateur doit au préalable déplacer l'avatar main dans l'EV afin de le positionner correctement sur le cube à sélectionner. L'avatar doit obligatoirement toucher le cube sans quoi celui-ci ne peut pas être sélectionné. Il en est de

même, si le numéro du cube n'est pas le bon. La sélection du cube est par contre différente selon le périphérique utilisé :

- Lorsque l'utilisateur est en train de manipuler avec la Cam 3D, la sélection s'effectue en fermant la main sur l'objet, comme si l'utilisateur voulait saisir un objet réel ;
- Lorsque l'utilisateur effectue la manipulation avec la Souris, la sélection se fait grâce au bouton gauche de la Souris.

Si l'utilisateur essaie de sélectionner le cube « 3 » alors qu'il doit sélectionner le « 2 » par exemple, un message d'erreur rouge apparaît et lui spécifie que le cube qu'il essaie de sélectionner n'est pas le bon.

### 1.5 Environnement virtuel : conception

L'application se compose d'un espace de jeu cubique de 30 cm d'arrête (30cm x 30cm x 30cm). L'environnement virtuel est un cube avec une face en moins sur un des côtés permettant ainsi à l'utilisateur de voir l'intérieur du cube ainsi que les différents objets positionnés à l'intérieur (Figure 26). La tâche à réaliser consiste à sélectionner 10 petites boîtes numérotées de 1 à 10, les unes après les autres, par ordre croissant. Une fois que l'utilisateur a positionné l'avatar de la main sur la bonne boîte, il doit la faire disparaître, soit par un clic (souris), soit en fermant la main (Cam3D). Il peut ensuite passer à la boîte suivante. L'utilisateur doit réaliser cette tâche avec chacun des deux systèmes.

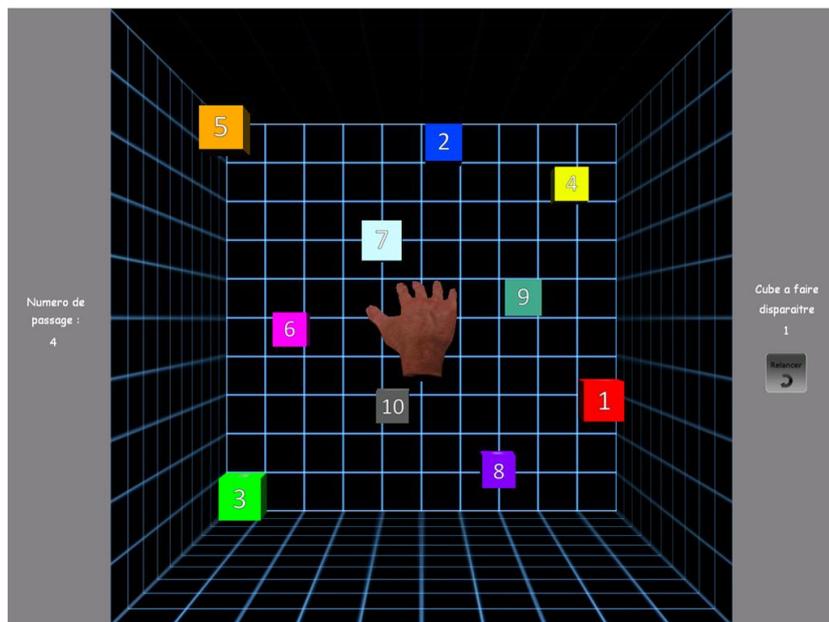


Figure 26 : Environnement virtuel - espace de jeu de l'expérimentation 1

Le jeu est affiché en perspective et l'utilisateur visualise l'environnement virtuel en stéréoscopie (active). Il peut ainsi apprécier de façon plus fine et plus précise les distances relatives entre chacune des boîtes à atteindre. La caméra est placée sur un des côtés du cube (devant la face manquante). Les positions des 10 boîtes sont fixes, en revanche l'ordre dans lequel ces dernières sont placées change entre la manipulation avec souris et la manipulation avec notre système. De cette manière, l'utilisateur ne peut donc pas mémoriser la position des cibles à atteindre, ce qui permet d'éviter tout biais dû à la mémorisation.

L'utilisateur est assis à une table, l'écran en face de lui. L'espace de jeu, dans lequel il interagit avec les objets virtuels, se situe au dessus de la table devant lui. La caméra qui capture les mouvements de

l'utilisateur est placée au-dessus de la table à environ 80 centimètres, elle ne gêne ni la vue, ni l'espace de jeu de l'utilisateur (cf. Figure 28 - p 77).

Nous avons utilisé pour cette expérimentation un ordinateur *Alienware (Intel® Core™ i7 930 cadencé à 2.80GHz avec 6 Giga-octets de RAM et une carte graphique de type NVidia GeForce GTX480)*, un écran *Samsung 120Hz (3D Vision Ready)* et des lunettes stéréo *NVidia* afin d'avoir une vision 3D stéréoscopique.

Pour acquérir la position et l'orientation de la main de l'utilisateur en 3D, nous avons utilisé, pour notre système, une caméra 3D de type *SwissRanger SR4000* de *Mesa Imaging*<sup>5</sup>.

### 1.6 Une interaction équivalente mais adaptée à chaque périphérique

Pour les deux systèmes (Souris et Cam 3D), l'interaction est la même mais a dû être légèrement adaptée dans certains cas. Visuellement, il n'y a aucune différence à l'écran, pour l'utilisateur, lors de l'utilisation des deux systèmes. En revanche, la modalité d'interaction est différente. En premier lieu, les déplacements de l'avatar avec la Cam3D s'effectuent à l'échelle 1:1, c'est-à-dire que les mouvements de la main de l'utilisateur sont retranscrits à l'identique dans l'EV. Un déplacement de la main de 10 cm en réel, correspond à 10 cm dans l'EV. Pour la souris, les déplacements de l'avatar sont indexés sur les mouvements de celle-ci à la façon d'un curseur. Cela pose deux problèmes principaux. D'une part, le déplacement n'est pas à l'échelle 1:1 car nous avons décidé de conserver les paramètres originaux de la souris. Cela, afin que l'utilisateur manipule dans les conditions « normales » d'utilisation, c'est à dire dans lesquelles il a l'habitude de travailler. Ainsi, pour déplacer l'avatar de quelques centimètres dans l'EV, un déplacement réel de la main de l'ordre de quelques millimètres est suffisant. D'autre part, le suivi des mouvements ne peut être fait qu'en deux dimensions (dans le plan de l'écran), il a donc fallu trouver une solution permettant de travailler aussi en profondeur. Dans ce cas, nous avons dû adapter la modalité d'interaction. Nous avons choisi, la transposition la plus simple de l'interaction courante : le déplacement en profondeur est assuré par un déplacement « avant / arrière » de la souris avec le maintien du bouton droit enfoncé. Cette méthode est utilisée dans certains logiciels pour effectuer des zooms par exemple. Enfin, la dernière différence se situe au niveau de la sélection de l'objet. Nous avons opté pour la fermeture de la main dans le système Cam3D afin de nous approcher le plus possible d'un mouvement réel de préhension. L'utilisateur « saisit » l'objet virtuel comme il saisirait un cube sur une table. Pour la souris, nous avons conservé la sélection par un clic le bouton gauche pour sélectionner l'objet, ce qui est déjà son utilisation la plus courante.

### 1.7 Implémentation et développements logiciels

Pour cette première expérimentation, il a fallu d'une part développer notre système de capture de mouvements à partir des données d'une caméra 3D, et d'autre part, développer une application pour l'expérimentation.

Notre système est un logiciel développé en C++, utilisant à la fois les données issues de la caméra 3D et une librairie permettant le traitement d'images. Les données provenant de la caméra 3D sont des données brutes qu'il faut nettoyer, puis leur appliquer des transformations avant de pouvoir les analyser. La caméra 3D envoie des « images » 2D dans lesquelles l'information de profondeur est donnée à la place des informations de couleur. Ainsi pour chaque pixel, une valeur (la distance entre l'objet et la caméra) en mètres est obtenue. Ce tableau en deux dimensions (x et y) constitue par

---

<sup>5</sup> <http://www.mesa-imaging.ch/>

conséquent une sorte d'image. Cela correspond en réalité à la carte de profondeurs de la scène. Si on essaie de repositionner chacun de ces points dans l'espace, on obtient un nuage de points représentant la scène. Malheureusement, ce nuage de point est imprécis et contient de nombreuses erreurs, en particulier au niveau des frontières des objets. La caméra n'arrive pas à déterminer si un point du contour appartient à l'objet ou au fond, ce qui fait que les données peuvent osciller entre les deux. Il faut donc dans un premier temps nettoyer le nuage de points afin d'éliminer les données aberrantes. D'autre part, la caméra peut donner une seconde image appelée carte de confiance qui permet de déterminer le degré de confiance de chaque pixel vis-à-vis des distances obtenues et une dernière image peut être obtenue. Elle représente l'intensité lumineuse reçue par chacun des pixels du capteur. A partir de cette dernière, on peut reconstituer une image en niveaux de gris de la scène.

Pour la première version de notre système, qui est utilisée pour cette expérimentation, une phase d'initialisation est requise. Cette initialisation dure en tout et pour tout moins d'une seconde car il suffit de prendre l'image de fond. En réalité cette phase est transparente pour l'utilisateur car il suffit que sa main ne soit pas dans le champ de vision au commencement. On capture tout d'abord une image de la scène sans aucun objet présent, c'est-à-dire que l'on prend non pas une « image de fond », mais une « carte de profondeur de fond ». De cette manière, on connaît la position et la distance de la table et des objets de l'environnement par rapport à la caméra 3D. Dans un second temps, lorsque la main de l'utilisateur entre dans le champ de vision de la caméra, on prend une image de la main (que l'on sauvegarde) puis on supprime de l'image toutes les données qui correspondent au fond. On ne garde, par conséquent, que les données de profondeur qui ont varié, c'est-à-dire dans notre cas la main de l'utilisateur. Comme les données ne sont pas précises au millimètre près, il est nécessaire de mettre un seuil de tolérance d'un ou deux centimètres.

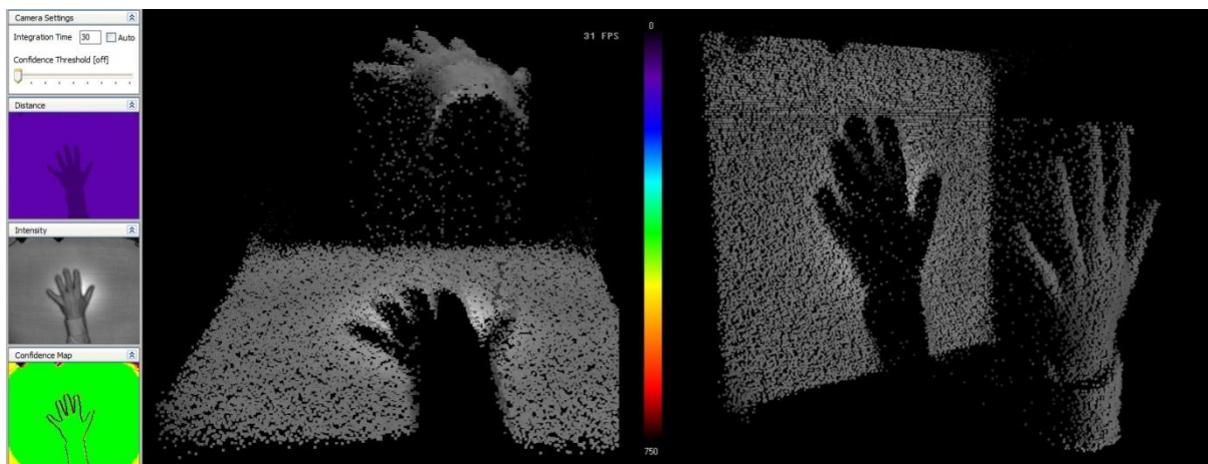


Figure 27 : Cartes de distances obtenues avec une SwissRanger SR4000. Le bruit est principalement présent au niveau des contours de la main.

L'image de la main sauvegardée nous sert à créer un modèle virtuel de la main en respectant les caractéristiques de l'utilisateur. De cette manière, on calcule la longueur totale de la main (distance entre l'extrémité du majeur et le milieu du poignet). A partir de cette valeur, il est possible de déterminer les longueurs moyennes des phalanges qui sont proportionnelles à la taille de la main. Pour réaliser cela, nous nous sommes basés sur les travaux de (Wagner, 1988). Ensuite, nous calculons les limites articulaires du modèle virtuel d'après les travaux de Pouliquen (Pouliquen, 2006) et (Chalfoun et al., 2003). Le modèle virtuel est donc adaptable à chaque utilisateur.

Lorsque la main de l'utilisateur est seule présente sur l'image résultante, une série de traitements est nécessaire afin de déterminer la position de celle-ci ainsi que son orientation. Nous avons utilisé pour cela la librairie *OpenCV* qui permet de réaliser de nombreux traitements d'images. Dans un premier temps, on détermine les contours de la main ainsi que le centre de celle-ci. On détermine ensuite 4 points autour de ce point central (en forme de losange) puis on récupère la valeur de distance correspondante pour chacun d'eux. La pente définie entre deux points opposés permet de déterminer le roulis, le tangage, ainsi que le lacet de la main. Cette méthode trouve ses limites lorsque la pente de la main est trop grande (i.e. si la main est perpendiculaire à la caméra), car il devient impossible de connaître son orientation. Dans le cas précis de notre expérimentation, ce problème n'est pas gênant étant donné que nous n'utilisons pas la donnée d'orientation. En effet, pour le déplacement de main et la sélection d'objets, la main est toujours « à plat » et donc face à la caméra.

Un dernier traitement est nécessaire pour déterminer si la main est ouverte ou fermée. Pour cela, nous nous basons uniquement sur le nombre de doigts visibles par la caméra. Si aucun doigt n'est détecté, on considère que la main est fermée, autrement elle est considérée comme étant ouverte. Ainsi si l'utilisateur n'a que deux doigts levés, la main sera considérée par le système comme ouverte. Il aurait été possible, bien sûr, de déterminer précisément le nombre de doigts levés et éventuellement déclencher différentes actions en fonction de celui-ci, mais nous ne nous servons uniquement des états « main ouverte » et « main fermée » pour la tâche de sélection.

Une fois que les valeurs d'orientation, de position dans l'espace, ainsi que l'état de la main (ouverte ou fermée) sont obtenues, les données sont envoyées par connexion UDP à un autre logiciel qui se charge de réaliser les interactions dans l'EV. Nous avons, pour cela, eu recours au logiciel *Virtools*.

## 2 Protocole expérimental

### 2.1 Généralités

Ce chapitre décrit l'étude empirique menée en vue d'évaluer les performances (rapidité d'exécution de la tâche) et l'acceptabilité (perception de l'utilité, de l'utilisabilité et de l'immersion) du système de capture de mouvements sans capteurs que nous avons créé comparativement à la souris.

### 2.2 Choix des participants

Cette étude a impliqué 71 participants volontaires (21 femmes et 50 hommes) d'âge compris entre 19 et 57 ans (âge moyen = 27,7 ans, écart-type = 8,5). La majorité des participants étaient étudiants de premier ou de deuxième cycle universitaire, issus du Master « Ingénierie du Virtuel et de l'Innovation », de l'« École Supérieure de Création Interactive Numérique » et de l'« Association pour la Formation Professionnelle des Adultes ».

Il n'a pas été possible d'équilibrer le nombre d'hommes et de femmes sans diminuer l'effectif total de participants. Nous avons tenu compte de cette variable dans l'étude des performances et des jugements des participants. Comme nos hypothèses de travail le soulignent, outre le sexe, les participants ont également été caractérisés selon la taille de leur main et leur degré d'expertise en Réalité Virtuelle. Ces mesures ont été faites a posteriori et sont expliquées dans la partie « Mesures ».

### 2.3 Conditions initiales et consignes

Les conditions initiales sont les mêmes quel que soit l'ordre de passation. Seuls les cubes dans l'environnement ne sont pas positionnés de la même façon selon si l'utilisateur commence par la souris ou par la Cam 3D. Nous avons décidé de garder les mêmes positions entre les différents participants mais de changer l'ordre des cubes entre la souris et la Cam 3D afin de ne pas introduire d'effet d'apprentissage dans les résultats. Le participant commence toujours par une phase d'apprentissage qui lui permet de comprendre les modalités d'interaction légèrement différentes entre les 2 périphériques ainsi que de se familiariser avec la tâche à réaliser. La durée d'entraînement est limitée à une minute sur chacun des systèmes.

L'utilisateur est assis à une table, l'écran en face de lui. L'espace de jeu, dans lequel il interagit avec les objets virtuels, se situe au-dessus de la table devant lui. Dans le cas de la Cam3D, la caméra utilisée pour capturer les mouvements de la main est placée au-dessus de la table c'est-à-dire à environ 80 centimètres de hauteur, elle ne gêne ni la vue, ni l'espace de jeu de l'utilisateur. Dans le cas de la souris, la table sur laquelle elle est posée est dégagée et permet à l'utilisateur d'utiliser tout l'espace qu'il souhaite (bien qu'une petite zone soit suffisante, du fait de la faible amplitude nécessaire à la souris pour parcourir l'espace de jeu).

Il était demandé aux participants de réaliser la tâche le plus rapidement possible.

Des consignes sont données en début d'expérimentation, afin d'expliquer en quoi consiste l'expérimentation et la tâche que doit réaliser le participant. Durant la phase d'apprentissage, les consignes et les instructions concernant l'interaction et la tâche à réaliser sont détaillées au participant. En plus de ces instructions orales, une explication écrite est donnée à l'utilisateur en guise de rappel sur la manière de déplacer l'avatar et de sélectionner les cubes, à la fois avec la souris et avec la Cam 3D.

Aucune consigne supplémentaire n'a été donnée pendant l'expérimentation. Néanmoins, si un participant éprouvait le besoin de poser une question, il pouvait le faire. Toutefois, aucun des 71 participants n'a eu besoin de poser des questions pendant l'expérimentation.

### 2.4 Déroulement de l'expérimentation

Le déroulement de l'expérimentation est identique, que le participant commence avec la souris ou avec la Cam3D. En premier lieu, le participant commence par remplir un questionnaire d'identification qui nous permet de recueillir les informations générales le concernant (âge, sexe, latéralisation, niveau d'étude) ainsi que les informations spécifiques nous permettant d'établir son niveau d'expertise en réalité virtuelle. Ce questionnaire est un formulaire au format *PDF* réalisé avec le logiciel *Adobe LiveCycle Form*.

Ensuite, l'utilisateur se voit présenter l'expérimentation et les enjeux de celle-ci. Une brève explication écrite est également donnée à chaque participant, en guise de rappel ; celle-ci est consultable à tout moment durant l'expérimentation. L'expérimentation proprement dite commence par une phase d'entraînement (limitée à une minute), dans laquelle les participants doivent sélectionner 10 fois de suite un cube positionné aléatoirement dans l'espace de jeu, avec chaque périphérique successivement (souris et cam3D). Afin de ne pas introduire de biais dans l'expérimentation, un seul cube est présent dans la phase de prise en main (contrairement à la tâche à réaliser). Lorsque l'utilisateur a sélectionné le cube, celui-ci disparaît et va se positionner à un autre endroit dans l'EV. Quand l'utilisateur a réussi dix fois de suite la sélection du cube, ou si limite temporelle est atteinte, la phase de prise en main s'arrête.

Les participants peuvent alors réaliser la tâche expérimentale. Celle-ci correspond à la sélection par ordre croissant de 10 cubes de couleur, numérotés, et positionnés dans l'espace de jeu. Le temps total a été fixé à 5 min pour éviter un sentiment de découragement de la part des participants. Le temps écoulé n'était pas affiché à l'écran pour ne pas induire de sentiment de stress aux participants et la seule indication visible correspondait au numéro du cube courant.

Lorsque le participant a effectué l'expérimentation avec la Souris et avec la Cam 3D, il doit remplir un dernier questionnaire *PDF* (réalisé lui aussi avec le logiciel *Adobe LiveCycle Form*). Ce questionnaire post-passation (cf. Annexe A – Questionnaire Post-passation) nous permet de recueillir les impressions subjectives du participant, ses suggestions d'améliorations, les points positifs et négatifs de chacun des systèmes ainsi que ses préférences vis-à-vis d'un système en particulier.

Enfin, une mesure de la taille de la main du participant est réalisée en vue de l'analyse des données.

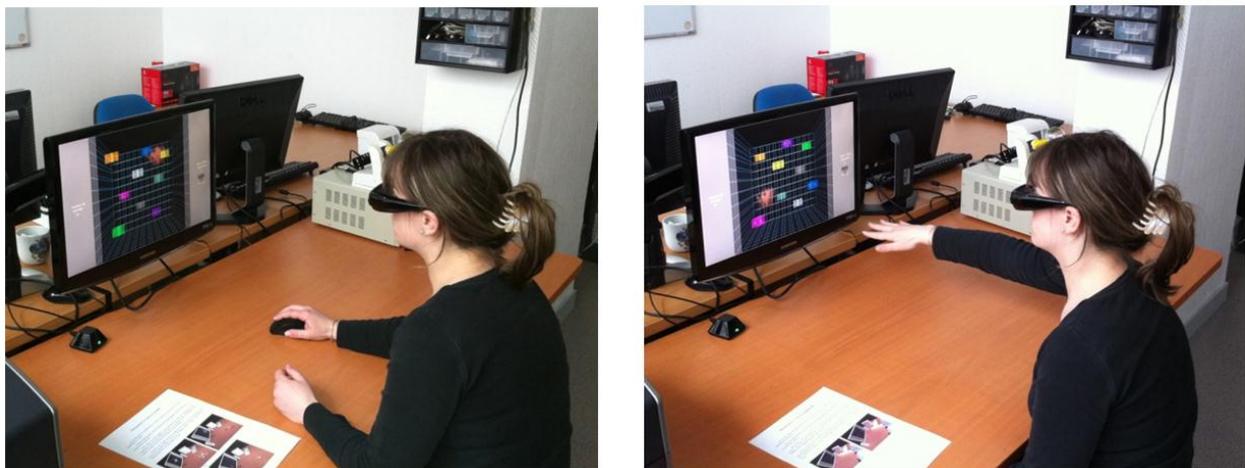


Figure 28 : Une participante pendant l'expérimentation, avec la souris (à gauche) et la cam 3D (à droite)

## 2.5 Données recueillies

Les interactions des participants avec l'application et avec les 2 périphériques ont été enregistrées tout au long de l'expérimentation : tous les événements, qu'ils soient générés par les participants (déplacements de l'avatar, sélections, etc.) ou par le système (début et fin d'expérimentation, les différents temps, etc.) ont été automatiquement marqués, datés et sauvegardés dans un fichier externe. Cette procédure de collecte a permis d'automatiser l'extraction des données importantes.

Avant la passation, les participants devaient remplir un questionnaire d'identification (sous la forme d'un formulaire *PDF*) permettant d'obtenir des informations de nature générale (âge, sexe, latéralisation,...), ainsi que des précisions sur leurs connaissances générales (diplômes obtenus ou en cours), leurs connaissances informatiques et leurs connaissances dans le domaine de la réalité virtuelle. Ce premier questionnaire nous a permis de « qualifier » les participants puis d'établir leur niveau d'expertise en réalité virtuelle.

Les jugements subjectifs et préférences des participants ont été obtenus par l'intermédiaire d'un questionnaire post-expérimentation (formulaire *PDF*). Les participants étaient invités à évaluer différentes caractéristiques de la Cam 3D et de la souris, sur des critères relevant de l'utilité, de l'utilisabilité ou de l'immersion. Ils pouvaient également évaluer la tâche à réaliser et suggérer des améliorations au système proposé. Les participants étaient amenés à répondre à des questions

ouvertes, des questions fermées ainsi qu'à des échelles de Likert à 6 modalités, pour lesquelles les participants devaient indiquer leur degré d'adéquation par rapport à un critère particulier (Likert, 1932).

Enfin, la taille de la main de chaque participant était mesurée.

## 2.6 Variables et mesures

Pour étudier les performances des participants, nous avons pris en compte le temps total d'exécution de la tâche de sélection des dix cubes. C'est le temps que met l'utilisateur pour accomplir les 10 sélections avec un des deux périphériques. Il débute après que l'utilisateur a lu la fiche explicative ainsi que les consignes. Il se termine quand le dixième cube a été sélectionné par le participant.

L'étude des préférences subjectives et des commentaires des participants est divisée en deux parties. D'une part, les appréciations globales concernant les préférences d'un système vis-à-vis de l'autre pour déterminer quel système est :

- le moins fatiguant ;
- le plus efficace (en termes de rapidité, de précision) ;
- le plus immersif ;
- le plus approprié pour la tâche ;
- le plus aisé à utiliser.

Et d'autre part, les appréciations concernant chacun des périphériques selon des échelles de Likert à 6 modalités ou des commentaires écrits :

- le degré de fatigue ressentie ;
- l'apport du système pour les interactions en EV ;
- suggestions d'améliorations ;
- problèmes rencontrés, difficultés.

Les réponses aux échelles de Likert sont traitées comme des variables numériques, chaque modalité étant codée par un nombre (1 = plus mauvais ; 6 = meilleur).

Premièrement, nous souhaitons comparer l'interaction avec la Cam3D et l'interaction avec la souris sous l'angle des performances et de l'acceptabilité. D'autre part, nous souhaitons également évaluer l'influence du sexe, de la taille de la main et de l'expertise en réalité virtuelle sur ces 2 critères. Nous allons dans un premier temps détailler la méthode de calcul de l'expertise en réalité virtuelle puis la mesure de la taille de la main. Dans un second temps, nous présenterons les mesures de performance et d'acceptabilité.

## 2.7 Méthodes d'analyse et tests statistiques

### 2.7.1 Expertise en réalité virtuelle

L'expertise a été calculée à partir des réponses des participants à certaines questions du questionnaire d'identification, toutes sous la forme d'échelles de Likert à 5 modalités, comme par exemple « Avez-vous des connaissances en réalité virtuelle ? » : les participants devant alors se positionner entre « pas du tout » et « excellentes connaissances ». Les autres questions prises en compte correspondaient, d'une part, aux périphériques utilisés et à leur fréquence d'utilisation, et d'autre part aux activités liées à la 3D pratiquées. L'ensemble des questions ont été regroupées en trois thématiques, les

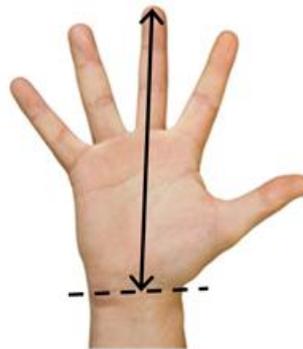
connaissances générales en RV, l'utilisation des périphériques liés à la RV et enfin l'utilisation des logiciels spécifiques liés à la RV. Chaque modalité a été convertie en un score : « 1 » = niveau le plus bas, « 5 » = niveau le plus élevé. Les scores ont été sommés pour chaque participant, suivant différentes pondérations. Trois groupes de niveau d'expertise ont alors été constitués, comme cela est résumé dans le tableau suivant (Tableau 6) :

**Tableau 6 : Niveau d'expertise en fonction du score calculé d'après les réponses aux échelles de Likert et au questionnaire d'identification.**

Niveau d'expérience	Novice	Initié	Expert
Score	de 0 à 1,5	de 1,5 à 3	de 3 à 5
Effectif par groupe	14	33	24

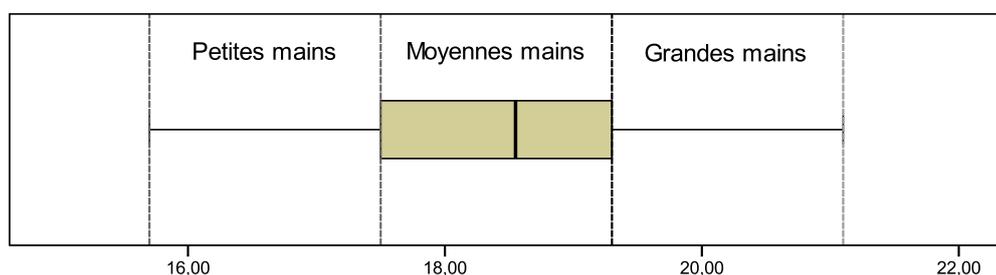
### 2.7.2 Taille de la main

La taille de la main de chaque participant a été obtenue en mesurant la distance, doigts écartés, entre l'extrémité du majeur et l'extrémité du poignet, sous la paume. La Figure 29 illustre la méthode de mesure :



**Figure 29 : Méthode de mesure de la taille de la main**

Pour répartir les participants entre 3 groupes, nous avons calculé la médiane (18,55) et les quartiles (17,50 et 19,30). Ainsi, une main de « taille moyenne » a sa mesure comprise entre 17,5 cm et 19,3 cm. La répartition est donc la suivante :



**Figure 30 : Répartition des participants en fonction de la taille de leur main, en 3 groupes**

Nous observons par ailleurs que la valeur médiane calculée pour l'ensemble de nos participants reste très proche des valeurs moyennes que l'on peut trouver dans la littérature (Ilayperuma et al., 2009). La répartition des participants est présentée dans le Tableau 7.

Tableau 7 : Taille de la main en fonction du score calculé d'après les réponses aux échelles de Likert

Taille de la main	Petite	Moyenne	Grande
Effectif par groupe	18	35	15

### 2.7.3 Performances / Acceptabilité

Le temps d'exécution total de la tâche T (en secondes) a servi de base à l'étude des performances des participants. Cela correspond au temps nécessaire pour sélectionner les 10 cubes et le faire selon l'ordre numéroté.

L'étude de l'acceptabilité repose sur les réponses des participants au questionnaire post-passation. Les questions posées étaient de deux types :

- des questions fermées, pour lesquelles les participants devaient marquer une préférence pour l'un ou l'autre périphérique, par exemple :

⇒ *Quel système, souris ou Cam 3D, vous a semblé le plus facile à utiliser ?*

Souris       Cam 3D

Question fermée prise en compte pour l'étude de l'acceptabilité

- des échelles de Likert à six modalités pour lesquelles les participants doivent donner leur degré d'adéquation par rapport un critère particulier, par exemple :

⇒ *Comment jugez-vous l'apport de la Cam 3D pour les interactions en environnement 3D ?*

Aucun                 Elevé

Question à échelle de Likert prise en compte pour l'étude de l'acceptabilité

Nous avons étudié l'acceptabilité sous l'angle de 3 critères : l'utilité, l'immersion et l'utilisabilité. Ce dernier critère sera évalué selon l'efficacité, la satisfaction et l'efficacité.

## 2.8 Validité interne et externe

Nous avons mis en œuvre différentes techniques pour assurer une validité interne et externe à cette expérimentation. Nous avons essayé de minimiser l'influence de certaines variables qu'il nous est difficile de contrôler parfaitement, comme par exemple le milieu social du participant. Les variables parasites provenant de l'environnement réel (éclairage, bruit, température) sont quant à elles maîtrisées, puisque l'expérimentation se tient dans une salle unique, climatisée et sans fenêtre.

### 2.8.1 Appariement

Compte tenu de l'effectif relativement faible des participants (et particulièrement du nombre de femmes), il semblait hasardeux de constituer 2 groupes distincts pour l'évaluation de chaque système. Par conséquent, chaque participant évalue alternativement chaque système. L'appariement permet de contrôler les facteurs parasites, car ils sont constants tout au long de l'expérimentation, mais présente l'inconvénient d'introduire des effets d'apprentissages non contrôlés (i.e. chaque participant acquiert,

au cours de l'expérimentation, un certain niveau de pratique, et peut ainsi accomplir la tâche plus rapidement).

### **2.8.2 Contre-balancement**

L'ordre de passation peut donc jouer un rôle non négligeable sur les performances, ainsi que sur les préférences (la motivation, comme la fatigue ne sont pas constantes d'un sujet à l'autre). Nous avons défini deux groupes de sujets, G1 et G2, pour contrebalancer cet ordre de passation. Ainsi, les sujets du groupe G1 commencent l'expérimentation en effectuant la tâche avec la souris, tandis que les sujets du groupe G2 commencent l'expérimentation avec la cam3D. Chaque groupe comprend autant d'hommes et de femmes (à une personne près). Le groupe 1 est constitué de 10 femmes et 26 hommes et le groupe 2 de 10 femmes et 25 hommes.

### **2.8.3 Affectation aléatoire**

L'affectation des participants à un groupe ou à l'autre (G1 ou G2) s'est faite de façon aléatoire, de manière à ne pas favoriser involontairement l'un ou l'autre des systèmes. En revanche, la position des cubes est déterminée à l'avance, et est la même pour tous les participants mais différente entre les deux périphériques.

## **3 Influence du périphérique sur les performances**

### **3.1 Tests préliminaires**

En premier lieu, nous avons effectué des tests de normalité (Kolgomorov-Smirnov) sur les variables relatives aux performances (temps total d'exécution de la tâche), pour chaque périphérique, afin d'évaluer le type de test statistique à mener par la suite. En cas de non normalité des distributions, nous avons eu recours à des transformations. Bien que leur usage soit sujet à discussion (Pallant, 2007), elles sont suggérées dans de nombreux manuels de statistiques et mises en œuvre par de nombreux auteurs. De plus, les tests non paramétriques se révèlent limités pour l'étude des interactions entre plusieurs effets sur des modèles mixtes. Le cas échéant, une analyse qualitative de la forme des distributions a permis de déterminer la transformation mathématique à appliquer. De fait, l'ensemble des tests réalisés pour les performances appartient à la famille des tests paramétriques.

Pour vérifier l'absence d'effet d'apprentissage, normalement annulé par le contre-balancement, nous avons réalisé un test T de Student pour échantillons indépendants (ordre de passation  $x$  périphérique).

L'étude principale a consisté en l'analyse de l'influence du type de périphérique sur les performances des participants (test T de Student pour échantillons appariés) et en l'analyse de l'influence du sexe, de la taille de la main et de l'expérience en réalité virtuelle sur les performances en lien avec le périphérique. Pour cette dernière analyse, nous avons effectué une ANOVA à plan mixte composée d'une variable intra-sujet (le périphérique) et d'une variable inter-sujet (sexe ou taille de la main ou expérience en RV). Pour l'analyse éventuelle des effets simples, nous avons eu recours, selon le nombre de modalités de la variable étudiée, soit à un test T pour échantillons indépendants, soit à une ANOVA à un facteur. Pour l'analyse éventuelle des effets principaux, nous avons mené des tests post-hoc de type LSD.

Nous nous sommes basés, pour tous ces tests, sur le seuil de significativité le plus courant, à savoir  $p < 0,05$ .

### 3.1.1 Normalisation

Les temps d'exécution mesurés pour la souris ne suivent pas une loi normale (Kolmogorov-Smirnov ;  $Z = 1,680$  ;  $p = 0,007$ ), contrairement à ceux mesurés pour la Cam3D. Nous avons donc fait le choix, en fonction de la forme de la distribution, d'appliquer une transformation logarithmique (Log10).

### 3.1.2 Etude de l'effet d'apprentissage

Pour vérifier s'il existe ou non un effet d'apprentissage, nous avons eu recours à un test T pour échantillons indépendants. Les résultats montrent qu'il n'y a pas d'effet d'apprentissage pour la souris ( $t = 0,044$  ; NS) ni pour la Cam3D ( $t = 0,858$  ; NS).

## 3.2 Résultats globaux : temps d'exécution

### 3.2.1 Influence du type de périphérique sur les performances

Il n'y a pas de différence significative de performances des participants entre les 2 périphériques utilisés, « Souris » et « Cam3D » ( $t = 1,363$  ;  $p = 0,177$ ). Les résultats sont présentés dans le Tableau 8.

Tableau 8 : Temps d'exécution avec la Cam3D et la Souris – données originales et transformées (Log10)

		Données originales		Données (Log10)	
		Cam3D	Souris	Cam3D	Souris
Moyenne		45,89	49,12	1,65	1,67
E.T.		13,29	19,15	,12	,15
t-Test	t			1,363	
	p			0,177	

L'étude qualitative confirme ces résultats en montrant que 19 des 71 participants sont plus rapides avec la souris (différence de plus de 15%), quand 26 participants sont plus rapides avec la Cam3D et 26 ont des performances équivalentes entre les 2 systèmes.

Notons néanmoins qu'en supprimant les participants ayant obtenu des temps supérieurs à 4 écarts-type de la moyenne (la règle en usage étant plutôt de considérer 1,5 ou 2 écarts-types - un seul participant est concerné ici), la significativité du test statistique est légèrement modifiée ( $t = 1,698$  /  $p = 0,094$ ). Ce dernier résultat indique une légère tendance en faveur de la Cam3D (moyenne = 1,64 ; E.T. = 0,11) par rapport à la souris (moyenne = 1,66 ; E.T. = 0,14).

Enfin, si les hommes obtiennent des résultats similaires avec les 2 systèmes ( $t = 0,458$  ;  $p = 0,649$ ), les femmes sont significativement plus rapides avec le système « Cam3D » ( $t = 3,417$  ;  $p = 0,003$ ).

### 3.2.2 Discussion

Ces résultats confirment l'hypothèse opérationnelle HO1 (cf : Tableau 5). Même s'ils peuvent apparaître relativement neutres, ils sont encourageants. Outre le fait que nous obtenons une légère tendance en faveur de la Cam3D, qu'il conviendrait de confirmer ou d'infirmer avec davantage de participants, nous pouvons d'ores et déjà affirmer que le système basé sur la Cam3D a le même niveau de performances que la souris pour la tâche considérée. Malgré le soin apporté à rendre le système « Cam3D » aussi simple que possible, la souris présente tout de même un avantage majeur sur tout autre système, celui de l'utilisation fréquente, qui tend à minimiser le temps d'apprentissage, même dans le cas d'une tâche

originale ou nouvelle. Les résultats obtenus ici nous incitent donc à réaliser une étude plus longue, qui permettrait de mettre davantage en valeur les différences entre les 2 systèmes.

### 3.3 Interaction entre le sexe et le périphérique sur les performances

Une ANOVA à plan mixte (sexe x périphérique) indique une différence significative selon le périphérique ( $F(1,69) = 6,62$  ;  $p = 0,012$ ). Nous notons également un effet du sexe ( $F(1,69) = 6,33$  ;  $p = 0,014$ ) et enfin une interaction entre ces deux facteurs ( $F(1,69) = 9,448$  ;  $p = 0,003$ ).

Une analyse des effets simples, menée via un test T à échantillons indépendants pour chaque périphérique, ne met pas en évidence de différence significative entre les hommes et les femmes pour la « Cam3D » ( $t = 1,008$  ; NS). A contrario, avec la souris, les hommes sont significativement plus rapides que les femmes ( $t = 3,345$  ;  $p = 0,001$ ). Ces résultats sont synthétisés dans le Tableau 9.

**Tableau 9 : Moyennes des temps d'exécution avec la Cam3D et la Souris selon le sexe – données originales et transformées (Log10)**

	Sexe	Données originales		Données (Log10)	
		moyenne	écart-type	moyenne	écart-type
Souris	H	44,13	12,56	1,63	0,114
	F	60,98	26,20	1,75	0,182
Cam3D	H	44,44	10,40	1,64	0,098
	F	49,35	18,30	1,67	0,152

En accord avec notre hypothèse de travail HO3, les performances avec notre système ne sont pas influencées par le sexe de l'utilisateur. Il est inattendu de trouver une différence significative de performances avec la souris. Ces derniers résultats sont bien sûr liés à l'étude réalisée. Ils mériteraient néanmoins d'être étudiés pour d'autres actions possibles en environnements virtuels, comme par exemple la navigation ou la manipulation d'objets, cependant nous n'avons pas développé ce type d'interaction car cela n'était pas l'objet de cette étude.

### 3.4 Interaction entre l'expertise en R.V. et le périphérique sur les performances

Nous avons défini trois groupes d'expertise en réalité virtuelle : novice, initié et expert. Une ANOVA à plan mixte (expertise en RV x périphérique) n'indique pas d'effet de l'expertise ni du périphérique, mais une interaction significative entre ces deux facteurs ( $F(2,69) = 4,150$  ;  $p = 0,020$ ). Une ANOVA à un facteur, pour étudier l'influence des effets simples (i.e. l'influence de l'expérience en RV sur les performances pour chaque périphérique), indique de façon significative que les personnes qualifiées d'expertes ou d'initiées en Réalité Virtuelle sont plus performantes que celles qualifiées de novices, aussi bien dans l'utilisation de la souris que de la Cam3D (Tableau 10Tableau 9). Ces différences sont illustrées dans la Figure 31.

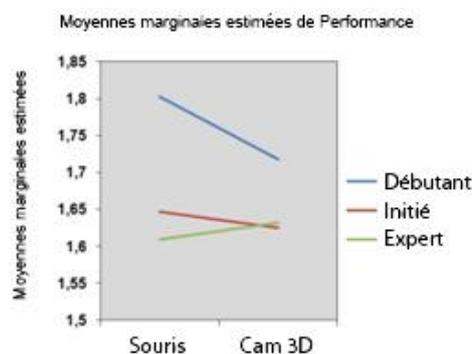


Figure 31 : Différences de performances pour chaque périphérique selon le niveau d'expertises en RV

Tableau 10 : Résultats de l'analyse des performances en fonction d'expertise en RV pour chaque périphérique

	Novice vs. Expert	Novice vs. Initié
Souris	$p < 0,005$	$p < 0,005$
Cam3D	$p = 0,027$	$p = 0,012$

Enfin, nous avons comparé les performances mesurées avec chaque périphérique indépendamment pour chaque groupe d'expertise. Il n'y a pas de différence significative entre les deux périphériques pour les groupes d'expertise « Initié » et « Expert ». En revanche, pour le groupe « Novice », les performances sont significativement meilleures avec la Cam3D ( $t = 2,650$  ;  $p = 0,020$ ).

## Discussion

Les différents résultats rapportés ici apportent plusieurs réponses à notre questionnement. D'une part, nous pouvons constater que, en accord avec notre hypothèse opérationnelle HO6, l'expertise en réalité virtuelle a une influence sur les performances, mais que cet effet est similaire avec les 2 périphériques proposés pour la tâche à réaliser. D'autre part, nous remarquons que la « Cam3D » apporte un avantage significatif pour les personnes qualifiées de « Novices ». La prise en main immédiate de notre système, sans recours à une métaphore d'interaction peut expliquer ce résultat. Les personnes ayant une expertise plus importante ont déjà rencontré la problématique de positionnement dans un espace 3D avec un périphérique simple (clavier + souris ou souris seule). De nouveaux périphériques ou visualisations peuvent provoquer des phénomènes de rejet, s'ils n'apportent pas un avantage immédiat à l'utilisateur. Notre dernier résultat est encourageant puisqu'il confirme l'intérêt de notre système en particulier pour les novices.

### 3.5 Interaction entre la taille de la main et le périphérique sur les performances

Une analyse ANOVA « taille de la main x périphérique » ne met pas en évidence d'interaction entre la taille de la main et le périphérique ( $F(2,65) = 1,288$  ; NS). Il existe par contre un effet principal de la taille de la main sur les performances ( $F(2,65) = 3,937$  ;  $p = 0,024$ ). Les tests LSD indiquent une différence significative de performances entre les participants ayant des mains caractérisées de petites et les participants ayant des mains caractérisées de grandes, en faveur de ces dernières ( $MD = 0,1117$  ;  $p = 0,007$ ).

Tableau 11 : Performances en fonction de la taille de la main pour chaque périphérique

		Petite	Moyenne	Grande
Souris	Moyenne	1,74	1,66	1,60
	Dev. std.	0,17	0,14	0,07
Cam 3D	Moyenne	1,68	1,65	1,60
	Dev. std.	0,16	0,10	0,07

### Discussion

Pour le périphérique « Cam3D », la taille de la main pouvait avoir une importance dans la capacité du système à reconnaître les mouvements « main ouverte » et « main fermée ». De mauvaises reconnaissances pouvaient alors impacter les performances, comme par exemple si la main sortait de l'espace de jeu. Cette particularité pouvait impacter majoritairement les gens avec des mains caractérisées de « grandes ». Il est surprenant de constater que les personnes avec de petites mains sont moins performantes. De plus, l'analyse statistique a indiqué qu'il n'y avait pas de relation entre la taille de la main et le sexe. Puisqu'on obtient le même effet pour la souris, il est difficile d'avancer une explication ou même de formuler une hypothèse pour tenter de l'expliquer. Ce résultat est cependant en désaccord avec notre hypothèse HO5, puisque la taille de la main a une influence sur les performances avec notre système. Néanmoins, le fait que les résultats soient semblables avec la souris nous force à rester prudents sur cette conclusion et impose de mener des tests complémentaires afin d'apporter une explication. Il apparaît en tous les cas clair que le type de périphérique n'est pas directement responsable de ces différences.

### 3.6 Discussion

Notre étude des performances s'est déroulée en plusieurs étapes, en lien avec nos hypothèses de travail. D'une part, nous avons étudié avec quel périphérique les participants obtenaient de meilleures performances, puis nous avons mené une analyse plus fine pour mesurer l'impact de trois facteurs sur les performances : le sexe, l'expertise en réalité virtuelle et la taille de la main.

En accord avec notre hypothèse HO1, le périphérique « Cam3D » permet d'obtenir des performances sensiblement équivalentes à celles obtenues avec la souris pour la tâche de sélection dans un environnement virtuel en 3D. Compte tenu de la nouveauté de ce type d'interaction, ce premier résultat est très encourageant. Concernant le premier facteur (le sexe), nous n'avons pas observé de différences dans les résultats pour la Cam 3D, contrairement à la souris (ce qui est en accord avec l'hypothèse HO3). Les résultats sont plus mitigés pour les deux autres facteurs : nous observons des résultats similaires pour les deux périphériques, avec de meilleures performances pour les personnes avec des mains caractérisées de « grandes » (contrairement à notre hypothèse HO5) et pour les personnes qualifiées d'« experts » ou d'« initiés » en Réalité virtuelle (en accord avec HO6). Le fait que nous obtenions des résultats similaires pour les deux périphériques semble limiter le rôle de ces deux facteurs.

Il serait donc intéressant d'étudier d'une part l'évolution des performances sur une période plus longue, afin de voir si une utilisation plus longue peut entraîner une supériorité sur la souris, qui, elle, bénéficie d'une pratique beaucoup plus importante. Pour une tâche de sélection, le périphérique « Cam3D » peut d'ores et déjà remplacer le périphérique le plus courant, sans perte de performance. Il reste néanmoins à étudier l'acceptabilité de ce système comparé à la souris. C'est en conjonction avec ce dernier critère que nous pourrions établir l'intérêt de la « Cam3D », et donc plus généralement de la capture de

mouvement sans marqueur pour des tâches simples comme la sélection d'objets en environnement virtuel.

## 4 Influence du périphérique sur les préférences subjectives : étude de l'acceptabilité

### 4.1 Résultats globaux

#### 4.1.1 Statistiques

Comme cela a été défini auparavant, l'acceptabilité a été étudiée sous l'angle de l'utilité, de l'immersion et de l'utilisabilité, ce dernier critère étant décomposé selon l'efficacité, l'efficience et la satisfaction.

Pour étudier l'acceptabilité comparativement avec les deux périphériques, des intervalles de confiance ont été calculés dans le cas des réponses aux questions fermées, afin de réaliser une inférence sur les proportions observées, en lien avec les hypothèses quant à la supériorité de la Caméra 3D comparativement à la souris. Plusieurs méthodes ont été développées pour calculer l'intervalle de confiance d'une proportion. La méthode standard traditionnelle, qui correspond à l'intervalle de Wald, est la plus couramment utilisée. Elle présente néanmoins l'inconvénient d'être « très anti-conservatrice, surtout lorsque  $p$  (la proportion) est près de 0 ou 1 ou lorsque  $n$  (l'effectif étudié) est petit » (Gagnon, 2006). Elle est donc plus adaptée à des valeurs de  $p$  proches de 0,5. Nous lui préférons donc la méthode de l'intervalle de Wilson, aussi appelée l'intervalle de score (Wilson, 1927), qui est considérée comme plus performante (Gagnon, 2006). Dans la suite du document, les intervalles de confiance seront toujours calculés sur la proportion des participants ayant choisi la réponse cam3D afin d'être en accord avec les hypothèses de recherche. Les intervalles de confiance seront illustrés par les données qualitatives associées aux questions et recueillies dans le même questionnaire post-passation. Dans le cas des échelles de Likert, nous avons attribué un score à chaque modalité. Nous avons alors comparé les moyennes entre les périphériques « Cam3D » et « souris » avec le test de Wilcoxon des rangs signés. La distribution des notations pour chaque variable mesurée a également été étudiée de manière qualitative avec des exemples provenant des questionnaires post-passation.

Pour étudier l'effet du sexe et de l'expertise en Réalité Virtuelle sur l'acceptabilité, nous avons eu recours, à deux sortes de tests. Pour l'analyse des réponses aux questions fermées, nous avons utilisé un test du Khi-deux de Pearson. Lorsque les effectifs théoriques étaient inférieurs à 5 (d'après les règles en usage) et pour des facteurs à deux modalités, nous avons utilisé le test exact de Fisher. Pour la réponse aux échelles de Likert permettant la comparaison des deux périphériques, nous avons réalisé une analyse de la variance (ANOVA) à un plan mixte, bien que la distribution de nos mesures de performances s'écarte de la normalité. Cela étant, nous nous appuyons sur la remarque de (Winer, 1971) quant à la robustesse de l'ANOVA aux erreurs de type 1.

Dans un premier temps, nous allons comparer l'acceptabilité perçue par les participants pour chaque périphérique. Nous étudierons ensuite l'influence du sexe et de l'expertise en Réalité Virtuelle sur l'acceptabilité.

#### 4.1.2 Influence du périphérique sur le ressenti de l'acceptabilité

Les proportions en faveur de la Cam 3D et de la souris, ainsi que les réponses aux échelles de Likert, sont synthétisées dans le Tableau 12. Elles seront ensuite décrites et discutées dans les parties suivantes.

Tableau 12 : Proportions en faveur de la Cam 3D et de la souris, ainsi que les réponses aux échelles de Likert, selon chaque critère

Critère	Question	Proportion de participants en faveur de		Moyenne échelle de Likert (sur 6)	
		Cam3D	Souris		
Utilité	Périphérique perçu comme le plus approprié pour la tâche de sélection de cibles	85,7 %	14,3 %		
	La Cam3D est perçue comme représentant un apport par rapport à la souris			5,06	
Utilisabilité	Efficacité	Périphérique perçu comme le plus efficace en termes de rapidité de positionnement	75,7 %	24,3 %	
		Périphérique perçu comme le plus efficace en termes de précision de positionnement	41,4 %	58,6 %	
		Périphérique perçu comme le plus efficace de façon globale	66,2 %	33,8 %	
	Satisfaction	Périphérique perçu comme le plus confortable	32,4 %	67,6 %	
		Périphérique avec lequel les participants se sentent le plus à l'aise	47,8 %	52,3 %	
	Efficience	Fatigue perçue avec la souris			1,32
		Fatigue perçue avec la Cam3D			2,58
Périphérique perçu comme le plus facile d'utilisation		59,2 %	40,8 %		
Immersion	Périphérique perçu comme le plus immersif	98,6 %	1,4 %		

#### 4.1.2.1 Perception de l'utilité

Le périphérique « Cam3D » est perçu comme étant globalement plus utile que la souris pour la tâche à effectuer.

D'une part, il est plus approprié pour la sélection de cibles, selon l'intervalle de confiance calculé (75,6% - 92%). Les données recueillies nous permettent de justifier qualitativement ce constat, la « Cam3D » étant jugée « *plus facile et plus rapide* » et permet un « *mouvement en parfaite cohérence avec la tâche à réaliser* ». En effet, pour « *saisir des objets, le schème préhension par la main est le plus naturel et le plus réaliste* ».

D'autre part, la « Cam3D » est associée à un apport élevé comparativement à la souris (note moyenne = 5,06 – E.T. = 0,886). L'analyse de la distribution des notations (Figure 32) montre par ailleurs que la majorité des participants ont jugé l'apport « *élevé* » (30/70 soit 42%) et « *très élevé* » (24/70 soit 34 %).

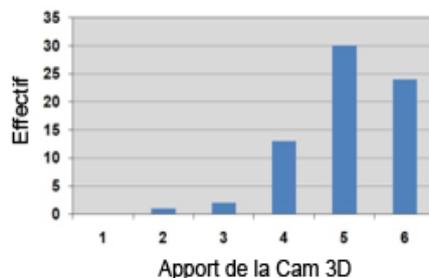


Figure 32 : Distribution des notes jugeant l'apport de la Cam3D par rapport à la souris

#### 4.1.2.2 Perception de l'utilisabilité

L'utilisabilité est étudiée sous l'angle de l'efficacité, de l'efficience et de la satisfaction. D'après la norme ISO 9241-11, qui donne les lignes directrices concernant l'utilisabilité, « *un système est utilisable lorsqu'il permet à l'utilisateur de réaliser sa tâche avec efficacité, efficience et satisfaction dans le contexte d'utilisation spécifié* ». Les tests d'utilisabilité doivent donc couvrir les trois domaines suivants :

- L'efficacité : Vérifier que les tâches données à l'utilisateur soient réalisées ;
- L'efficience : Mesurer les ressources nécessaires pour réaliser ces tâches et la facilité d'apprentissage de l'interface ;
- La satisfaction : Déterminer si le système est agréable à utiliser ;

Le périphérique « Cam3D » est perçu comme étant plus efficace que la souris pour la rapidité de positionnement (intervalle de confiance 64,5% - 84,2%) et pour l'efficacité globale (intervalle de confiance 54,4% - 76,3%). Pour la dernière question concernant la précision de positionnement, il n'est pas possible d'affirmer la supériorité de l'un ou l'autre périphérique (intervalle de confiance 30,6% - 53,1% pour la Cam3D). Les participants justifient la supériorité de la Cam3D pour la rapidité de positionnement par « *la réaction en temps réel aux variations de mouvements [qui permet de] gérer plus facilement les mouvements de la main* » et par « *un déplacement en profondeur plus facile et rapide qu'avec la souris* ». Pour l'efficacité globale, c'est parce que « *d'un point de vue technique, le fait que la Cam3D puisse suivre nos mouvements à la fois dans le plan et en profondeur, facilite la tâche et optimise l'efficacité de celle-ci* ».

Le périphérique « Cam3D » n'est pas perçu comme étant plus satisfaisant que la souris. En effet, en termes de confort, la Cam3D est jugée inférieure à la souris (intervalle de confiance 22,7% - 43,9%). Pour le sentiment d'aisance, les participants sont partagés (intervalle de confiance 36,4% - 59,4% pour la Cam3D). Si la Cam3D est jugée « *intuitive, rapide, performante, naturelle, réaliste, immersive, agréable, directe et sans intermédiaire* », les participants trouvent qu'il est « *plus fatigant de garder son bras tendu en l'air que de le laisser posé sur la table* ».

Enfin, la supériorité de la Cam3D en termes d'efficience n'est que partiellement établie. Celle-ci est majoritairement perçue comme étant plus simple d'utilisation que la souris (intervalle de confiance 47,6% - 69,9%), car « *les interactions sont plus naturelles* » et « *le déplacement est beaucoup plus intuitif qu'avec la souris* ». Par contre, la Cam3D est perçue comme étant plus fatigante, selon le test de Wilcoxon pour échantillons appariés sur les scores obtenus avec les échelles de Likert ( $Z = 5,486$  ;  $p < 0.005$ ). Ce résultat est confirmé par la distribution des notes concernant la fatigue perçue avec chaque périphérique (cf. Figure 33).

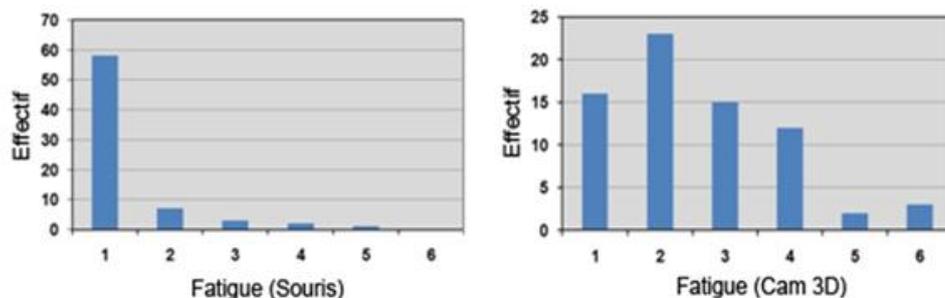


Figure 33 : Distribution des notes concernant la fatigue perçue avec chaque périphérique  
« 1 » représente une absence de fatigue, « 6 » une fatigue élevée

### 4.1.2.3 Perception de l'immersion

La cam3D est perçue comme étant globalement plus immersive que la souris par 70 des 71 participants à l'expérimentation (intervalle de confiance 92,4% - 99,8%). Les données recueillies nous permettent également de justifier qualitativement ce constat. Les participants ayant jugé la cam3D plus immersive que la souris l'ont justifié par le fait que « *la perspective associée au système Cam3D donne la sensation d'atteindre la limite du décor en avançant la main* » ainsi que par le fait que « *de n'avoir aucun périphérique aide à se concentrer davantage sur la représentation visuelle* ».

### 4.1.2.4 Discussion

De manière globale, l'acceptabilité est meilleure avec la Cam3D, ce qui est partiellement en accord avec l'hypothèse HO2. C'est indéniable en termes de perception de l'utilité et du sentiment d'immersion. Concernant, l'utilisabilité, la supériorité de la Cam3D est moins évidente, tout du moins sur les critères de l'efficacité et de la satisfaction. Le principal reproche est la fatigue et l'absence de confort, dues à la position de l'utilisateur, qui est assis et doit tendre le bras, sans support pour le reposer. Dans le cas de la comparaison avec la souris, la différence est effectivement très importante. Cet écueil sera évidemment identique avec d'autres moyens de capture, comme par exemple un gant de données ou des marqueurs permettant un tracking via des caméras infrarouges. Il faut néanmoins garder à l'esprit l'utilisation potentielle – et visée – de notre système dans des environnements immersifs à l'échelle 1, de type « CAVE », dans lesquels l'utilisateur est debout. Il est alors plus facile d'adopter une posture plus reposante, avec les bras partiellement en repos au niveau du buste. L'avantage du meilleur sentiment d'immersion permis avec notre système prend alors tout son sens. Notre tâche, pourtant simple, de sélection d'un objet dans un environnement en trois dimensions, souligne les faiblesses de la souris en terme de rapidité et de simplicité de positionnement et confirme l'intérêt de la conception ou de l'utilisation de périphériques adaptés aux tâches en environnement virtuels.

## 4.2 Interaction entre le sexe et le périphérique

Il n'existe pas de différence selon le sexe pour la perception de l'utilité, d'une part quant à l'appréciation du périphérique le plus approprié pour la tâche de sélection de cibles (test exact de Fisher  $p = 1,00$ ) ni vis-à-vis du jugement de l'apport de la cam3D<sup>6</sup> (Mann-Whitney ;  $U = 457,5$  ; NS), comme cela est visible sur la Figure 34.

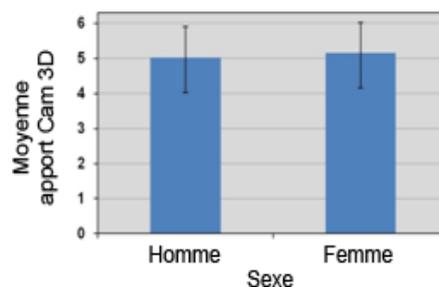


Figure 34 : Évaluation de l'apport de la Cam3D par les hommes et les femmes

De manière générale, le sexe des participants influence peu la perception de l'utilisabilité. Ainsi, nous ne notons pas de différence significative entre les hommes et les femmes pour l'efficacité en termes de rapidité ( $\chi^2 = 0,04$  ;  $p = 0,951$ ) et de précision de positionnement ( $\chi^2 = 0,137$  ;  $p = 0,711$ ), pour la satisfaction en termes de confort d'utilisation ( $\chi^2 = 3,156$  ;  $p = 0,076$ ) et de sentiment d'aisance

<sup>6</sup> hommes : (moyenne = 5,02 ; E.T. = 0,892), femmes : (moyenne = 5,15 ; E.T. = 0,875)

( $\chi^2 = 2,398$  ;  $p = 0,121$ ). Par contre, nous pouvons noter une différence significative pour l'efficacité, c'est-à-dire pour la facilité d'utilisation ( $\chi^2 = 5,864$  ;  $p = 0,015$ ). L'ANOVA réalisée sur l'appréciation de la fatigue en fonction du périphérique et du sexe des participants montre une interaction significative entre ces deux facteurs ( $F(1,69) = 7,965$  ;  $p = 0,006$ ). Nous notons, par conséquent, qu'il y a une influence du sexe sur la fatigue ressentie avec la cam3D ( $U = 369$  ;  $p < 0,043$ ), contrairement à la souris ( $U = 472,5$  ; NS). Les hommes (moyenne = 2,78 ; E.T. = 1,345) ont perçu une plus grande fatigue que les femmes (moyenne = 2,10 ; E.T. = 1,136) avec la caméra 3D. Ces résultats sont présentés sur la Figure 35.

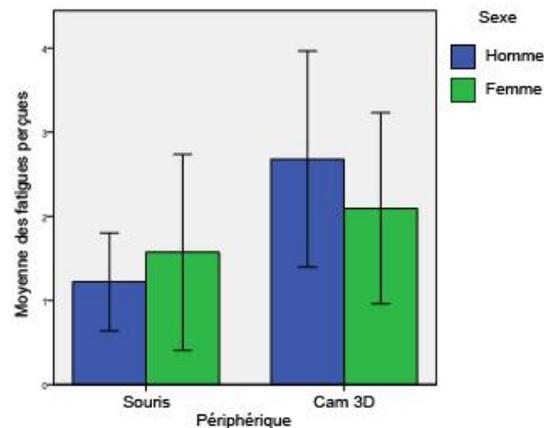


Figure 35 : Fatigue perçue pour chaque périphérique, selon le sexe des participants

Enfin, selon le test exact de Fisher, il n'y a pas de différence significative entre les hommes et à les femmes quant à la perception du périphérique le plus immersif ( $p = 0,286$ ). De façon globale, ces résultats confirment l'hypothèse HO4 avec toutefois une réserve émise sur l'utilisabilité.

#### 4.3 Interaction entre l'expertise en R.V. et le périphérique

Il n'existe pas de différence significative selon le niveau d'expertise en RV quant à l'appréciation du périphérique le plus approprié pour la tâche de sélection de cibles ( $\chi^2 = 0,096$  ;  $p = 0,953$ ), ni pour le jugement de l'apport de la cam3D (Kruskal-Wallis ;  $H = 1,993$  ; NS).

L'expertise en RV des participants n'influence pas la perception de l'utilisabilité. Concernant l'efficacité, il n'y a pas de différence significative entre les niveaux d'expertise en RV que ce soit dans l'appréciation du périphérique le plus efficace en termes de vitesse de positionnement ( $\chi^2 = 0,350$  ;  $p = 0,839$ ), dans l'appréciation du périphérique le plus efficace en termes de précision de positionnement ( $\chi^2 = 0,651$  ;  $p = 0,722$ ) ou encore dans l'appréciation du périphérique jugé globalement le plus efficace ( $\chi^2 = 1,459$  ;  $p = 0,482$ ). Il en est de même pour la perception de la satisfaction, pour le confort ( $\chi^2 = 1,312$  ;  $p = 0,519$ ) et le sentiment d'aisance ( $\chi^2 = 0,673$  ;  $p = 0,714$ ). Enfin, pour l'efficacité, il n'y a pas de différence significative selon le niveau d'expertise en RV quant à l'appréciation du périphérique le plus facile d'utilisation ( $\chi^2 = 0,428$  ;  $p = 0,807$ ). Une ANOVA ne permet pas de mettre en évidence une interaction significative entre la fatigue ressentie pour chacun des périphériques et l'expertise en RV des participants ( $F(2,68) = 0,946$  ; NS).

Il n'existe pas d'effets significatifs de l'expertise en RV sur l'acceptabilité de l'artefact évalué, ce qui confirme l'hypothèse HO7.

## 4.4 Synthèse des résultats

Nous avons reproduit le schéma d'étude mis en œuvre pour l'analyse des performances, à savoir une étude comparative entre la Cam3D et la souris, puis une étude fine de l'influence du sexe et de l'expertise en Réalité Virtuelle sur ces résultats. D'après les participants, la Cam3D apporte un avantage significatif en termes d'utilité perçue et d'immersion ressentie, mais reste inférieure à la souris en termes d'efficacité et de satisfaction. La principale raison est une fatigue jugée plus importante, cela étant inhérent à la position assise associée au maintien du bras tendu. S'il est évident que ce dernier point est critique, il est davantage lié à la tâche de sélection en environnement 3D qu'à l'usage strict du périphérique. Il sera nécessaire de trouver une solution pour permettre un repos partiel du bras sans pour autant nuire au réalisme et à l'immersion. Enfin, il convient de noter que la fatigue pourrait être atténuée pour l'interaction dans un CAVE, où l'utilisateur se tient debout. Dans ce cas, la Cam 3D peut favorablement remplacer un périphérique à tenir en main (e.g. *ART Flystick*).

Le sexe et l'expertise en Réalité Virtuelle sont des facteurs qui n'influencent pas l'acceptabilité, ce qui est en accord avec les hypothèses HO4 et HO7, mis à part la fatigue qui est jugée supérieure avec la Cam3D pour les hommes.

Dans les questionnaires, les participants n'ont pas fait état de difficultés quant à l'utilisation de notre système. Le fait de ne pas recourir à des métaphores d'interaction représente un intérêt. L'utilisation d'une technique de suivi de la main sans marqueurs ne nous permet pas de recourir à des techniques de retour d'effort : hormis le retour visuel sur la position de l'avatar de leur main dans l'espace, les utilisateurs ne disposent pas d'autres indices proprioceptifs. Pourtant, les participants n'ont pas porté de jugements sur l'absence de retour sur les collisions avec d'autres cubes lors de la sélection. Ces résultats démontrent d'une part la valeur que peut apporter notre système pour des tâches sur écran d'ordinateur où le sentiment d'immersion reste habituellement faible et où le positionnement dans l'espace peut être problématique et d'autre part l'intérêt que pourrait avoir le portage de notre système dans un environnement immersif stéréoscopique à l'échelle 1.

## 5 Conclusions générales

L'objectif de cette étude est de démontrer, par l'intermédiaire d'une tâche simple de sélection en environnement virtuel en trois dimensions, l'intérêt de notre système par rapport à un périphérique utilisé couramment dans ce type de tâche. Nous avons pour cela mené une expérimentation auprès de 71 participants (50 hommes, 21 femmes), la tâche consistant à sélectionner successivement 10 cubes dans un environnement borné, alternativement avec la souris et avec une caméra 3D. Nous avons formulé l'hypothèse générale que la caméra 3D (Cam3D) apportait un avantage significatif, vis-à-vis d'un périphérique courant comme la souris, en termes d'acceptabilité, tout en permettant des performances au moins similaires et cela indépendamment de la taille de la main de l'utilisateur, de son sexe ou de son expérience en réalité virtuelle. Cette hypothèse a donc guidé nos choix pour les analyses quantitative et qualitative. La performance s'appuie sur la mesure du temps d'achèvement de la tâche entière de sélection. L'analyse qualitative repose sur l'étude de l'acceptabilité, et plus précisément de l'utilité, de l'utilisabilité et du sentiment d'immersion, ces données étant recueillies grâce à un questionnaire post-passation. Différents tests statistiques ont été menés pour assurer la validité et la reproductibilité des résultats.

Les performances avec la Cam3D ne sont pas sensiblement meilleures que celles obtenues avec la souris. Ce résultat est à mettre en perspective avec le temps d'expérimentation très court, qui constitue

un désavantage pour la Cam3D, la souris bénéficiant d'un niveau de pratique supérieur, malgré la phase de prise en main au début de l'expérimentation. Ce premier résultat confirme notre hypothèse. Le sexe n'influence pas les performances obtenues avec la cam3D, contrairement à l'expertise en réalité virtuelle et à la taille de la main. Nos résultats indiquent que les personnes expertes et initiées obtiennent de meilleurs résultats que les personnes novices. De la même manière, les personnes avec des mains qualifiées de grande taille ou de taille moyenne sont plus rapides à achever la tâche que les personnes avec de petites mains. La perception de l'acceptabilité est divisée : la Cam3D est jugée plus utile et apporte un meilleur sentiment d'immersion mais l'utilisabilité est par contre jugée moindre en raison de la fatigue ressentie comme supérieure avec la Cam3D. C'est la position de l'utilisateur, assis et bras tendu sans point d'appui qui justifie ce jugement. Le sexe ou l'expertise en réalité virtuelle ne sont pas des critères de nature à influencer l'acceptabilité perçue par les participants.

Pris conjointement, ces résultats quantitatifs et qualitatifs démontrent clairement l'intérêt d'un périphérique sans marqueur pour les tâches de sélection en environnement virtuel. En effet, si les performances sont équivalentes, le sentiment d'immersion est supérieur avec notre système. Le schème d'interaction, sans métaphore intermédiaire, apporte une compréhension immédiate à l'utilisateur et par là même une grande simplicité d'utilisation. Ces différents résultats nous mettent face à un paradoxe : un temps d'expérimentation et donc d'utilisation plus important permettrait de limiter le biais du temps d'apprentissage et pourrait impacter positivement les performances, cependant, il dégraderait dans le même temps l'utilisabilité perçue et plus précisément occasionnerait une grande fatigue (elle est déjà importante pour une durée d'utilisation de moins de 5 minutes). Même si l'intérêt de notre système est avéré, ce dernier serait plus approprié dans deux scénarii d'usages différents permettant de diminuer la fatigue perçue. D'une part, dans le cadre d'une utilisation debout dans un environnement immersif en vision 3D stéréoscopique : l'utilisateur ne serait pas obligé de travailler le bras tendu et pourrait reposer son bras sur le haut du corps. D'autre part, dans le cadre d'une utilisation assise, pour des gestes épisodiques, comme par exemple dans le cas de la commande d'appareils à distance (domotique).

Notre hypothèse générale est donc vérifiée. Certains résultats restent néanmoins difficiles à expliquer : il est curieux de constater une différence de performances relative à la taille de la main des participants. Tout du moins, le résultat observé est contraire à ce que nous aurions pu attendre. En effet, les utilisateurs avec des mains de grande taille, en sortant partiellement de l'espace de jeu, auraient pu entraîner des erreurs de détection de la position de la main et donc un temps d'exécution plus long. Pourtant, les performances sont inférieures avec des personnes avec de petites mains. Les meilleures performances obtenues par les utilisateurs familiers des applications de réalité virtuelle, si elles contredisent nos hypothèses, peuvent s'expliquer par une meilleure propension de ce type d'utilisateurs à s'adapter à de nouveaux systèmes d'interaction. Il convient de souligner que, pour les novices, les résultats sont meilleurs avec la Cam3D qu'avec la souris. Ce résultat est intéressant dans le cadre de l'accès au plus grand nombre aux nouvelles technologies.

Après cette première étude, s'appuyant sur une tâche simple, nous pouvons donc dire que la capture de mouvements sans marqueurs semble prometteuse car elle apporte des avantages indéniables dans le cadre de l'interaction en environnement virtuel, où le sentiment d'immersion est un moteur important du réalisme et de la présence ressentie (Loup-Escande, 2010). Ce travail est un premier pas vers la démonstration scientifique des apports potentiels d'une telle technologie. Il constitue un point de départ et apporte donc une justification à la poursuite des travaux de recherche portant sur d'autres types d'activités 3D temps réel, sur des expérimentations plus longues, et également comparant notre système avec des périphériques plus évolués.

## Chapitre 4

# Gant de données vs. Cam3D pour la manipulation et la navigation en environnement virtuel : étude 2

---



# 1 Méthodologie

## 1.1 Présentation générale

La première étude nous a permis de mettre en évidence les avantages et les inconvénients d'un système de capture de mouvements sans marqueurs pour une tâche simple et courante (i.e., la sélection d'objets). Nous avons décidé de continuer dans cette voie pour répondre à notre problématique générale, en évaluant des tâches plus complexes et en améliorant notre système.

Nous avons, pour l'étude précédente, utilisé une caméra time-of-flight de type « *SwissRanger SR 4000* » dont la limite principale était la résolution, ce qui avait pour conséquence principale de limiter l'amplitude des mouvements de l'utilisateur. Cette caméra avait été choisie car son framerate était de 60FPS et sa résolution était l'une des meilleures par rapport à ses concurrentes lors de nos développements et de l'expérimentation. Lors de l'analyse des résultats de notre première étude, un nouveau périphérique est apparu sur le marché, destiné en premier lieu aux applications ludiques (les jeux vidéo sur *Xbox 360* de *Microsoft*). Ce périphérique, nommé Kinect, est constitué d'une caméra couleur de résolution 640 x 480, de microphones, d'accéléromètres et surtout d'une caméra 3D de résolution 640x480 fonctionnant à 30 images par seconde. Étant données les caractéristiques de cette caméra, et la naissance d'une communauté de développeurs, nous avons décidé d'avoir recours à cette caméra pour l'étude présentée dans ce chapitre.

Très rapidement, l'utilisation de la Kinect a été détournée par des particuliers afin qu'elle puisse fonctionner sur ordinateur. Voyant les possibilités offertes par ce périphérique dans des domaines autres que le ludique, ainsi que le fort intérêt manifesté par les communautés de programmeurs et de chercheurs pour ce nouveau périphérique, la société qui fabrique ces caméras a décidé de mettre en accès libre les pilotes ainsi que l'algorithme permettant de capturer les mouvements du corps. Pour cela, un squelette est appliqué sur chaque utilisateur et il est possible de « tracker » jusqu'à 16 personnes simultanément.

Nous avons donc décidé d'utiliser la Kinect pour améliorer le système basé initialement sur la caméra SR 4000. Les développements qui ont été menés permettent désormais au système de capturer les mouvements de tous les membres de l'utilisateur ; nous avons par ailleurs limité le nombre de personnes pouvant travailler simultanément à 2 car au-delà la résolution ne permet pas à trois personnes de travailler côte à côte sans se gêner. Par ailleurs, notre système permet de détecter si chacune des mains est ouverte ou fermée. Pour cette deuxième étude, nous avons décidé de ne faire manipuler qu'une seule personne à la fois, l'aspect collaboratif n'étant pas traité dans le cadre de nos travaux. Il sera néanmoins possible dans des travaux futurs.

L'objectif de cette seconde étude est d'aller plus loin dans l'interaction avec l'environnement virtuel. En effet, si la première étude permettait à l'utilisateur de réaliser des tâches de sélection dans un environnement virtuel en 3 dimensions, cette seconde étude s'intéresse en plus aux actions de désélection, déplacement, rotation et redimensionnement d'objets ainsi qu'à la navigation en environnement virtuel. Ainsi, nous aurons couvert l'ensemble des actions courantes de manipulation en environnement virtuel.

Les interactions sont réalisées avec une main ou de manière bi-manuelle, selon l'action choisie et sont réalisées avec un taux de rafraîchissement de 30 FPS, rendant possible une interaction 3D temps réel en environnement virtuel tels qu'expliqués par (Bowman et al., 2004; Coquillart et al., 2003).

Cette expérimentation vise à montrer d'une part que le système développé avec la Kinect permet d'obtenir des résultats *a minima* équivalents aux systèmes couramment utilisés en réalité virtuelle, comme des capteurs magnétiques (en ce qui concerne la position dans l'espace des mains) et des gants de données (pour détecter l'ouverture ou la fermeture des mains), tout en améliorant le confort, l'efficacité et l'efficacité ressentis par les utilisateurs. D'autre part, nous nous attacherons à démontrer que les modalités d'interaction développées dans notre système sont efficaces, intuitives et adaptées aux tâches de déplacement, de redimensionnement, de rotation et de navigation demandées à l'utilisateur.

## 1.2 Terminologie

Voici les différents acronymes et termes que nous allons utiliser dans la suite de ce chapitre :

- Environnement virtuel : environnement 3D dans lequel les actions sont effectuées ;
- Système Cam3D : système utilisant une caméra 3D Kinect comme interface transparente pour récupérer les actions et les mouvements effectués par l'utilisateur ;
- Système Gants3D : système utilisant 2 gants 5DT et 2 capteurs électromagnétiques pour récupérer les actions et les mouvements effectués par l'utilisateur ;
- Main dominante : main de prédilection de l'utilisateur. La main droite sera la main dominante pour une personne droitère ;
- Main secondaire : main qui n'est pas la main dominante de l'utilisateur. La main gauche sera la main secondaire pour une personne droitère.

Dans la suite de ce chapitre, afin de simplifier la lecture, nous remplacerons environnement virtuel par **EV**, système Cam3D par **Cam3D**, système Gants + Capteurs par **Gants3D**, tâche 1 par **T1**, tâche 2 par **T2** et ainsi de suite jusqu'à **T9** (pour les 9 tâches à réaliser dans l'expérimentation).

## 1.3 Hypothèses

Nous avons démontré dans l'étude précédente qu'un système de capture sans marqueur pouvait se subsister à la souris pour des tâches de sélection simple, sans entraîner de baisse de performances et avec un apport en termes de préférences subjectives. Pour des tâches de manipulation en environnement 3D, la souris n'est plus suffisante : nous avons donc choisi comme matériel de référence les composants les plus courants, à savoir le gant de données associé à un système de tracking électromagnétique. Si l'avantage en terme de coût est indéniable (environ un facteur 40) grâce à la large diffusion de la Kinect entraînant un prix raisonnable, il reste à prouver que notre système « Cam3D » permettra d'obtenir des performances aux moins équivalentes au système « Gants3D » pour la réalisation des tâches de manipulation (déplacement, rotation, redimensionnement, navigation). Les deux systèmes sont fonctionnellement équivalents et sont donc évalués sur les mêmes modalités d'interaction, que nous avons par ailleurs développées. L'objectif scientifique premier de cette étude est donc de démontrer l'intérêt de notre système par rapport au matériel courant utilisé en réalité virtuelle, en termes de performances (temps d'exécution, précision), d'acceptabilité (efficacité, confort, efficacité) et de préférences subjectives.

L'objectif secondaire est l'évaluation des modalités d'interaction proposées.

La finalité, si l'intérêt de notre système est démontré, et en fonction de l'évaluation des modalités d'interaction, est l'amélioration de ces dernières (voir le chapitre 5 p. 143) afin d'être en mesure de

fournir un système complet (périphérique et modalités d'interaction) qui soit directement utilisable dans les applications de réalité virtuelle.

Nos hypothèses visent à répondre au premier objectif. Pour y répondre, nous avons mené une étude comparative de deux systèmes (i.e., Gants3D et Cam3D) pour des tâches de manipulation en environnement virtuel. Nous avons formulé 4 hypothèses de travail :

- H1 : le système Cam3D permet d'obtenir une meilleure précision dans les tâches de manipulation que les Gants3D, du fait de l'absence de port de matériel, donc par l'absence d'entrave des mouvements et de l'absence de poids de l'équipement ;
- H2 : le système Cam3D permet d'obtenir des temps d'exécution au moins similaires au système Gants3D, du fait d'une détection des mouvements aussi précise et aussi fiable ;
- H3 : le système Cam3D doit recueillir un meilleur niveau d'acceptabilité globale que les Gants3D, en raison d'un meilleur confort (pas d'équipement, pas de poids à porter/supporter), d'une meilleure efficacité (majoritairement en termes de précision et de stabilité ressentie) et d'une meilleure efficacité ;
- H4 : les préférences des participants sont en faveur de la Cam3D, du fait du plus haut degré d'immersion supposé, ainsi que d'une plus grande aisance d'utilisation.

Ces hypothèses, si elles sont confirmées, permettront de conclure sur la supériorité de la Cam3D sur les Gants3D pour les tâches de manipulation en environnement virtuel rendues possibles par les modalités d'interaction que nous avons développées.

#### **1.4 Tâches de manipulation et de navigation**

On entend par tâches de manipulation l'ensemble des tâches pour lesquelles l'utilisateur doit effectuer des actions sur un objet, celles-ci étant de trois types :

- Le déplacement, c'est-à-dire le positionnement dans l'environnement virtuel d'un objet par translation selon les axes X, Y et Z ;
- La rotation, c'est-à-dire l'orientation de l'objet selon les axes X, Y et Z ;
- Le redimensionnement, c'est-à-dire la modification de l'échelle d'un objet. L'utilisateur peut soit agrandir, soit rétrécir l'objet selon les 3 axes à la fois. De cette façon l'objet n'est pas déformé mais garde bien ses proportions initiales.

On entend par tâches de navigation les tâches pour lesquelles l'utilisateur se déplace dans l'environnement virtuel. Afin d'imposer la navigation, nous avons décidé de concevoir plusieurs scènes entre lesquelles le participant devra « se déplacer ». La navigation s'effectue par translation de l'environnement virtuel selon les axes X, Y et Z.

Nous avons décidé d'adopter une complexité graduelle pour les tâches de manipulation, afin de pouvoir réaliser des mesures plus fines. Dans un premier temps, les participants doivent réaliser, dans la première scène, chaque type de tâche de manière isolée sur des objets distincts. Puis, sur la deuxième scène, nous composons deux types de tâche (e.g., déplacement et rotation) et enfin les trois types de tâches (i.e., rotation et déplacement et redimensionnement) doivent être réalisés dans la troisième scène. Ce choix permettra également au participant une meilleure appropriation des modalités d'interaction proposées. Nous avons, de fait, réalisé 3 scènes différentes entre lesquelles le participant doit naviguer. Les scènes et les objets sont différents afin de rendre l'expérimentation plus attractive et motivante pour le participant.

## 1.5 Environnements virtuels et tâches à réaliser : conception

Chaque scène contient des objets à manipuler qu'il faut faire correspondre au mieux à un modèle. Pour chaque objet, nous avons donc reproduit le même mécanisme : afin de montrer à l'utilisateur où positionner, comment orienter et comment redimensionner les objets, des représentations fantômes (en transparence) des objets sont affichées. L'utilisateur doit alors faire correspondre au mieux les objets à manipuler avec leurs représentations sous forme de « ghost » et peut également s'y reprendre autant de fois qu'il le souhaite pour accomplir la tâche.

La première scène ne contient que des objets associés à des tâches simples (i.e. ne requérant qu'une seule action), elle correspond à un environnement « campagne » :

- Tâche 1 (T1) : déplacement d'un puits ;
- Tâche 2 (T2) : redimensionnement d'un cheval ;
- Tâche 3 (T3) : orientation d'une maisonnette.

La Figure 36 représente la scène initiale à gauche (avec les représentations fantômes et les objets à manipuler avec leur position / orientation / taille de départ) et la scène finale, à droite, avec chaque objet correctement positionné / orienté / redimensionné. Il n'est possible d'appliquer à un objet que la manipulation à laquelle il est associé (e.g., il n'est pas possible d'orienter le cheval).

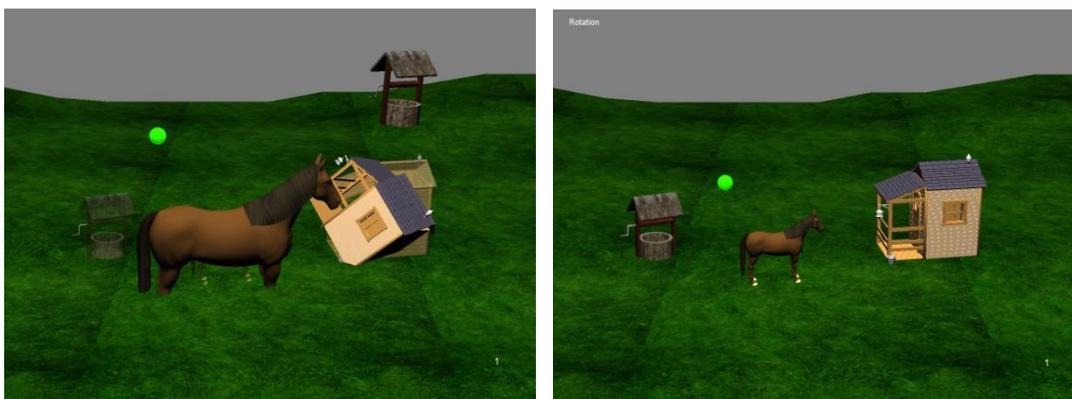


Figure 36 : Scène 1 avec les 3 objets à manipuler (puits, cheval et maisonnette) avant et après manipulation

La scène 2 correspond à l'environnement « bureau ». Elle contient également 3 objets à manipuler, mais cette fois avec des actions composées de deux types d'actions élémentaires :

- Tâche 5 (T5) : déplacement et redimensionnement d'un globe ;
- Tâche 6 (T6) : déplacement et orientation d'une horloge ;
- Tâche 7 (T7) : orientation et redimensionnement d'un ordinateur portable.

Là encore, chaque objet est restreint aux actions que doit lui appliquer l'utilisateur : par exemple l'ordinateur portable ne peut pas être déplacé, il peut uniquement être redimensionné et orienté. La Figure 37 représente la scène initiale (à gauche) et la scène finale (à droite).

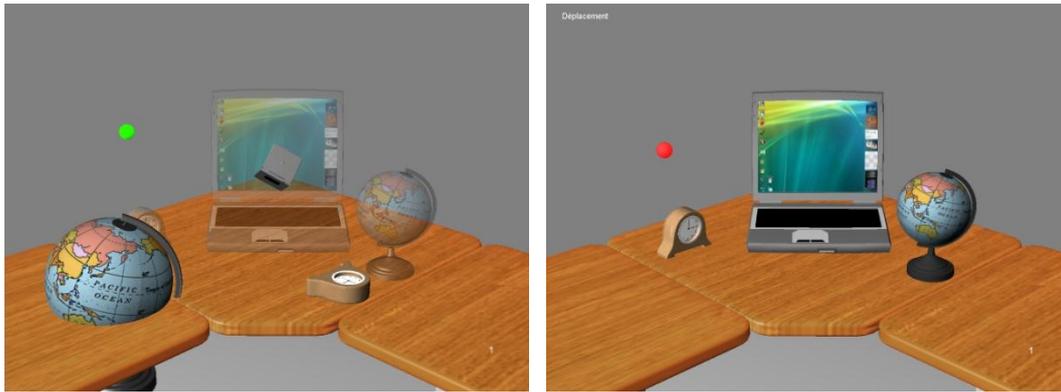


Figure 37 : Scène 2 avec les 3 objets à manipuler (globe, horloge et ordinateur) avant et après manipulation

La dernière scène, « mer », ne contient qu'un seul objet, un bateau qui doit être manipulé en composant les actions de rotation, déplacement et redimensionnement (T9). La Figure 38 représente la scène initiale (à gauche) et la scène finale (à droite).

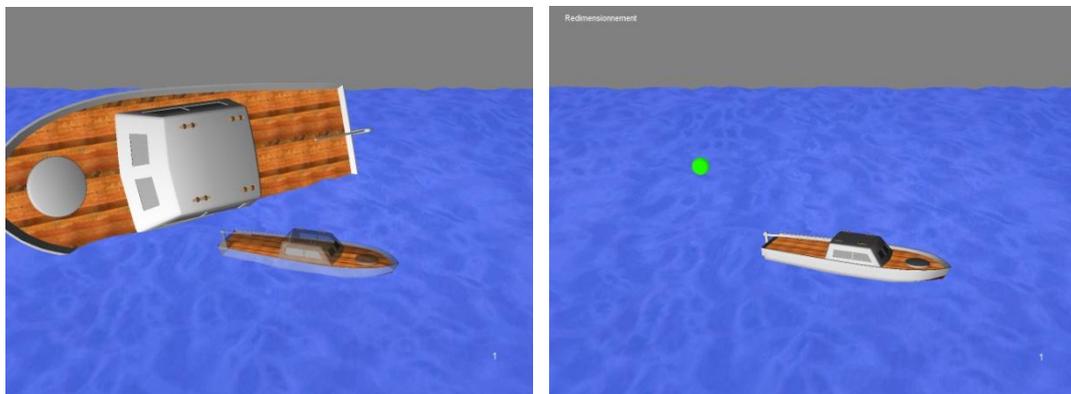


Figure 38 : Scène 3 avec l'objet à manipuler (bateau), avant et après manipulation

Pour passer d'une scène à l'autre, l'utilisateur doit réaliser des actions de navigation (les scènes sont distantes dans l'environnement virtuel).

Le jeu est affiché en perspective, permettant à l'utilisateur de mieux juger des distances et des orientations des objets. La caméra est placée légèrement en hauteur et avec suffisamment de recul pour permettre au participant de voir entièrement la scène dans laquelle il se situe.

La place des objets est fixée au départ et reste identique pour les 2 systèmes à tester. Nous avons décidé de laisser les objets dans les mêmes conditions (position, orientation, dimension) d'un périphérique à l'autre pour obtenir des données comparables entre les systèmes étudiés. Si les participants peuvent affiner leur stratégie entre les 2 conditions (par exemple, utilisation de points de repères), ce n'est pas tant la position, l'orientation ou la taille initiale de l'objet qui rentrera en ligne de compte, que le temps de manipulation et donc d'appropriation. De plus, pour les tâches simples (une seule action), l'utilisateur ne met pas a priori de stratégie pour arriver au but fixé. Dans tous les cas, nous avons mis en place des mécanismes pour assurer la validité interne de notre expérimentation (cf. 2.8 p. 113).

## 1.6 L'utilisation d'un pie-menu adapté à l'environnement

Étant donné que l'utilisateur peut réaliser 3 actions différentes de manipulation, plus une action de navigation, il est nécessaire de rendre possible le changement d'action (e.g., passer de la rotation au déplacement). Pour cela, plusieurs solutions sont envisageables :

- Langage gestuel : étant donné qu'un apprentissage des mouvements est déjà nécessaire pour réaliser les actions de sélection / désélection / manipulation / navigation, nous avons jugé que cette alternative rendrait l'expérimentation complexe et risquerait d'induire une certaine forme de confusion dans l'esprit du participant.
- Commande vocale : même si la reconnaissance vocale est aujourd'hui performante et fiable, elle reste encore peu utilisée. De plus, il nous semble hasardeux d'induire de la multimodalité dans notre expérimentation, l'utilisateur travaillant en interaction bi-manuelle, avec une charge cognitive déjà relativement importante.
- Périphérique externe : l'utilisation d'un périphérique externe (e.g., le clavier) a été proscrite pour éviter une rupture de l'immersion dans la tâche à réaliser.

Nous avons décidé d'afficher des boutons directement à l'écran, et donc de réaliser un menu. Nous n'avons pas jugé opportun de disposer des boutons en permanence dans l'interface, d'une part pour éviter au participant des gestes supplémentaires pour se positionner sur un bouton en particulier, et d'autre part pour ne pas gêner la perception 3D de l'environnement par un plan 2D superposé.

Le choix que nous avons fait a été d'afficher un menu à la demande (qui disparaît une fois effectué le choix de l'action), et à l'endroit où se situe la main de l'utilisateur, pour éviter tout déplacement supplémentaire de la main. La solution d'un « pie-menu » répond à ces différentes contraintes. Des travaux ont attesté l'intérêt de ces menus par rapport aux menus linéaires (Callahan et al., 1988) car ils permettent une rapidité accrue tout en minimisant les erreurs de sélection. De plus, ils permettent de rendre identique la distance entre le point d'activation et les différents items à atteindre, du fait de l'organisation circulaire. Ce type de menu permettra l'adoption éventuelle de techniques avancées de sélection ou mode expert (par exemple les « marking menus » (Kurtenbach & Buxton, 1991)) où seule la gestuelle permet de sélectionner une action, sans affichage du menu.

Nous avons réalisé un menu circulaire composé de 4 secteurs (voir Figure 39). Les 4 secteurs correspondent aux 4 actions possibles (déplacement, redimensionnement, rotation et navigation). Nous avons mis des couleurs différentes afin d'aider le participant à bien différencier les 4 actions.

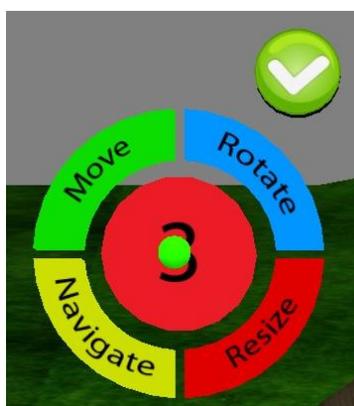


Figure 39 : Menu circulaire avec 4 choix possibles + un bouton validation pour la fin de l'expérimentation

Nous avons également associé au menu une icône de validation (au dessus à droite). Elle n'est utilisée qu'une seule fois pour permettre à l'utilisateur d'indiquer qu'il a terminé l'expérimentation (à l'issue de la scène 3).

Pour faire apparaître le menu, l'utilisateur doit fermer sa main et l'ouvrir rapidement, puis garder la main ouvert immobile pendant 2 secondes. Le menu apparaît alors à l'endroit exact où l'utilisateur a fermé la main. Pour choisir une action, il doit déplacer sa main en la maintenant ouverte vers l'action qu'il souhaite réaliser et rester sur le menu choisi pendant 2 secondes (la main est figurée par un avatar prenant la forme d'une sphère verte). Un compte à rebours s'affiche au centre du menu lui indiquant que son choix est en cours de validation. Tant que le choix n'est pas validé, l'utilisateur peut se déplacer vers un autre menu (par exemple s'il a changé d'avis ou s'il s'est trompé) et le compte à rebours est alors réinitialisé. Le menu reste affiché tant que l'utilisateur n'a pas fermé la main ou qu'une action n'a pas été choisie. Une fois le menu validé, un message en haut à gauche de l'écran affiche l'action que l'utilisateur a choisie et qu'il peut donc effectuer. L'action choisie reste valide tant que l'utilisateur n'en a pas demandé d'autre.

### **1.7 Manipulation : pistes envisagées et solutions retenues**

Dans la vie de tous les jours, nous interagissons avec les objets qui nous entourent à l'aide de nos mains. Si un petit objet peut être manipulé avec une seule main, les objets plus volumineux nécessitent l'usage des 2 mains. Notre système ne permet pas à l'heure actuelle de déterminer l'orientation des mains mais uniquement leur position dans l'espace et leur état (ouverte ou fermée). Par conséquent, il nous a fallu trouver des modalités d'interactions (bi-manuelles pour certaines) permettant de réaliser des rotations, des déplacements et des redimensionnements de manière simple et intuitive pour l'utilisateur.

Nous avons également ajouté un retour visuel pour indiquer au participant la position ainsi que l'état de ses deux mains dans l'environnement virtuel. L'utilisation de modèles réalistes de mains n'a pas été retenue pour 2 raisons :

- Pour la manipulation d'objets réels de grandes dimensions, comme une maison ou un cheval, l'utilisation de modèle de mains n'aurait pas été cohérente car il n'y aurait pas eu adéquation entre la taille des mains et la taille de l'objet. Cela aurait pu être perturbant pour les utilisateurs.
- Pour des tâches de précision, un modèle de main pose deux problèmes. D'une part, l'intersection avec l'objet peut être multiple, d'autre part la taille de modèle peut gêner la vision de l'objet, en masquant ce dernier.

Nous avons donc choisi une métaphore simplificatrice, en représentant les avatars des mains sous forme de petites sphères. La couleur de celles-ci indique l'état de la main :

- Vert : main ouverte
- Rouge : main fermée

Plusieurs actions ne nécessitent que l'usage de la main dominante : la sélection d'objet, la désélection, l'affichage du menu, le déplacement d'objet et la navigation dans l'environnement virtuel. En revanche, les tâches plus complexes comme le redimensionnement et la rotation exigent l'utilisation des 2 mains. Nous allons décrire dans un premier temps les tâches simples, puis les tâches complexes dans un second temps.

### 1.7.1 Tâches simples

Pour la sélection d'objet, l'utilisateur doit déplacer sa main dominante sur l'objet à sélectionner puis fermer sa main pendant 2 secondes. Lorsque l'objet est sélectionné, il devient bleu. La Figure 40 décrit la procédure de sélection, qui a également été donnée sous forme papier aux participants. Les procédures pour les autres modalités sont présentées en Annexe B (Fiches explicatives).

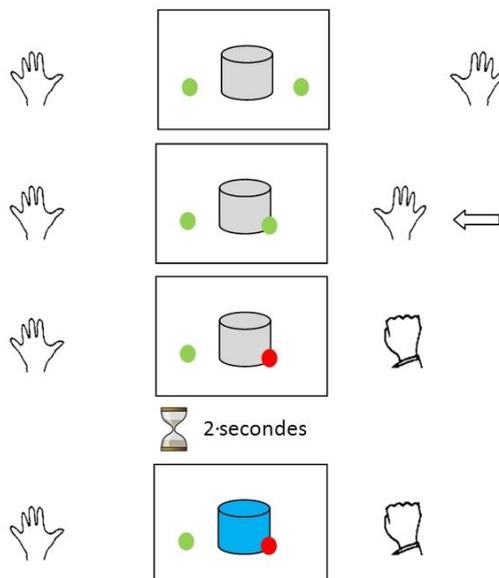


Figure 40 : Procédure de sélection d'un objet (pour un droitier)

Pour désélectionner un objet, l'utilisateur doit déplacer sa main dominante sur l'objet à désélectionner, maintenir sa main ouverte dessus pendant 2 secondes. Lorsque l'objet reprend sa couleur d'origine, il est alors désélectionné.

Pour appeler le menu, l'utilisateur doit fermer la main dominante et l'ouvrir rapidement sans la déplacer. En la maintenant ouverte pendant 3 secondes, le menu apparaît. Pour choisir une action, il doit déplacer sa main (toujours ouverte) sur un des sous-menus pendant 3 secondes. Un compte à rebours apparaît au centre du menu. Une fois arrivé à 0, le sous menu est sélectionné. Un texte en haut à gauche de l'écran indique alors quelle action est actuellement possible.

Enfin, pour déplacer un objet préalablement sélectionné, il faut positionner la main dominante (fermée) dessus. L'objet s'attache à la main et suit alors ses mouvements tant que celle-ci reste fermée. L'utilisateur doit déplacer sa main dominante à l'endroit où il souhaite positionner l'objet. Une fois que ce dernier est correctement positionné, l'utilisateur doit ouvrir la main pour relâcher l'objet.

### 1.7.2 Tâches complexes

On considère que les tâches de rotation et de redimensionnement sont plus complexes car elles demandent une coordination des 2 mains.

La modalité du redimensionnement d'un objet s'apparente dans le monde réel à un ressort duquel on tiendrait une extrémité dans chaque main et que l'on pourrait comprimer ou étirer pour en modifier la taille. Le redimensionnement étant homogène sur les trois axes, la modification de la taille sur l'axe X a pour effet d'appliquer le même facteur sur les axes Y et Z. Une fois l'objet préalablement sélectionné, l'utilisateur doit positionner sur celui-ci sa main dominante fermée et fermer l'autre main pour autoriser le redimensionnement. En écartant ou rapprochant les mains, l'objet s'agrandit ou se réduit. Lorsque l'objet est redimensionné l'utilisateur doit ouvrir les mains pour « verrouiller » la taille de l'objet.

Pour la rotation, une fois l'objet sélectionné, l'utilisateur doit positionner sa main dominante (fermée) sur celui-ci. L'avatar de la main disparaît alors et se positionne automatiquement au centre de l'objet. Il faut fermer l'autre main pour autoriser la rotation. En déplaçant la main secondaire selon l'axe des X, cela permet à l'objet de tourner dans cette même direction. Il en est de même pour les axes Y et Z. Lorsque la rotation est achevée, l'utilisateur doit ouvrir la main dominante d'abord puis la deuxième afin d'éviter de petits décalages lors de l'ouverture des mains.

Le redimensionnement reste assez simple étant donné qu'il suffit de fermer les deux mains et de les rapprocher ou de les éloigner pour agrandir ou rétrécir l'objet. Par contre, la rotation est plus difficile étant donné que la main dominante sert à verrouiller l'objet sur lequel on veut travailler et l'autre main à l'orienter dans l'espace : les 2 mains font donc des mouvements distincts ce qui demande plus de concentration pour l'exécution.

### 1.8 La métaphore du joystick pour la navigation

Une fois le menu « navigation » sélectionné, un joystick apparaît au centre de l'écran. Nous avons choisi la métaphore du joystick pour la navigation car c'est un périphérique dont beaucoup de personnes connaissent le principe de fonctionnement.

Pour naviguer, il faut positionner la main dominante sur la manette du joystick et la fermer pendant 2 secondes afin de sélectionner la manette. Lorsque la manette est sélectionnée, un cadre apparaît ainsi que des flèches pour indiquer à l'utilisateur dans quelle direction il doit aller ainsi que la position finale où il doit se rendre (Figure 41). Un cadre est positionné sur chaque scène, il permet non seulement d'indiquer la position finale mais aussi de s'assurer que l'utilisateur sera bien positionné pour effectuer les manipulations d'objets sans être ni trop près d'eux, ni trop loin et assurer ainsi un confort et une facilité d'utilisation maximum.

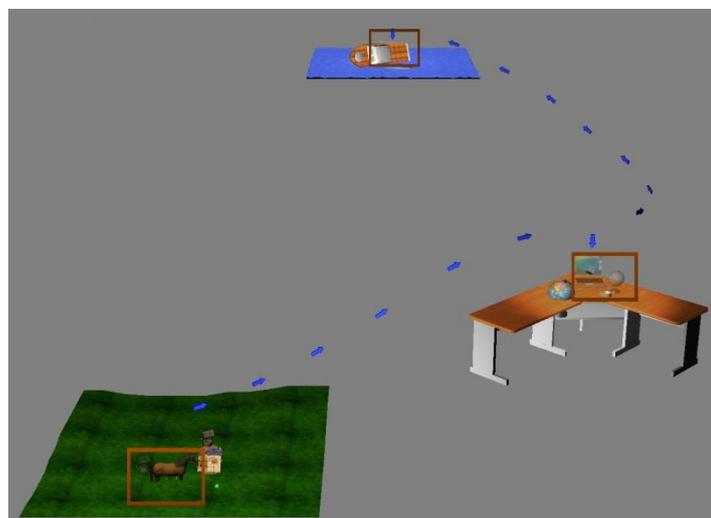


Figure 41 : Navigation entre les 3 scènes. Les flèches bleues indiquent l'itinéraire à suivre, les cadres marrons permettent au participant de se positionner avec précision sur la nouvelle scène

Lorsque les flèches, indiquant la direction à suivre, apparaissent, l'utilisateur déplace sa main (toujours fermée) dans cette même direction. Une fois au bon endroit, il lui suffit d'ouvrir la main pour stopper la navigation. Enfin, l'utilisateur doit maintenir la main ouverte sur la manette pendant 2 secondes afin de la désélectionner.

Lorsque l'utilisateur ferme la main sur la manette, on prend cette position pour référence et le déplacement de la main dans une direction sert de vecteur directionnel pour la navigation. On utilise la distance au cube entre la position de la main et la position de référence pour indiquer la rapidité du déplacement. Ainsi une amplitude de mouvement faible permet un déplacement précis et lent, tandis que des mouvements de grandes amplitudes permettent de se déplacer rapidement d'une scène à l'autre. De plus le déplacement fonctionne de façon incrémentale à la façon d'une « Spacemouse », c'est-à-dire que la navigation ne se fait pas en fonction de la position de la main dans l'espace mais de sa distance par rapport à la position de référence.

## 1.9 Implémentation et développements logiciels

Nous avons utilisé pour cette expérimentation un ordinateur *Alienware* (*Intel*® *Core*™ *i7* CPU930 cadencé à 2.80GHz avec 6 Giga-octets de RAM et une carte graphique de type *NVidia GeForce GTX480*) ainsi qu'un écran *Samsung 120Hz (3D Vision Ready)*. Pour acquérir la position et l'état des mains (ouvertes ou fermées), nous avons utilisé 3 périphériques différents : une *Kinect* de *Microsoft* pour notre système d'une part et 2 gants 5DT ainsi que 2 capteurs magnétiques *Polhemus Patriot* pour le système « Gants3D » d'autre part.

Pour cette expérimentation, nous avons amélioré le système Cam3D précédemment créé. Nous avons essayé de tenir compte des remarques des utilisateurs, aussi bien en termes d'ajout de fonctionnalités qu'en termes d'améliorations de détection. Parmi les suggestions récoltées dans les questionnaires post-passation, certaines idées étaient récurrentes comme par exemple le fait de : « Pouvoir manipuler les objets et non plus seulement les sélectionner » ou encore « effectuer des opérations bimanuelles ». De même, l'algorithme de détection de main fermée / main ouverte a été amélioré.

Comme nous l'avons expliqué précédemment, nous avons délaissé la Caméra 3D *SwissRanger* pour une caméra 3D *Kinect* de *Microsoft*. En effet, bien que le *framerate* ne puisse dépasser les 30 images par seconde, elle offre cependant une résolution nettement supérieure (d'environ 4 fois, 640 x 480). La précision en profondeur reste du même ordre bien que la technologie utilisée soit différente (cf. chapitre 2 partie 4.3.3 p. 50). De plus, ce périphérique possède une seconde caméra couleur de même résolution et a de nombreux capteurs (audio, accéléromètres, etc.). Enfin, elle est motorisée et permet d'orienter la caméra vers le haut ou le bas si nécessaire. Ces nombreux avantages en font un excellent périphérique pour la capture de mouvements (ce qui est logique, étant donné qu'elle a été créée initialement pour cela). De même, les nombreux capteurs pourront être utilisés par la suite pour améliorer le système, par exemple en y ajoutant des commandes vocales. Lorsque la *Kinect* a été vendue en novembre 2010, de nombreux développeurs se sont lancés dans la course pour essayer d'en détourner l'usage et de l'utiliser sur ordinateur. Un concours a même été lancé et une récompense de 1000\$ (initialement) puis 3000\$ (finalement) a été offerte à la première personne étant en mesure de fournir des drivers pour pc. En seulement 3 heures après la sortie du périphérique en Europe, le premier driver a été mis au point et moins d'une semaine après, il était disponible sur internet. Très rapidement, différentes équipes ont réalisé leurs propres versions et il a fallu faire un choix parmi les 2 principales : *LibFreenect* fournie par la communauté « *OpenKinect* » et *CL NUI* par la compagnie « *Code Laborator*. Nous avons décidé d'utiliser *CL NUI* pour nos développements car les possibilités offertes semblaient un peu plus évoluées et qu'elle permettait de programmer en C/C++.

Le driver et la librairie fournis avec *CL NUI* permettaient d'obtenir la carte de profondeur, une image couleur ainsi que le pilotage du moteur et les informations des accéléromètres à un taux de rafraîchissement de 30 fps. Nous avons donc, dans un premier temps, adapté notre code initial pour

qu'il fonctionne avec *CL NUI*, puis nous l'avons amélioré. Avec cette première version améliorée, nous pouvions détecter toutes les mains qui se présentaient à la caméra. Cependant, il était impossible de déterminer si chacune d'elle était une main gauche ou droite et encore moins à qui elles appartenaient. Les mains étaient considérées comme indépendantes les unes des autres et non rattachées à une personne. Une fois le système porté sur *CL NUI*, il a permis d'obtenir de meilleures performances que notre première version. Mais peu de temps après, un nouveau driver est sorti.

Vu l'engouement des développeurs pour ce périphérique, la société *PrimeSense* (qui est à l'origine du développement d'une partie de la *Kinect*) a sorti un *framework* complet sous le nom d'*OpenNI*, laissant penser que les drivers utilisés sont les plus fiables disponibles. L'installation en est fastidieuse car il faut installer les bibliothèques et les drivers dans un ordre précis : *OpenNI* en premier, puis *NITE* et enfin les drivers de la caméra (*SensorKinect Module*) ainsi que les drivers audio. De plus, les différentes versions ne sont pas toujours compatibles et provoquent parfois des conflits.

Cependant, nous avons décidé de nous tourner vers cette nouvelle solution car elle offre, en plus des informations de profondeur, de couleur, etc. (déjà obtenues par les autres drivers), un algorithme permettant d'appliquer un squelette sur les personnes présentes dans le champ de la caméra. Cet algorithme est par ailleurs utilisé par la *Kinect* dans les jeux *Xbox 360*. Nous avons fait le choix de porter la nouvelle version de notre système sur *OpenNI* uniquement pour pouvoir obtenir le squelette des utilisateurs. Dans notre algorithme précédent, il était nécessaire de parcourir l'image en entier pour déterminer où se situaient les mains des utilisateurs. Grâce à cela, il nous est possible de connaître les positions des mains des utilisateurs dans l'espace et de ne travailler que dans des zones 3D précises. Cela permet de gagner énormément de temps de calcul par rapport à la version précédente de notre algorithme dans laquelle nous devions « parcourir » l'image intégralement pour connaître la position des mains. La solution offerte par *CL NUI* est vraiment robuste : un utilisateur peut sortir et revenir du champ de vision, passer devant un autre utilisateur ou encore des objets (tables, chaises, etc.) sans que le tracking soit perdu. Cela est possible grâce à la caméra couleur, qui permet d'établir un histogramme de couleurs de chaque personne. Nous sommes donc en mesure d'identifier d'une part la position des mains dans l'espace mais également les associer à un utilisateur en particulier. D'autre part nous pouvons déterminer si c'est une main gauche ou une main droite. Ce point est déterminant dans la suite de nos traitements, car la main dominante joue un rôle plus important que la main secondaire dans la manipulation d'objet ainsi que dans la navigation. Notre système s'adapte aussi bien pour une personne droitrière que gauchère. Il suffit de choisir en début de manipulation la latéralisation de l'utilisateur et les modalités d'interactions développées s'adaptent automatiquement. La première version de notre système permettait un espace de jeu d'environ 30 x 30 x 30 cm et nécessitait que la caméra soit positionnée au-dessus de la table (à environ 80 cm de celle-ci), la version améliorée permet maintenant un espace de jeu de près de 150 x 150 x 100 cm. Le volume de jeu est donc considérablement agrandi et le nombre de mains détectées n'est plus de une mais en théorie 32. *OpenNI* permet en effet de *tracker* au maximum 16 personnes à la fois. Cependant, la zone de tracking est insuffisante pour permettre à plus de 2 personnes de travailler en même temps tout en étant à l'aise. De plus, pour cette deuxième expérimentation, nous avons décidé qu'une seule personne effectuerait la manipulation. Le caractère collaboratif, bien qu'étant pris en compte dans nos développements et donc tout à fait possible, n'est pas abordé dans cette thèse.

Notre système amélioré permet maintenant de récupérer les positions des mains de chaque utilisateur en les identifiant (gauche ou droite) et de connaître leur état (ouverte ou fermée). Ces informations sont envoyées par connexion UDP au logiciel *Virtools* qui permet d'interagir avec un environnement virtuel. Un serveur (créé en VSL – langage de *Virtools*) permet de récupérer ces informations et d'effectuer les

interactions correspondantes en fonction des mouvements et actions de l'utilisateur. Étant donné que l'utilisateur peut utiliser ses deux mains, nous pouvons effectuer des tâches simples comme la sélection, désélection, déplacement d'objet et navigation mais également des tâches plus complexes comme la rotation ou le redimensionnement d'objet. Pour que l'utilisation du système soit facilitée et permette à tous une meilleure prise en main, une meilleure précision ainsi qu'une meilleure immersion, nous avons tenu compte de la latéralisation de l'utilisateur.

Pour cette deuxième expérimentation, deux ordinateurs ont été utilisés. Un pour le système Cam3D et un pour les gants et les capteurs magnétiques. Les capteurs magnétiques permettent d'obtenir les informations de position 3D des mains, tout en les identifiant (main gauche / main droite) et les gants permettent d'évaluer si les mains sont ouvertes ou fermées. Les modalités d'interaction développées sont exactement identiques entre les deux systèmes.

Très récemment, *Microsoft*, voyant le marché potentiel que représente la communauté des développeurs, a accepté de fournir une version officielle des pilotes ainsi qu'un SDK. Il suffit maintenant d'installer un exécutable et toutes les bibliothèques et drivers sont installés automatiquement, ce qui simplifie énormément l'installation et les problèmes de compatibilité. L'algorithme permettant de déterminer le squelette des joueurs est également fourni. De plus, il n'y a plus besoin de phase d'initialisation de quelques secondes avant que le squelette soit appliqué sur la personne, contrairement à *OpenNI* où la personne doit positionner ses bras en l'air pour former avec sa tête la lettre psi ( $\Psi$ ), la reconnaissance est maintenant immédiate. Des points supplémentaires ont également été ajoutés au squelette. Avec *OpenNI*, seule la main possède un point qui se déplace légèrement si la main est ouverte ou fermée. Avec le SDK de Microsoft, un point situé au niveau du poignet est donné, ainsi qu'un point au centre de la main. Dans les développements futurs, cela simplifiera énormément notre détection de l'état de la main (ouverte ou fermée).

## 2 Protocole expérimental

### 2.1 Généralités

Pour cette expérimentation, nous avons repris des éléments du protocole expérimental de la première étude, en particulier concernant les procédures de recueil ainsi que les sources de données des préférences subjectives et commentaires des participants. Cette étude étant plus longue, et permettant d'extraire davantage de données pour les analyses, nous avons choisi de limiter le nombre de participants par rapport à l'étude précédente, afin de disposer d'un temps suffisant pour réaliser l'ensemble de l'exploitation des résultats, certaines données devant être extraites à la main. Nous conservons par contre le principe d'une évaluation par les participants des deux systèmes, et non par des groupes distincts de participants. Certains mécanismes sont mis en place pour assurer de ce fait la validité interne de nos résultats.

### 2.2 Choix des participants

L'expérimentation a impliqué 20 participants volontaires, 8 femmes et 12 hommes, âgés entre 21 et 34 ans. Tous étaient des utilisateurs expérimentés d'ordinateurs et d'un niveau d'étude BAC+3 minimum et disposaient de connaissances variées en réalité virtuelle. La plupart des participants étaient des étudiants en cours de préparation du Master « Ingénierie du Virtuelle et Innovation » ou des membres de l'équipe « Presence & innovation » du « Laboratoire des Arts et Métiers ParisTech d'Angers » (Doctorants, Ingénieurs de Recherche). Le nombre de participants étant relativement restreint, nous

avons veillé à restreindre la tranche d'âge afin de limiter la variabilité interindividuelle quant aux performances et préférences subjectives (moyenne = 26 ; E.T. = 3,7). Les effectifs relatifs des femmes (moyenne d'âge = 25,4 ; E.T. = 2,3) et des hommes (moyenne = 26,4 ; E.T. = 4,0), proches même s'ils ne sont pas strictement équivalents (8 contre 12), nous ont tout de même permis de contrebalancer l'ordre de passation suivant le sexe. Le sexe, comme l'expertise en réalité virtuelle, seront des variables dont l'impact sera évalué dans l'étude statistique conduite dans ce chapitre.

### 2.3 Conditions initiales et consignes

Les conditions initiales sont identiques quel que soit l'ordre de présentation des périphériques (Gants3D ou Cam3D) : le participant commence avec l'expérimentation proprement dite par la phase d'apprentissage nécessitant le placement, le redimensionnement et l'orientation correcte d'un avion de ligne. Pour chaque tâche, la position et/ou orientation et/ou taille de l'objet et du modèle sont identiques, pour les 2 périphériques.

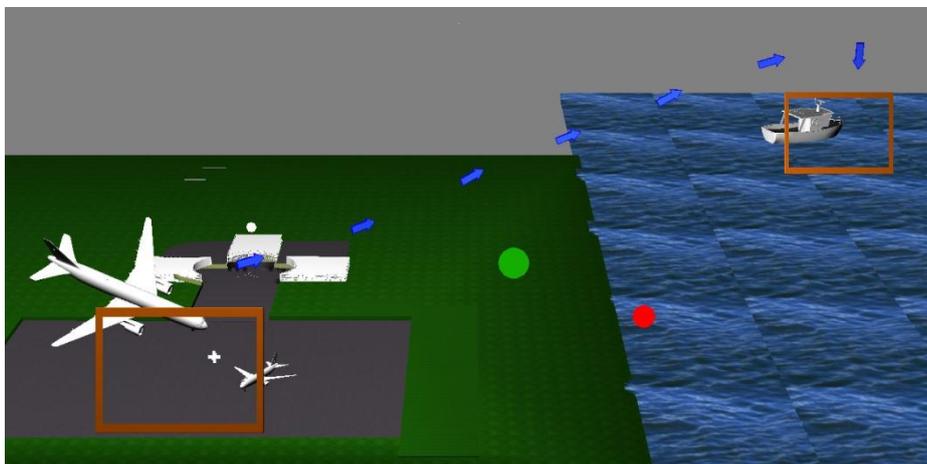


Figure 42 : Environnement virtuel de la phase d'apprentissage de l'expérimentation 2

L'utilisateur est assis à une table, l'écran en face de lui. L'espace de jeu, dans lequel il interagit avec les objets virtuels, se situe au-dessus de la table devant lui. Dans le cas de la Cam3D, la caméra qui capture les mouvements de l'utilisateur est placée au-dessus de l'écran c'est-à-dire à environ 40 centimètres de la table, elle ne gêne ni la vue, ni l'espace de jeu de l'utilisateur. Dans le cas du système Gants3D, un capteur magnétique était fixé sur chaque gant et l'ensemble était relié de manière filaire à l'ordinateur. L'antenne nécessaire aux capteurs magnétiques était posée sur la table devant le participant, sans que cela ne masque l'écran ou ne gêne les gestes du participant.

Il était demandé aux participants d'essayer de faire correspondre avec précision l'objet et le modèle. Il ne leur a pas été demandé explicitement de privilégier la précision sur le temps d'exécution, afin de ne pas induire un temps d'expérimentation exagérément long qui pourrait entraîner une fatigue ou un énervement du participant.

Des consignes sont données en début d'expérimentation, afin d'expliquer l'expérimentation et les tâches à réaliser par le participant. Durant la phase d'apprentissage, sont détaillées au participant, les consignes et les instructions concernant l'interaction et les tâches à réaliser. En plus de ces instructions orales, une explication écrite est donnée à l'utilisateur en guise de rappel sur la manière de sélectionner ou désélectionner les objets, comment appeler le menu et comment réaliser les différentes actions (déplacement, rotation, redimensionnement, et navigation).

Les consignes en cours d'expérimentation étaient des rappels sur les actions à réaliser sur chaque tâche (e.g., redimensionnement et rotation pour l'ordinateur portable) et l'ordre préférentiel à adopter (e.g., rotation puis redimensionnement pour ce même objet). Il s'agissait également de rappel sur l'interaction (principalement pour la rotation) ainsi qu'une aide lorsque le système « perdait » le squelette du participant.

## 2.4 Déroulement de l'expérimentation

Le déroulement de l'expérimentation est identique, que le participant commence avec le système Gants3D ou le système Cam3D.

Le participant commence par remplir le questionnaire d'identification, nous permettant de recueillir les informations générales sur les participants (âge, sexe, latéralisation, niveau d'étude) et les informations spécifiques nous permettant d'établir le niveau d'expertise en réalité virtuelle. Ce questionnaire est un formulaire au format pdf réalisé avec le logiciel Adobe LiveCycle Form.

L'expérimentation commence par une phase de prise en main qui permet au participant de se familiariser avec les modalités d'interactions et les tâches à réaliser. Afin de ne pas introduire un biais dans l'expérimentation, le premier environnement est distinct de ceux utilisés pour l'évaluation des 2 systèmes. Pour cette première tâche, le participant doit tester l'ensemble des actions possibles : déplacement, redimensionnement, orientation, navigation ainsi que les actions de sélection/désélection et l'utilisation des menus. Lorsque le participant a réussi à manipuler l'avion pour le faire correspondre au modèle et de naviguer de la scène 1 à la scène 2, on considère qu'il a correctement assimilé le fonctionnement du périphérique et les tâches à réaliser.

Lorsque le participant se sent prêt et que l'expérimentateur juge que la prise en main est réussie (i.e. le participant réussit à utiliser le système sans difficulté particulière), l'expérimentation sur laquelle on enregistre les données débute par l'affichage d'un écran explicatif (voir Figure 43). Une fois la lecture des consignes achevée, le participant doit placer sa main dominante au centre de l'écran pendant 3 secondes. C'est à partir de ce moment que commence l'enregistrement des informations utiles à l'analyse des données, et l'utilisateur commence par la scène 1.

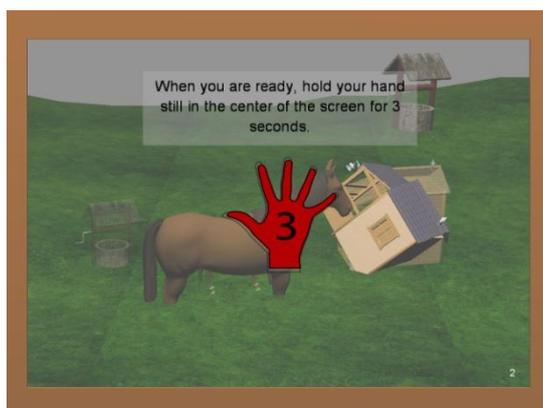


Figure 43 : Écran des consignes au début de l'expérimentation.

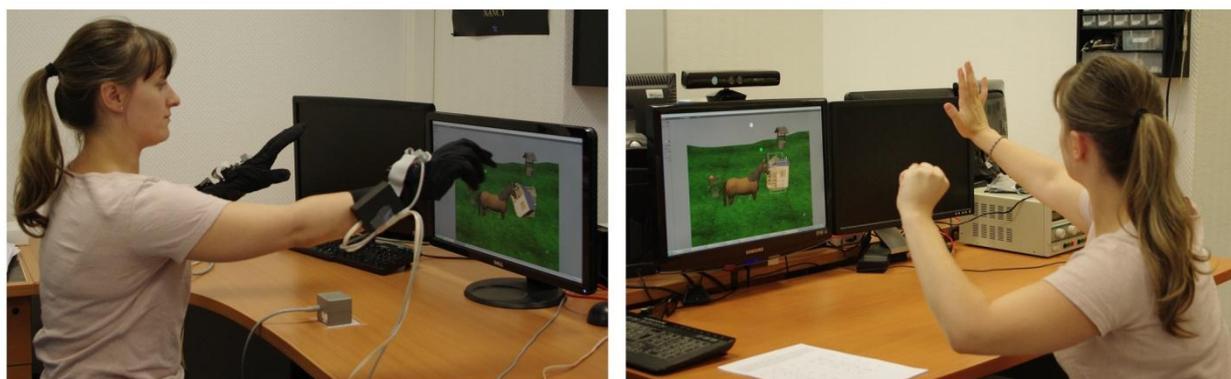
Les participants devaient alors effectuer les 9 tâches demandées, dans l'ordre indiqué par l'expérimentateur (le même ordre pour tous les participants). Chaque tâche se terminait lorsque le participant jugeait correcte la correspondance entre l'objet et le « modèle » ou, correcte, la position

terminale après la navigation. Le temps n'était pas limité pour l'exécution de chacune des 9 tâches, mais le temps total d'expérimentation était limité à 30 minutes (à l'exclusion des éventuels problèmes d'ordre technique). Nous avons estimé qu'au-delà de ce temps, le participant aurait pu éprouver du découragement ou de la fatigue et que cela aurait introduit un biais au niveau des données recueillies. Il faut tout de même noter qu'aucun des participants n'a dépassé cette limite de 30 minutes. Le Tableau 13 présente l'enchaînement des 9 tâches.

**Tableau 13 : Enchaînement des tâches (D = déplacement, T = redimensionnement, R = rotation, Navi. = navigation) - Les cases grisées sont des actions non réalisables pour les tâches correspondantes**

Scène 1									Tâche 4	Scène 2									Tâche 8	Scène 3		
Tâche 1			Tâche 2			Tâche 3				Tâche 5			Tâche 6			Tâche 7				Tâche 9		
Puits			Cheval			Maisonnette				Globe			Horloge			Ordinateur				Bateau		
D	T	R	D	T	R	D	T	R	Navi.	D	T	R	D	T	R	D	T	R	Navi.	D	T	R

A l'issue de la réalisation des 9 tâches, le participant fait apparaître le menu une dernière fois et sélectionne le bouton « valider » pour terminer l'expérimentation. Une fois le menu validation sélectionné, l'expérimentation prend fin et la collecte des informations s'arrête. Il n'est pas possible de revenir en arrière. Le participant doit remplir un questionnaire final (également réalisé sous la forme d'un formulaire pdf). Ce questionnaire nous permet de recueillir les impressions, commentaires et jugements subjectifs des participants.



**Figure 44 : Participante passant l'expérimentation 2 avec le système Gants3D (à gauche) et le système Cam3D (à droite)**

## 2.5 Sources de données

Comme pour la première expérimentation, l'ensemble des interactions des participants avec les 2 systèmes ont été enregistrées : les actions des participants (sélection, désélection, choix d'un item du menu), l'état (ouvert, fermé) des mains des participants, ainsi que la position, l'orientation et la taille des objets. Toutes ces données ont été automatiquement marquées, datées et enregistrées. Ces données nous permettront d'extraire les variables qui rendront possible l'analyse des performances.

Les données permettant l'identification des participants et l'évaluation de l'expertise en réalité virtuelle sont obtenues à partir du questionnaire initial, qui contient des questions ouvertes ainsi que des questions fermées et des échelles de Likert.

Les commentaires et jugement subjectifs sont recueillis dans le questionnaire post-expérimentation. Celui-ci contient des questions ouvertes, des questions fermées ainsi que des questions sous la forme d'échelle de Likert. Les données recueillies concernent l'évaluation comparative de deux systèmes, l'évaluation de l'interaction ainsi que les remarques générales quant à l'expérimentation et aux améliorations éventuelles.

## 2.6 Variables étudiées

Les mesures suivantes ont été prises en compte et ont servi de base à l'étude des performances des participants :

- Le temps d'exécution total des 9 tâches (en secondes). C'est le temps mis par l'utilisateur pour accomplir les 9 tâches avec un des deux périphériques. Il débute après que l'utilisateur a lu les consignes, la fiche explicative et qu'il a positionné sa main 3 secondes au centre de l'écran ; il s'arrête à la sélection du menu « validation ».
- Le temps d'exécution pour chaque tâche (en secondes) et chaque système.
- Le temps d'action total des 9 tâches (en secondes) pour chaque système. Ce temps correspond aux périodes pendant lesquelles le participant effectue une action de rotation, de dimensionnement, de déplacement ou de navigation. Il exclut donc les phases de choix dans le menu, les périodes de « non-activité » ainsi que les moments où aucun objet n'est sélectionné.
- Le temps d'action (en secondes) pour chaque tâche et chaque système. Ces temps sont différenciés le cas échéant selon l'action réalisée.
- L'erreur de précision : elle est soit globale, c'est-à-dire moyennée sur les 3 axes, soit calculée pour chaque axe.
  - o Pour le redimensionnement, il n'y a qu'une valeur, car il est homogène sur les 3 axes. L'erreur correspond au pourcentage de différence par rapport à l'échelle de référence.
  - o Pour la rotation, l'erreur correspond à un pourcentage calculée d'après le décalage angulaire. Une erreur de 100% correspond à un décalage angulaire de 180°.
  - o Pour le déplacement, l'erreur correspond à un pourcentage calculé par rapport à la position de référence, normalisé selon la dimension de l'objet indépendamment pour les 3 axes. Une précision nulle sur un axe pour le déplacement correspond à un décalage au moins équivalent à la dimension de l'objet.
- Le ratio précision / temps d'action pour chaque tâche et chaque système. Ce ratio met en relation la précision atteinte avec le temps de manipulation du participant, et permet d'évaluer la performance du participant. Il s'agit de la précision moyenne sur les 3 axes pour la rotation et le déplacement.
- Le nombre d'actions différentes (i.e., la séquence d'actions) pour les tâches demandant de conjuguer 2 ou 3 actions (T5 à T7 et T9). Cela correspond à la combinaison d'actions élémentaires nécessaires pour réaliser une tâche.

L'expertise en Réalité Virtuelle a été calculée d'après les données du questionnaire initial, sur la même base que pour l'expérimentation 1 : prise en compte de réponses à des échelles de Likert et groupement selon 3 thèmes de poids équivalents (connaissances générales en RV, utilisation de périphériques spécifiques à la RV et utilisation de logiciels spécifiques à la RV). De la même manière que pour l'expérimentation 1, trois groupes ont été constitués (Tableau 14).

Tableau 14 : Niveaux d'expertises basés sur les scores calculés d'après les réponses aux échelles de Likert

Niveau d'expertise	Débutant	Moyen	Expert
Score	de 0 à 1,5	de 1,5 à 3	de 3 à 5
Effectif par groupe	2	2	16

Comparativement à la première expérimentation, où les effectifs des groupes étaient relativement distribués, nous remarquons que la majorité des participants est ici qualifiée d'experte. Cela tient à l'effectif plus restreint de cette expérimentation et à l'origine des participants (étudiants en RV, membres de l'équipe), ceci comme résultat d'avoir une hétérogénéité restreinte en termes de différences inter-individuelles.

L'étude des préférences subjectives et commentaires des participants est divisée en deux parties. D'une part, les appréciations concernant les modalités d'interaction et d'autre part une évaluation de chaque système de manière séparée puis comparative.

Pour la première partie, qui fera l'objet d'un traitement séparé dans les résultats, nous avons demandé aux participants d'évaluer chaque modalité d'interaction selon des échelles de Likert à 5 niveaux pour les critères suivants :

- l'efficacité globale
- la rapidité
- la précision
- la simplicité
- l'intuitivité de l'action

Les participants ont pu également donner leurs impressions grâce à des questions ouvertes : points positifs / négatifs et suggestions éventuelles d'amélioration pour chaque modalité.

L'étude de l'acceptabilité repose quant à elle sur l'évaluation de chaque système selon des échelles de Likert à 5 modalités, pour 10 critères regroupés en 3 catégories :

- confort : poids, maniabilité, liberté de mouvement, fatigue engendrée, confort d'utilisation
- efficacité : efficacité globale, précision, stabilité, fiabilité de la reconnaissance
- efficience : facilité d'utilisation

Les participants étaient interrogés sur leurs préférences globales au moyen de questions à choix multiples (Gants – Cam3D – Egal) sur les critères suivants :

- préférence générale
- périphérique le plus approprié
- meilleur sentiment d'immersion
- meilleure précision
- meilleure aisance d'utilisation

Enfin, nous avons demandé aux participants de juger l'adéquation entre le périphérique et les modalités d'interaction au regard de la précision et de la facilité d'utilisation.

Les Tableau 15 et Tableau 16 résument les variables étudiées ainsi que leur format.

Tableau 15 : Étude des systèmes – variables étudiées et données recueillies

<b>Acceptabilité</b> <i>Evaluation séparée des systèmes</i>		<b>Préférences globales</b> <i>Evaluation comparative des systèmes</i>
<b>Echelle de Likert (5 modalités)</b>		<b>Gants 3D – Cam3D – Egal</b>
Confort	Poids	Préférences générales
	Maniabilité	Système le + approprié
	Liberté de mouvement	Sentiment d'immersion
	Fatigue	Précision
	Confort d'utilisation	Aisance d'utilisation
Efficacité	Efficacité globale	
	Précision	
	Stabilité	
	Fiabilité	
Efficiency	Facilité d'utilisation	

Tableau 16 : Étude des modalités d'interaction – variables étudiées et données recueillies

<b>Adéquation</b> <b>périphérique – modalités d'interaction</b>		<b>Evaluation des modalités</b> <b>d'interaction</b>
<b>Gants 3D – Cam3D – Egal</b>		<b>Echelle de Likert</b>
<b>Déplacement</b>	Précision	Efficacité globale
	Facilité d'utilisation	Rapidité
<b>Rotation</b>	Précision	Précision
	Facilité d'utilisation	Simplicité
<b>Redimensionnement</b>	Précision	Intuitivité de l'action
	Facilité d'utilisation	
<b>Navigation</b>	Précision	
	Facilité d'utilisation	

Les réponses aux échelles de Likert seront traitées comme des variables numériques, chaque modalité étant codée par un nombre (1 = plus mauvais ; 5 = meilleur).

Les zones de textes libres, les justifications et les propositions d'améliorations serviront à illustrer les résultats.

## 2.7 Conditions de passation

Pour éviter au maximum toute variabilité dans les indications données aux sujets, une fiche descriptive a été mise en place pour expliquer en quoi consiste l'expérimentation d'une part et d'autre part pour expliquer comment fonctionnent les modalités d'interaction afin de réaliser les différentes tâches demandées. Cette fiche est expliquée oralement à chacun des participants et leur est ensuite laissée pour qu'ils puissent s'y référer à tout moment pendant la manipulation, si nécessaire.

Une fiche comprenant des consignes précises a été donnée à chaque participant avant l'expérimentation. Cette fiche décrit étape par étape ce que doit faire l'utilisateur pendant l'expérimentation afin qu'il n'y ait pas de différence quant aux consignes initiales données aux participants par l'examineur. Le nombre de tâches à réaliser ainsi que les éventuels problèmes techniques (e.g., problème de reconnaissance de la Kinect) imposait la présence d'un expérimentateur, que nous avons réussi à éviter dans la première expérimentation. L'expérimentateur était assis à côté du participant, légèrement en retrait pour éviter que ce dernier puisse se tourner vers lui. Il vérifiait que la reconnaissance du squelette et la détection de l'état des mains (ouverte / fermée) étaient correctes et donnait les instructions en cas de problèmes. Au début de chaque tâche, l'expérimentateur rappelait

l'objectif (e.g., l'ordinateur portable) ainsi que les actions à réaliser (e.g., rotation et redimensionnement) et la séquence recommandée (e.g., rotation puis redimensionnement). En dehors de ces interventions, l'expérimentateur restait silencieux mais pouvait répondre aux questions des participants, le cas échéant.

La collecte des informations est entièrement automatisée, ce qui facilite la mise en forme de celles-ci en vue de leur analyse par la suite.

## **2.8 Validité interne et externe**

Pour assurer la validité interne et externe, nous avons eu recours aux mêmes techniques que dans l'expérimentation précédente. Pour assurer la validité interne, nous avons essayé de restreindre l'origine (en termes de formation et d'expertise en réalité virtuelle) et l'âge des participants et essayer de maîtriser les variables parasites provenant de l'environnement réel (éclairage, bruit, température), l'expérimentation se tenant dans une salle unique, climatisée et sans fenêtre.

### **2.8.1 Appariement**

Étant donné l'effectif relativement restreint, nous avons fait le choix de faire évaluer successivement les 2 systèmes à chaque participant plutôt que de disposer de 2 groupes distincts. L'appariement nous permet d'obtenir des résultats comparables pour les 2 systèmes et de contrôler les facteurs parasites qui demeurent constants durant l'expérimentation.

### **2.8.2 Contre-balancement**

Nous avons mis en œuvre le contre-balancement pour pallier l'éventuel apprentissage induit par l'appariement. Cela nous permet de prendre en compte la variabilité éventuelle des résultats en cours d'expérimentation (amélioration ou dégradation). Nous avons réparti les sujets en 2 groupes, nommés G1 et G2 dans la suite de ce chapitre. Les sujets appartenant au groupe G1 ont commencé par l'utilisation de la Cam3D, ceux du groupe G2 ont commencé par l'utilisation des Gants3D. Chaque groupe contient 4 femmes et 6 hommes.

### **2.8.3 Affectation aléatoire**

L'affectation des participants à un groupe ou à l'autre (G1 ou G2) s'est faite de façon aléatoire. En revanche, la position, l'orientation et l'échelle des objets dans les scènes sont prédéterminées et est identique pour chaque participant et pour chaque système à tester.

## **3 Analyse comparative des performances des participants**

Nous allons nous intéresser dans cette partie aux résultats comparés des participants selon l'utilisation de chaque système. Après le compte-rendu des tests préliminaires et le détail des tests statistiques réalisés, l'étude de la comparaison des 2 systèmes portera en premier lieu sur les données globales, c'est-à-dire sur l'ensemble de 7 tâches. Nous évaluerons également l'influence éventuelle du sexe et de l'expertise sur ces résultats. En second lieu, nous nous intéresserons à chaque tâche séparément. Étant donné le nombre de tâches et le nombre de variables, nous ne réaliserons pas l'étude de l'interaction éventuelle entre sexe / expertise en RV et périphérique dans ce cas. Nous terminerons la présentation des résultats par les études annexes, portant sur la comparaison de la précision décomposée suivant les axes (rotation et déplacement), sur le ratio temps / précision ainsi que sur l'impact de l'effet d'apprentissage. Enfin, nous mettrons en perspective l'ensemble des résultats dans la discussion.

### 3.1 Tests préliminaires et tests statistiques

Nous avons conservé pour les statistiques la même démarche que dans la première expérimentation. Nous avons donc effectué des tests de Kolmogorov-Smirnov pour vérifier la normalité des distributions des valeurs obtenues pour chaque variable prise en compte pour les 2 systèmes étudiés. Seules quelques variables ne suivent pas une loi normale, pour celles-ci, nous choisirons les tests statistiques non paramétriques mis à part dans le cas d'étude de l'interaction entre plusieurs facteurs auxquels nous privilégierons l'analyse de la variance, étant donnée la robustesse de l'ANOVA aux erreurs de type 1 (Winer, 1971).

Pour analyser l'influence du système sur les performances des participants, nous avons utilisé le test T de Student pour échantillons appariés, ainsi que le test de Wilcoxon en cas de non-normalité d'une des variables comparées. Lors de la comparaison des valeurs prises par 3 variables, nous avons utilisé le test T de Student paire à paire (ou le test de Friedman puis le test de Wilcoxon paire à paire en cas de non-normalité). L'analyse des effets du sexe et de l'expertise en RV en relation avec le système repose sur une ANOVA mixte avec une variable intra-sujet (le système) et une variable inter-sujet (le sexe ou l'expertise en RV). L'étude des effets simples s'appuie sur un test T de Student pour variables indépendantes (2 modalités) ou sur une ANOVA simple (3 modalités). Enfin, des tests post-hoc LSD de Fisher permettent d'étudier les éventuels effets principaux. Enfin, l'étude de l'impact de l'éventuel effet d'apprentissage sur les performances s'appuie sur des tests T de Student ou des tests non paramétriques pour échantillons indépendants.

Nous avons adopté les règles suivantes concernant la significativité des tests statistiques :

- $p \leq 0,05$  : test significatif
- $0,05 < p \leq 0,1$  : tendance
- $p > 0,1$  : test non significatif

### 3.2 Résultats généraux

Le Tableau 17 présente les données correspondant à la comparaison du temps d'expérimentation total ainsi que du temps de manipulation total entre les 2 systèmes.

Tableau 17 : Résultats concernant les temps totaux d'expérimentation et de manipulation

		Temps total d'expérimentation (s)		Temps total de manipulation (s)	
		Gants3D	Cam3D	Gants3D	Cam3D
Moyenne		902,4	954,4	514,5	522,4
Ecart-type		266,7	274,0	176,6	193,5
Test T	T	1,000		0,179	
	p	0,330		0,860	

Nous ne notons pas de différence entre les systèmes pour le temps total d'expérimentation et le temps total de manipulation.

#### Etude de l'influence du sexe sur les temps totaux d'expérimentation et de manipulation

Une ANOVA à plan mixte (sexe x système) sur le temps total d'expérimentation n'indique pas d'effet du système ( $F(1, 16) = 0,03$  ; NS), ni du sexe ( $F(1, 16) = 1,359$  ; NS) et pas d'interaction entre ces 2 facteurs ( $F(1, 16) = 0,164$ ). Concernant le temps total de manipulation, nous n'observons pas d'effet du système

( $F(1, 16) = 0,0437$  ; NS), ni du sexe ( $F(1, 16) = 1,835$  ; NS) et pas d'interaction entre ces 2 facteurs ( $F(1, 16) = 0,084$ ).

### **Etude de l'influence de l'expertise en RV sur les temps totaux d'expérimentation et de manipulation**

Une ANOVA à plan mixte (expertise en RV x système) sur le temps total d'expérimentation n'indique pas d'effet du système ( $F(1, 16) = 0,03$  ; NS), ni de l'expertise en RV ( $F(2, 16) = 0,875$  ; NS) et pas d'interaction entre ces 2 facteurs ( $F(1, 16) = 0,677$ ). Concernant le temps total de manipulation, nous n'observons pas d'effet du système ( $F(1, 16) = 0,0437$  ; NS), ni d'expertise en RV ( $F(1, 16) = 1,503$  ; NS) et pas d'interaction entre ces 2 facteurs ( $F(1, 16) = 0,640$ ).

### **Synthèse**

Nous n'obtenons de résultats significatifs pour les résultats globaux, ni d'effet du sexe ou de l'expertise en RV sur ceux-ci. De ce fait, nous pouvons affirmer que les Gants3D ne permettent pas de réaliser l'ensemble des tâches plus rapidement que la Cam3D. De plus, le temps de manipulation, c'est-à-dire le temps cumulé pendant lequel les participants manipulent un objet (redimensionnement, rotation, déplacement) ou naviguent, n'est pas significativement inférieur. Ce premier résultat permet de valider partiellement l'hypothèse H2 car, sans être meilleure, la Cam3D permet d'obtenir des résultats équivalents. La validation est partielle car les variables étudiées sont macroscopiques et ne permettent pas déterminer s'il existe des différences de temps d'exécution entre les différentes modalités d'interaction. Ces résultats seront présentés dans la partie suivante.

### **3.3 Gants vs. Cam3D selon chaque tâche**

Pour chaque tâche et chaque système, nous étudions les variables suivantes (elles ont été détaillées dans la partie 2.6 (p. 110) :

- Le temps d'exécution (TE) en secondes ;
- Le temps de manipulation (TM) en secondes ; celui-ci est différencié selon le déplacement (TMD), la rotation (TMR) et le redimensionnement (TME) le cas échéant ;
- L'erreur de précision, donnée selon 4 variables distinctes : déplacement (Err\_D), rotation (Err\_R), navigation (Err\_N) et redimensionnement (Err\_E) ;
- Le nombre d'actions différentes (Seq).

Étant donné le nombre important de variables manipulées et dans le but d'assurer une meilleure lisibilité de nos résultats, nous ne présenterons dans les tableaux que les variables pour lesquelles nous obtenons des différences significatives entre les 2 systèmes.

#### **3.3.1 Tâche 1 – Puits**

La tâche 1 correspond au déplacement d'un objet (puits). Seul le temps d'exécution avec les Gants3D ne suit pas une loi normale ( $Z = 1,420$  ;  $p = 0,035$ ). Nous avons donc réalisé le test T de Student pour échantillons appariés sur les variables TM, TE, Err\_D et Seq.

Il n'y a pas de différence significative entre les Gants3D et la Cam3D pour les variables étudiées.

#### **3.3.2 Tâche 2 – Cheval**

La tâche 2 correspond au redimensionnement d'un objet (cheval). Seul le nombre d'actions différentes (Seq) avec les Gants3D ( $Z = 2,014$  ;  $p = 0,001$ ) et la Cam3D ( $Z = 1,820$  ;  $p = 0,003$ ) ne suit pas une loi

normale. Nous avons donc réalisé le test de Wilcoxon pour échantillons appariés sur la variable Seq et le test T de Student pour échantillons appariés sur les variables TM, TE et Err\_R.

Le temps de manipulation (TM) et le temps d'exécution (TE) sont significativement plus courts avec les Gants3D. On note une tendance en faveur des Gants3D concernant la séquence d'actions différentes. Le Tableau 18 présente ces résultats.

**Tableau 18 : Différences significatives entre Cam3D et Gants3D pour la tâche 2**

	TE		TM		Seq	
	Gants3D	Cam3D	Gants3D	Cam3D	Gants3D	Cam3D
<b>Moyenne</b>	33,1	41,3	10,0	14,9	1,3	0,571
<b>Ecart-type</b>	10,4	17,2	3,2	8,1	1,65	0,933
<b>Test</b>	(T) 2,093		(T) 2,405		(Z) 1,643	
<b>p</b>	0,05		0,027		0,10	

### 3.3.3 Tâche 3 – Maisonnette

La tâche 3 correspond à la rotation d'un objet (maisonnette). L'ensemble des variables suivent une loi normale, nous avons donc appliqué le test T de Student pour échantillons appariés sur les variables TM, TE, Err\_E et Seq.

Il n'y a pas de différence significative entre les Gants3D et la Cam3D pour les variables étudiées.

### 3.3.4 Tâche 4 – Navigation 1

La tâche 4 correspond à une navigation entre la scène 1 et la scène 2. L'ensemble des variables suivent une loi normale, nous avons donc appliqué le test T de Student pour échantillons appariés sur les variables TM, TE, Err\_N et Seq.

Il n'y a pas de différence significative entre les Gants3D et la Cam3D pour les variables étudiées.

### 3.3.5 Tâche 5 – Globe

La tâche 5 correspond au redimensionnement et au déplacement d'un objet (globe). Seul le nombre d'actions différentes (Seq) avec les Gants3D ( $Z = 1,442$ ;  $p = 0,031$ ) et la Cam3D ( $Z = 1,731$  ;  $p = 0,005$ ) ne suit pas une loi normale. Nous avons donc réalisé le test de Wilcoxon pour échantillons appariés sur la variable Seq et le test T de Student pour échantillons appariés sur les variables TE, TMD, TME, Err\_E, Err\_D.

Il n'y a pas de différence significative entre les Gants3D et la Cam3D pour les variables étudiées.

### 3.3.6 Tâche 6 – Horloge

La tâche 6 correspond à la rotation et au déplacement d'un objet (horloge). Seul le nombre d'actions différentes (Seq) avec les Gants3D ( $Z = 1,880$ ;  $p = 0,002$ ) et la Cam3D ( $Z = 2,093$  ;  $p < 0,001$ ) ne suit pas une loi normale. Nous avons donc réalisé le test de Wilcoxon pour échantillons appariés sur la variable Seq et le test T de Student pour échantillons appariés sur les variables TE, TMD, TMR, Err\_R, Err\_D.

Le temps de déplacement (TMD) et l'erreur de précision en rotation (Err\_R) sont significativement plus faibles avec les Gants3D. Le Tableau 19 présente ces résultats.

Tableau 19 : Différences significatives entre Cam3D et Gants3D pour la tâche 6

	TMD		Err_R	
	Gants3D	Cam3D	Gants3D	Cam3D
<b>Moyenne</b>	24,0	52,6	2,76	3,66
<b>Ecart-type</b>	6,9	30,4	1,71	2,91
<b>Test</b>	(T) 4,268		(T) 2,046	
<b>p</b>	< 0,001		0,055	

### 3.3.7 Tâche 7 – Ordinateur portable

La tâche 7 correspond à la rotation et au redimensionnement d'un objet (ordinateur portable). Seul le nombre d'actions différentes (Seq) avec les Gants3D ( $Z = 1,565$  ;  $p = 0,015$ ) ne suit pas une loi normale. Nous avons fait le choix de réaliser le test T de Student pour échantillons appariés sur les variables TE, TME, TMR, Err\_R, Err\_E.

Le temps de redimensionnement (TME) tend à être plus court avec les Gants 3D alors que la durée de rotation (TMR) est significativement inférieure avec la Cam3D. Le Tableau 20 présente ces résultats.

Tableau 20 : Différences significatives entre Cam3D et Gants3D pour la tâche 7

	TME		TMR	
	Gants3D	Cam3D	Gants3D	Cam3D
<b>Moyenne</b>	20,0	29,1	80,1	46,4
<b>Ecart-type</b>	11,7	18,3	49,4	32,9
<b>Test</b>	(T) 1,920		(T) 3,052	
<b>p</b>	0,07		0,007	

### 3.3.8 Tâche 8 – Navigation 2

La tâche 8 correspond à une navigation entre la scène 1 et la scène 2. L'ensemble des variables suivent une loi normale, nous avons donc appliqué le test T de Student pour échantillons appariés sur les variables TM, TE, PD, Err\_N et Seq.

Il n'y a pas de différence significative entre les Gants3D et la Cam3D pour les variables étudiées.

### 3.3.9 Tâche 9 – Bateau

La tâche 9 correspond à la rotation, au déplacement et au redimensionnement d'un objet (bateau). Seul le nombre d'actions différentes (Seq) avec les Gants3D ( $Z = 1,565$  ;  $p = 0,015$ ) ne suit pas une loi normale. L'ensemble des variables suivent une loi normale, nous avons donc appliqué le test T de Student pour échantillons appariés sur les variables TE, TME, TMR, TMD, PR, PE, PD, Err\_R, Err\_E, Err\_D.

L'erreur de déplacement est significativement inférieure avec les Gants3D

Tableau 21 : Différence significative entre Cam3D et Gants3D pour la tâche 9

	Err_D	
	Gants3D	Cam3D
<b>Moyenne</b>	5,06	10,42
<b>Ecart-type</b>	3,31	10,75
<b>Test</b>	(T) 2,229	
<b>p</b>	0,038	

### 3.3.10 Synthèse

Pour 5 des 9 tâches, nous n'observons aucune différence significative entre les Gants3D et la Cam3D. Pour l'ensemble des 4 autres tâches, seules 8 variables permettent de différencier ces 2 systèmes, principalement concernant le temps de déplacement, de rotation ou de redimensionnement le cas échéant. Sur ces 8 variables considérées, 6 indiquent une différence en faveur des Gants3D.

Pour la tâche simple de redimensionnement (tâche 2, cheval), le temps d'exécution et le temps de manipulation sont plus faibles avec les Gants3D.

Pour la tâche demandant des actions de déplacement et de rotation (tâche 6, horloge), nous observons un temps de déplacement plus faible avec les Gants3D, ainsi qu'une erreur plus faible pour la rotation avec ce même système.

Pour la tâche demandant des actions de redimensionnement et de rotation (tâche 7, ordinateur portable), on observe un temps plus court pour le redimensionnement avec les Gants3D mais un temps de rotation plus court avec la Cam3D.

Enfin, pour la tâche mêlant les 3 actions élémentaires (tâche 9, bateau), on observe une erreur de déplacement plus faible avec les Gants3D.

Pris dans leur ensemble, ces résultats ne démontrent pas la supériorité des Gants3D, car ils restent très partiels, n'étant basés que sur quelques différences statistiquement significatives, ou encore des tendances. Les Gants3D semblent être à même de minimiser le temps d'exécution pour les actions de redimensionnement. Cette tâche demande peu de précision comparativement aux autres, et le meilleur temps obtenu avec les Gants3D peut s'expliquer par l'absence de problème de reconnaissance de l'état de la main d'une part, et par le fait que les mains (et donc les capteurs de position) ne s'éloignent pas de l'antenne, ce qui se traduit par une bonne stabilité de la valeur de position d'autre part. Pour les autres variables, il est impossible de mettre en valeur l'un ou l'autre système.

Ces résultats viennent infirmer l'hypothèse H1 et confirmer l'hypothèse H2. En effet, nous nous attendions à une supériorité de la Cam3D sur les valeurs de précision, étant donnée l'absence de port de matériel sur les mains, et donc de poids à supporter. Au final, la précision reste équivalente entre les systèmes. Le temps d'expérimentation, relativement court, n'a peut-être pas permis d'obtenir une discrimination sur la précision, en limitant la fatigue engendrée par le port des gants et des capteurs. A l'opposé, nous attendions des temps d'exécution *a minima* équivalents entre les 2 systèmes. Globalement, mis à part quelques variables, cette hypothèse est vérifiée. Nous pouvons donc affirmer qu'aux stades des développements actuels, pour des tâches n'imposant pas de reconnaître des gestes de la main, la Cam3D est au niveau d'un périphérique courant de réalité virtuelle en ce qui concerne les performances (temps, précision).

## 3.4 Étude de l'erreur d'orientation et de positionnement décomposée selon les axes X, Y et Z

### 3.4.1 Déplacement

Nous n'avons considéré que les tâches mettant en jeu des déplacements : tâche 1 (puits), tâche 5 (globe), tâche 6 (horloge) et tâche 9 (bateau). Le Tableau 22 synthétise l'ensemble des résultats pour les Gants3D.

Tableau 22 : Décomposition de l'erreur de positionnement selon chaque axe pour les Gants3D

Tâche	1 (puits)			5 (globe)			6 (horloge)			9 (bateau)		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moy	8,47	9,89	20,09	1,88	2,47	6,39	44,82	18,85	8,93	4,28	7,96	2,96
E.T.	6,33	9,02	17,51	2,05	1,90	3,79	3,40	19,58	6,94	4,27	6,60	2,58
Paire	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z
Test	0,954	3,561	4,143	1,052	5,289	4,186	5,539	2,450	20,123	2,412	3,583	1,276
p	0,352	<b>0,002</b>	<b>0,001</b>	0,306	<b>&lt;,001</b>	<b>0,001</b>	<b>&lt;,001</b>	<b>0,024</b>	<b>&lt;,001</b>	<b>0,026</b>	<b>0,002</b>	0,217

Pour les tâches 1 et 5, l'erreur est significativement plus importante sur l'axe Z que sur les axes X et Y. Ce constat s'inverse pour les tâches 6 et 9, l'erreur est significativement moins importante sur l'axe Z.

Le Tableau 23 synthétise l'ensemble des résultats pour la Cam3D.

Tableau 23 : Décomposition de l'erreur de positionnement selon chaque axe pour la Cam3D

Tâche	Tâche 1 (puits)			Tâche 5 (globe)			Tâche 6 (horloge)			Tâche 9 (bateau)		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moy	9,62	8,70	18,58	2,20	2,30	6,84	43,51	27,74	12,75	3,21	22,98	5,07
E.T.	11,57	11,79	16,10	2,11	3,13	7,89	3,18	28,15	10,64	2,95	27,76	4,57
Paire	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z
Test	1,082	4,429	4,030	0,184	2,953	3,290	2,620	2,387	12,387	3,386	3,108	2,168
p	0,293	<b>&lt;,001</b>	<b>0,001</b>	0,856	<b>0,008</b>	<b>0,004</b>	<b>0,017</b>	<b>0,028</b>	<b>&lt;,001</b>	<b>0,003</b>	<b>0,006</b>	<b>0,043</b>

Les conclusions sont les mêmes que précédemment pour les tâches 1, 5 et 6. Pour la tâche 9, l'erreur est significativement plus importante avec l'axe Y, puis l'axe Z que l'axe X.

### 3.4.2 Rotation

Nous n'avons considéré que les tâches mettant en jeu des rotations : tâche 3 (maisonnette), tâche 6 (horloge), tâche 7 (ordinateur portable) et tâche 9 (bateau). Le Tableau 24 synthétise l'ensemble des résultats pour les Gants3D.

Tableau 24 : Décomposition de l'erreur d'orientation selon chaque axe pour les Gants3D (\* le test utilisé est le test de Friedman, pour cause de non-normalité)

Tâche	3 (maisonnette)			6 (horloge)			7 (ordinateur)			9 (bateau) *		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moy	1,42	1,89	2,78	2,56	3,25	2,47	1,61	0,94	0,78	2,33	3,22	1,94
E.T.	1,01	1,44	2,61	2,70	2,07	2,91	1,42	0,92	0,66	3,14	2,49	2,15
Paire	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z
Test	1,182	1,301	2,486	0,890	1,340	0,096	2,058	0,688	2,498	Khi2 = 4,9		
p	0,252	0,209	<b>0,022</b>	0,385	0,196	0,924	<b>0,054</b>	0,50	<b>0,022</b>	0,086		

Pour une tâche, l'erreur est significativement plus faible avec l'axe X qu'avec l'axe Z (tâche 3). Pour la tâche 7, l'erreur est significativement supérieure avec l'axe X qu'avec les axes Y et Z. Il n'y a pas de différences significatives pour les autres tâches.

Le Tableau 25 synthétise l'ensemble des résultats pour la Cam3D.

Tableau 25 : Décomposition de l'erreur d'orientation selon chaque axe pour la Cam3D

Tâche	3 (maisonnette)			6 (horloge)			7 (ordinateur)			9 (bateau)		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
Moy	1,14	2,03	1,53	3,67	3,72	3,58	2,42	0,69	0,69	1,50	3,22	2,14
E.T.	0,94	1,84	1,59	4,64	2,80	3,31	2,27	0,84	0,78	1,22	2,74	1,73
Paire	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z	X - Y	Y - Z	X - Z
Test	1,946	0,978	1,266	0,054	0,182	0,109	3,562	0	3,288	3,151	1,656	1,405
p	0,067	0,340	0,221	0,958	0,857	0,914	<b>0,002</b>	1	<b>0,004</b>	<b>0,005</b>	0,114	0,176

Pour la tâche 7, l'erreur d'orientation est significativement plus faible avec les axes Y et Z qu'avec l'axe X. Pour la tâche 9, l'erreur d'orientation est significativement plus faible avec l'axe X qu'avec l'axe Y.

### 3.4.3 Synthèse

Concernant le déplacement, nous obtenons des résultats similaires entre les 2 systèmes. Pour les premières tâches (le puits et le globe), les déplacements sont significativement moins précis avec l'axe Z qu'avec les axes X ou Y. Pour le puits, cela peut s'expliquer par l'amplitude du mouvement demandé associé aux problèmes inhérents aux systèmes : tremblement lorsque l'on s'éloigne de l'antenne avec les Gants3D et problème de reconnaissance de l'état de la main avec la Cam3D, avec l'éventualité de sortir du champ de la Kinect. Il en est de même pour le globe et l'horloge. Sur ces deux objets, nous avons également observé qu'il était possible d'avoir l'impression d'un bon positionnement alors qu'en réalité il ne s'agissait que d'un effet de perspective lié aux décalages sur plusieurs axes. Ce problème était moins présent du fait de l'orientation du bateau et de l'horloge. Nous pouvons également remarquer que lorsque le déplacement est associé à une rotation, les résultats obtenus sur l'axe X sont moins bons.

Pour la rotation, l'interprétation des résultats est plus difficile compte tenu de la variabilité de ceux-ci. Par exemple, nous observons des tendances similaires entre les Gants3D et la Cam3D pour les tâches 7 (ordinateur portable) et 9 (bateau) mais avec des résultats opposés pour ces deux tâches : pour la tâche 7, la rotation selon l'axe X (c'est-à-dire gérée par un mouvement en profondeur de la main) est la plus imprécise alors qu'elle se révèle meilleure pour la tâche 9. Ce qui confirme nos suppositions pour cet axe. En revanche, nous pensions que la rotation selon l'axe Z (c'est-à-dire un mouvement de la main vertical) aurait également souffert d'un manque de précision : pour une question de complexité du mouvement. En effet, lorsque l'on souhaite décomposer les mouvements pour isoler la rotation suivant les axes, le mouvement selon l'axe Y (pour une rotation selon l'axe Z) est celui pour lequel nous avons la plus faible amplitude, et il est difficile d'abaisser la main sans avancer le bras vers l'avant, ce qui a pour effet de réaliser une seconde rotation en parallèle. Il est de ce fait surprenant de ne pas observer ce « problème » sur la plupart des tâches mettant en jeu la rotation.

A la vue des résultats précédents, il existe des différences, mais qui sont principalement liées à l'interaction et à la tâche à réaliser. Pour les déplacements, nous pouvons préconiser de trouver une technique permettant de restreindre l'espace de jeu pour éviter au participant de ne plus être dans l'angle de vue de la Kinect. Pour les rotations, la difficulté de bien dissocier les axes peut-être une justification de la variabilité des résultats. Ce point reste à confirmer avec l'analyse des préférences et des commentaires des participants.

### 3.5 Étude de l'impact d'effet d'apprentissage

Nous avons contrebalancé l'ordre de passation des participants sur les 2 modalités (Cam 3D puis Gants3D ou Gants3D puis Cam 3D) pour éviter le biais de l'effet d'apprentissage sur l'ensemble des résultats. Pour obtenir un complément d'information sur notre système, il serait néanmoins intéressant d'étudier s'il existe un effet d'apprentissage, ou de manière plus précise, si on obtient de meilleurs résultats lorsque le système est testé en second. Étant donné que l'interaction est la même pour les 2 systèmes, il serait normal que les participants s'approprient celle-ci et deviennent plus performants avec le deuxième système utilisé, dans le cas idéal. Il reste bien entendu la possibilité que des problèmes inhérents au système testé n'entraînent pas de performances meilleures voire les dégradent.

Nous allons évaluer la présence éventuelle de cet effet sur l'ensemble des variables étudiées précédemment, soient les données globales et les données relatives à chaque tâche. Celles-ci correspondent aux mesures d'erreur (déplacement, redimensionnement, rotation), au temps total de chaque tâche et temps d'exécution pour chaque interaction.

Étant donné les effectifs limités des 2 groupes, soient 10 participants, nous avons choisi d'effectuer des tests non paramétriques. Nous avons donc utilisé le test U de Mann-Whitney pour échantillons indépendants.

#### 3.5.1 Données globales

On observe deux tendances sur les variables du temps total d'expérimentation avec la Cam3D ( $Z = 1,890$  ;  $p = 0,059$ ) et du temps total de manipulation avec la Cam3D ( $Z = 1,814$  ;  $p = 0,075$ ) :

- les participants ayant terminé avec la Cam3D ont eu un temps d'expérimentation plus court (moyenne = 841,4 ; écart-type = 190,5) que ceux ayant commencé avec ce système (moyenne = 1067,4 ; écart-type = 306,2).
- les participants ayant terminé avec la Cam3D ont eu un temps de manipulation plus court (moyenne = 437,7 ; écart-type = 120,4) que ceux ayant commencé avec ce système (moyenne = 607,0 ; écart-type = 220,6).

#### 3.5.2 Données relatives à chaque tâche

Concernant les valeurs d'erreurs de précision, il n'y a qu'une seule différence significative, concernant l'erreur de précision de redimensionnement du cheval (tâche 2) avec la Cam3D ( $Z = 2,419$  ;  $p = 0,015$ ) : les participants ayant terminé avec la Cam3D ont fait une erreur de précision plus faible (moyenne = 1,48 ; écart-type = 1,30) que ceux ayant commencé avec ce système (moyenne = 2,77 ; écart-type = 1,10).

Concernant les temps mesurés, nous obtenons des différences significatives pour certaines variables des tâches 6 (horloge), 7 (ordinateur portable), 9 (bateau) et pour tâche 4 (navigation 1). Les résultats sont synthétisés dans le Tableau 26.

Tableau 26 : Comparatif de l'effet d'apprentissage entre les 2 systèmes. TX = tâche n°x ; C = Cam3D ; G = Gants3D ; T = temps total de la tâche ; TD = temps de déplacement ; TR = temps de rotation ; TE = temps de redimensionnement ; TN = temps de navigation

		T6 : C_T	T6 : C_TR	T7 : C_TE	T4 : C_T	T4 : C_TN
<b>Test M.-W.</b>	<b>Z</b>	2,495	1,663	2,570	1,701	2,080
	<b>P</b>	<b>0,013</b>	<i>0,096</i>	<b>0,009</b>	<i>0,089</i>	<b>0,035</b>
<b>Cam3D en 1er</b>	<b>μ</b>	193,4	63,4	39,3	63,2	44,2
	<b>E.T.</b>	67,6	40,2	17,6	21,7	19,5
<b>Gants3D en 1er</b>	<b>μ</b>	<b>126,2</b>	<b>33,6</b>	<b>18,8</b>	<b>48,5</b>	<b>29,6</b>
	<b>E.T.</b>	30,9	15,2	12,8	13,7	8,9

Sur l'ensemble des valeurs significatives, soient 8 variables, toutes concernent la Cam3D et montrent une amélioration des résultats des participants lorsqu'ils testent la Cam3D en deuxième, après avoir déjà réalisé l'expérimentation avec les Gants3D.

Il reste surprenant d'observer ces différences pour la Cam3D mais pas pour les Gants3D, l'interaction étant identique dans les 2 cas. Ces résultats excluent l'hypothèse de l'élaboration de stratégies pour les tâches demandant de conjuguer plusieurs manipulations (e.g., déplacement + rotation). Lors de l'essai avec la première condition, les participants auraient en effet pu évaluer la meilleure façon de terminer chaque tâche, pour être plus performants lors de la seconde condition. L'hypothèse d'une plus grande difficulté de prise en main avec la Cam3D associée à l'appropriation des modalités d'interaction (commune aux 2 modalités), ne peut être confirmée par les résultats rapportés ici.

Ces résultats, mis en perspective avec ceux décrivant les performances générales (pour la plupart similaires entre les 2 systèmes), suggèrent que la Cam3D, sur certains points, du fait d'une progression entre les 2 conditions, pourrait permettre d'obtenir de meilleures performances que les Gants3D in fine, mais là encore uniquement sur certains points.

Pour conclure, les différences observées ne concernant qu'une faible part des variables étudiées (8 sur 74), les résultats obtenus restent très partiels et ne permettant pas d'établir l'existence d'un effet apprentissage.

### 3.6 Étude du rapport précision / temps

Dans les tests précédents, nous avons évalué séparément les mesures de temps nécessaire et d'erreurs (avec pour cette dernière une distinction selon les 3 types de manipulations permises) afin d'obtenir des valeurs absolues pour chaque système. Dans l'optique d'identifier quel système permet de minimiser le temps de manipulation (et par ce biais, la fatigue), tout en conservant une précision maximale, il est nécessaire de s'intéresser au ratio « précision / temps ». Ce rapport se doit d'être le plus faible possible pour garantir le meilleur compromis entre le temps d'exécution et la précision.

Pour prendre en compte le pourcentage de précision et non le pourcentage d'erreur comme précédemment, nous avons simplement retranché les valeurs précédentes à 100. Nous allons considérer la précision globale (moyenne des 3 axes) de déplacement, de rotation et de redimensionnement (précision homogène sur les 3 axes).

Nous avons effectué le test de normalité de Kolmogorov-Smirnov sur l'ensemble des variables étudiées. Seules 2 variables sur les 24 ne suivent pas une loi normale : le rapport précision / temps pour le déplacement dans la tâche 5 avec les Gants3D ( $Z = 1,421$  ;  $p = 0,035$ ) et pour le déplacement dans la

tâche 9 avec les Gants3D ( $Z = 1,485$  ;  $p = 0,024$ ). Nous avons donc décidé de mener des tests T de Student pour échantillons appariés pour l'ensemble des variables.

Seules 2 valeurs présentent des différences significatives entre les Gants3D et la Cam3D et une valeur présente une tendance. Les résultats sont synthétisés dans le Tableau 27.

**Tableau 27 : Différences significatives entre Cam3D et Gants3D pour le ratio « précision / temps »**

	Déplacement T6		Rotation T7		Déplacement T9	
	Gants3D	Cam3D	Gants3D	Cam3D	Gants3D	Cam3D
<b>Moyenne</b>	<b>3,40</b>	1,98	1,73	<b>3,44</b>	<b>6,26</b>	3,93
<b>Ecart-type</b>	0,96	1,33	1,04	2,68	4,66	2,71
<b>T-Test</b>	4,057		3,471		1,743	
<b>p</b>	<b>0,001</b>		<b>0,003</b>		0,097	

Comme pour l'étude de l'effet d'apprentissage, seules quelques variables présentent des différences significatives (2 sur 24) ou des tendances (1 sur 24). Il n'est donc pas possible de conclure à un quelconque effet du système (Cam3D ou Gants3D) sur le ratio « précision / temps », d'autant plus que les résultats sont soit en faveur de la Cam3D (Rotation Tâche 7), soit en faveur des Gants3D (Déplacements Tâches 6 et 9).

Pour effectuer ce calcul, nous avons appliqué la même pondération pour la précision et le temps. Nous avons demandé aux participants de viser avant tout la précision, sans parler du temps d'expérimentation, afin de ne pas introduire de stress. Pour autant, la dimension temporelle est importante car elle agit directement sur la fatigue des participants, ceux-ci devant travailler, quelque soit le système, avec les 2 bras levés, bien qu'ils soient assis. Globalement, nous avons pu observer que les valeurs de précisions étaient assez élevées dans les différentes tâches, ce qui peut expliquer que les valeurs de ratio soient « lissées ».

### 3.7 Discussion sur les performances

Dans cette première partie relative aux performances, nous avons étudié les performances globales (temps total d'expérimentation, temps total de manipulation) ainsi que les variables relatives à chaque tâche (erreur de précision, temps, séquences d'actions élémentaires). Nous avons également réalisé des études complémentaires. La première, pour déterminer si les déplacements ou les rotations suivant certains axes entraînaient une précision moindre, pour nous permettre d'améliorer notre système. La seconde, afin d'évaluer la présence d'un effet d'apprentissage, et de manière connexe, s'il était possible pour les participants d'améliorer leurs performances avec la pratique afin d'évaluer les perspectives d'évolution dans l'usage. Enfin, pour déterminer lequel des deux systèmes permet d'avoir le meilleur ratio « précision / temps ». L'étude des performances devait permettre de valider ou d'infirmer les hypothèses H1 et H2 :

- H1 : le système Cam3D permet d'obtenir une meilleure précision dans les tâches de manipulation que les Gants3D, du fait de l'absence de port de matériel, donc par l'absence d'entrave des mouvements et de l'absence de poids de l'équipement ;
- H2 : le système Cam3D permet d'obtenir des temps d'exécution au moins similaires au système Gants3D, du fait d'une détection des mouvements aussi précise et aussi fiable.

Les résultats infirment l'hypothèse H1 et confirment l'hypothèse H2. Sur les variables étudiées et sur les tâches que nous avons conçues, la Cam3D ne permet pas d'obtenir une précision supérieure aux

Gants3D (H1), mais permet d'obtenir des temps d'exécution et de manipulation similaires (H2). Le système que nous avons conçu associant le périphérique, les algorithmes de reconnaissance et les modalités d'interaction permet d'obtenir des performances équivalentes à un système couramment utilisé en réalité virtuelle, associant les gants de données et des capteurs électromagnétiques. Ces premières conclusions rejoignent ainsi celles de la première expérimentation, pour une simple tâche de sélection. Il est néanmoins surprenant d'obtenir des performances équivalentes, étant données les différences entre les 2 systèmes. Le fait de ne pas porter d'équipement et de ne pas être gêné dans ses mouvements aurait dû avoir un impact plus important sur les performances. Deux raisons peuvent expliquer l'absence de différence : d'une part un temps d'expérimentation relativement court qui nivelle la différence de fatigue accumulée avec les deux systèmes. D'autre part, des problèmes inhérents à chaque système, comme la sensibilité à la distance entre les capteurs et l'antenne pour les Gants3D, et d'autre part la sensibilité à la reconnaissance de l'état ouvert / fermé des mains.

Nous n'avons observé un effet limité de l'apprentissage, et seulement avec la Cam3D. En effet, les participants ayant terminé avec ce système ont parfois eu de meilleurs résultats que ceux ayant commencé avec ce système. Nous ne pouvons donc pas conclure à un effet d'apprentissage, même si les résultats observés seraient à étudier plus en détails, afin de voir si la progression observée continue dans le temps ou si ce n'est qu'un effet isolé.

Les actions de rotation et de déplacement (ainsi que de navigation) nécessitent de faire des mouvements dans l'espace, selon les dimensions X, Y et Z. Ainsi, nous avons étudié les erreurs de précision selon ces 3 axes, afin d'évaluer si certains déplacements de la main, plus difficiles que d'autres, peuvent pénaliser la précision. En effet, certains mouvements sont plus difficiles à réaliser, comme par exemple monter et descendre le bras verticalement parallèlement à l'axe de l'écran (axe Y) qui présente une amplitude limitée par rapport à un déplacement selon l'axe X par exemple. De même la difficulté de faire un mouvement en profondeur en ramenant la main vers soi : celle-ci est bloquée par le tronc. Les résultats obtenus sont partagés : si pour certaines tâches les mouvements selon l'axe Z sont les moins précis, cette tendance s'inverse pour d'autres tâches. La précision décomposée selon les axes X, Y et Z reste avant tout fortement liée à la tâche et également à la géométrie de l'objet, qui peut offrir des points de repère facilitant le positionnement et/ou l'orientation.

Enfin, l'étude du ratio « précision / temps » ne permet pas d'obtenir suffisamment de différences significatives entre les 2 systèmes au regard du nombre de variables étudiées.

## **4 Influence du périphérique sur les préférences subjectives**

Comme cela a été présenté dans la partie 2.6 (p. 110), nous étudions les préférences subjectives à partir des réponses des participants au questionnaire final, et selon 2 points de vue. D'une part, avec une évaluation des systèmes proposés, au travers d'une étude de l'acceptabilité pour chaque système indépendamment et d'autre part via une analyse des préférences globales, prenant la forme d'une évaluation comparative des systèmes.

### **4.1 Analyse statistique**

L'étude de l'acceptabilité s'appuie sur une étude statistique des réponses données par les participants à des questions posées sous la forme d'échelle de Likert à 5 modalités. Les critères ont été répartis en 3 thèmes :

- Le confort : maniabilité, liberté de mouvement, fatigue, confort d'utilisation, (poids) ;

- L'efficacité : efficacité globale, la précision, la stabilité, la fiabilité ;
- L'efficience : facilité d'utilisation.

Nous avons donc codé les réponses aux échelles de Likert par des entiers compris entre 1 et 5 (1 = appréciation la plus mauvaise ; 5 = appréciation la plus élevée), pour chaque critère et chaque participant. Le poids n'a pas été pris en compte pour la comparaison des 2 systèmes, car il n'est pas évalué pour la Cam3D, l'utilisateur ne portant pas de matériel ; la valeur moyenne sera donnée à titre indicatif. Nous avons également calculé les scores moyens pour chaque thème et chaque participant.

Nous avons en premier lieu effectué un test de normalité sur chaque variable à l'aide du test de Kolmogorov-Smirnov à un échantillon (loi normale). Sur les 18 variables étudiées, 8 suivent une loi normale :

- Gants : fiabilité (Z = 1,352 ; p = 0,052), stabilité (Z = 1,380 ; p = 0,044), facilité (Z = 1,566 ; p = 0,015) ;
- Cam3D : confort (Z = 1,344 ; p = 0,054), maniabilité (Z = 1,789 ; p = 0,003), efficacité (Z = 2,025 ; p = 0,001), fiabilité (Z = 1,352 ; p = 0,052), facilité (Z = 1,464 ; p = 0,027).

Nous avons décidé d'effectuer des tests non paramétriques pour l'ensemble des comparaisons, à l'aide du test de Wilcoxon pour échantillons appariés. Les règles suivantes ont été adoptées concernant la significativité des tests statistiques :

- $p \leq 0,05$  : test significatif ;
- $0,05 < p \leq 0,1$  : tendance ;
- $p > 0,1$  : test non significatif.

Les questions portant sur les préférences globales sont à choix multiples, les réponses étant « Gants », « Cam3D » et « Egal ». Il n'a donc pas été possible de s'appuyer sur les intervalles de confiance utilisés dans la première expérimentation. Nous nous appuyons donc sur une étude statistique descriptive basée sur les comparaisons d'effectifs. Ces données seront illustrées de verbalisations issues des questionnaires finaux.

## 4.2 Etude de l'acceptabilité

Les données sont synthétisées dans le Tableau 28.

**Tableau 28 : Valeurs moyennes pour chaque critère d'étude de l'acceptabilité, d'après les échelles de Likert à 5 modalités (1 = très mauvais ; 5 = très bon)**

		Cam3D		Gants3D	
		moyenne	écart-type	moyenne	écart-type
<b>Confort</b>	<b>Maniabilité</b>	4,00	0,459	3,50	0,761
	<b>Liberté de mouvement</b>	3,55	0,945	3,10	0,718
	<b>Fatigue</b>	2,60	0,754	2,55	0,759
	<b>Confort d'utilisation</b>	3,90	0,788	3,26	0,733
<b>Efficacité</b>	<b>Efficacité globale</b>	3,80	0,768	3,65	0,745
	<b>Précision</b>	3,45	0,759	3,10	0,968
	<b>Stabilité</b>	3,10	0,788	2,65	0,745
	<b>Fiabilité</b>	3,25	0,639	4,25	0,639
<b>Efficience</b>	<b>Facilité d'utilisation</b>	3,80	0,616	3,70	0,571

La « note » moyenne du critère poids, pour les Gants3D, est de 2,55 (écart-type = 0,605).

Il n'y a pas de différence significative entre les 2 systèmes concernant la fatigue ( $Z = 0,302$ ), l'efficacité ( $Z = 1,136$ ), la précision ( $Z = 1,218$ ), la stabilité ( $Z = 1,510$ ), l'efficacité globale ( $Z = 0,286$ ) et la facilité ( $Z = 0,707$ ).

Le confort ( $Z = 2,448$  ;  $p = 0,014$ ), la maniabilité ( $Z = 2,673$  ;  $p = 0,008$ ) et le confort global ( $Z = 2,652$  ;  $p = 0,008$ ) sont significativement meilleurs avec la Cam3D ; on observe de plus une tendance en faveur de ce système pour la liberté de mouvement ( $Z = 1,654$  ;  $p = 0,098$ ).

La fiabilité est significativement meilleure avec le système Gants3D ( $Z = 3,038$  ;  $p = 0,002$ ). La fiabilité est jugée meilleure avec les Gants3D, cela en raison de problèmes de détection de l'état de la main (ouverte / fermée) avec la Cam3D, soulignés par plusieurs participants :

*La Cam3D a des difficultés pour repérer la main ouverte et fermée, et ça peut couper l'action (Sujet 6)*

La liberté de mouvements est jugée supérieure avec la Cam3D car les participants ne portent pas de matériel avec ce système et que le système Gants3D, en plus du matériel proprement dit, impose d'être relié par 4 câbles de connexion à l'ordinateur (2 pour les gants, 2 pour le système de tracking magnétique) :

*On peut déplacer son corps sans gêne contrairement aux gants où nous sommes rattachés par des fils (Sujet 6)*

Le fait que la maniabilité ait été jugée inférieure avec le système Gants3D peut s'expliquer par la taille unique des gants de données, qui ne conviennent pas à toutes les mains, comme cela a été mentionné par un participant :

*Le gant était plus grand que ma main et me faisait de grands doigts qui ne facilitaient pas mes actions (Sujet 10)*

Le poids des gants est également un facteur qui peut influencer la perception de la maniabilité :

*[Avec la Cam3D] on n'a pas le poids des gants qui entrave légèrement la rotation (Sujet 4)*

L'ensemble des scores est représenté pour une lecture simple sur la Figure 45. Les critères encadrés sont ceux pour lesquels la différence est significative ou illustre une tendance (en bleu).

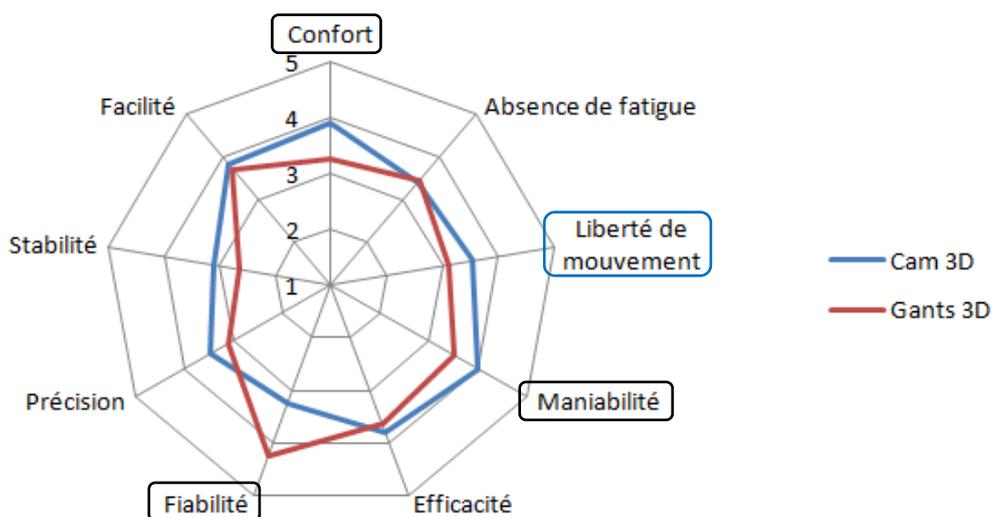


Figure 45 : représentation des scores moyens de chaque critère pour les 2 systèmes étudiés. Les variables pour lesquelles les différences sont significatives sont encadrées (en bleu, tendances).

Le graphique représenté sur la Figure 45 permet de bien mettre en évidence les différences entre les 2 systèmes, avec une supériorité de la Cam3D pour le confort, la maniabilité et la liberté de mouvement, en revanche la fiabilité est meilleure avec les Gants3D. Il est surprenant de constater que la fatigue ressentie est équivalente entre les deux systèmes. Cela implique que le poids des gants et des capteurs magnétiques est négligeable par rapport à la fatigue engendrée par la position des bras tendus, sans support pour se reposer. L'acceptabilité est meilleure avec la Cam3D, ce qui est justifié par les commentaires et jugements subjectifs des participants. Le point négatif, illustré par la fiabilité, correspond à la mauvaise reconnaissance de l'état de la main (ouverte ou fermée) qui peut se produire parfois.

### 4.3 Étude des préférences globales

L'étude des préférences globales repose sur l'évaluation comparative des systèmes selon les 5 critères suivants :

- La préférence générale ;
- L'adéquation aux tâches à réaliser ;
- Le sentiment d'immersion ;
- La précision ;
- L'aisance d'utilisation.

Les participants pouvaient choisir les gants, la Cam3D ou ne pas émettre de préférence (« égal »).

Le Tableau 29 présente les effectifs de participants ayant choisi chaque possibilité pour chaque critère. Ces résultats sont représentés sur le Figure 46.

**Tableau 29 : Effectifs de participants selon la préférence attribuée à chaque critère**

	Gants	Cam3D	Egal
<b>Préférence générale</b>	3	12	5
<b>Adéquation aux tâches</b>	1	13	6
<b>Sentiment d'immersion</b>	2	8	10
<b>Précision</b>	6	11	3
<b>Aisance d'utilisation</b>	5	9	6

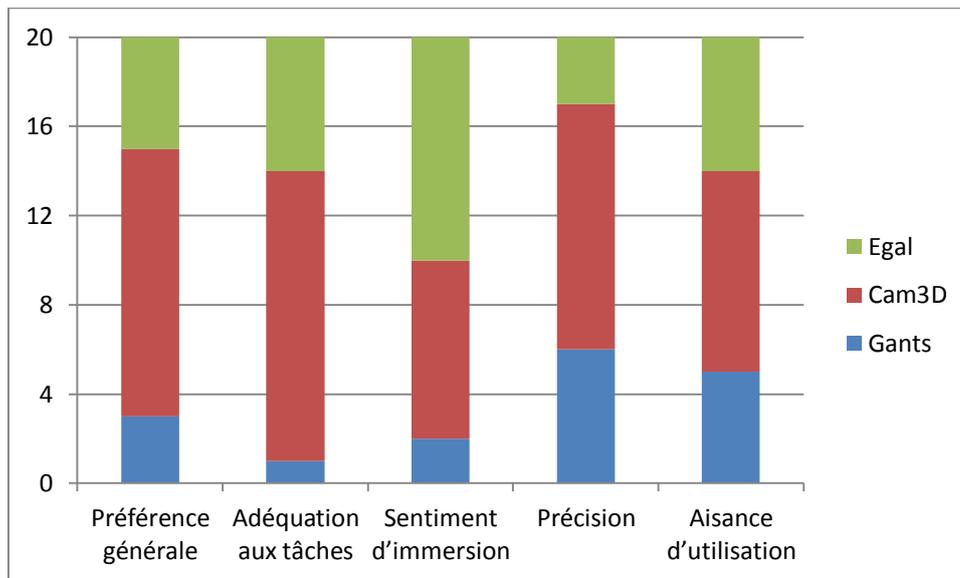


Figure 46 : Histogramme empilé de la répartition des participants selon le choix du participant pour chaque critère étudié

Le système Gants3D est moins apprécié que le système Cam3D pour l'ensemble des critères étudiés (de 1 à 6 participants sur les 20, contre 8 à 13 pour la Cam3D).

Pour 3 des 5 critères (préférences générales, adéquation aux tâches et précision), plus de la moitié des participants donnent leur préférence au système Cam3D.

Concernant l'aisance d'utilisation, cela rejoint les résultats de la partie précédente sur les critères de maniabilité, liberté de mouvements et maniabilité. La Cam3D semble apporter un avantage significatif sur ce point par rapport aux Gants3D.

Il apparaît clairement, même s'il convient de nuancer au regard du nombre de participants, que la Cam3D apporte un sentiment d'immersion égal voire supérieur à celui provoqué par les Gants3D (8 contre 2, 10 sans préférences). L'interaction étant identique entre les 2 systèmes étudiés, nous pouvons avancer l'hypothèse que l'absence de port de matériel (et donc l'absence de poids au niveau des mains et l'aisance de mouvement) est de nature à renforcer le sentiment d'immersion. La faible portée du système de tracking magnétique (qui tremble lorsque trop éloigné de l'antenne) était également de nature à perturber l'engagement du participant dans la tâche et donc à diminuer le sentiment d'immersion.

La répartition des participants quant à la précision traduit une faiblesse des Gants3D sur ce point particulier. Cette observation est corroborée par les verbalisations des participants :

*Il y a trop de trop de tremblement avec les gants, c'est beaucoup moins précis (Sujet 13)*

La préférence générale des participants est nettement en valeur de la Cam3D, puisque sur les 20 participants, 12 choisiraient la Cam3D, contre seulement 3 les Gants (et 5 ne donnent pas de préférence particulière).

#### 4.4 Synthèse

L'étude de l'acceptabilité et des préférences globales des participants, étudiées d'après une analyse statistique (tests de Wilcoxon) d'une part et d'une analyse statistique descriptive d'autre part, permet de conclure à la supériorité de la Cam3D sur les Gants3D. Ce dernier système apporte néanmoins une meilleure fiabilité, ce qui tend à nuancer légèrement le constat précédent, en raison d'une meilleure

reconnaissance de la main ouverte ou fermée. En effet, les gants fonctionnent qu'elle que soit l'orientation et la position de la main, contrairement à toute caméra, 2D ou 3D. Et il reste difficile pour les participants d'avoir la perception d'être en dehors du champ de la Kinect ou avec la main mal orientée, car ils sont absorbés par la tâche à réaliser. Enfin, il faut noter que la Cam3D apporte un meilleur confort, une plus grande liberté de mouvements et une meilleure maniabilité, ces deux derniers critères faisant référence à l'absence de fils, de capteurs et de gants sur les mains des participants. Ces résultats viennent confirmer les hypothèses H3 et H4, concernant l'acceptabilité et les préférences générales.

## 5 Évaluation des modalités d'interaction proposées

Au-delà de l'étude de l'acceptabilité des systèmes, nous nous sommes intéressés à l'évaluation des modalités d'interaction et de l'adéquation entre le périphérique et celles-ci. Ce second axe d'analyse permettra d'alimenter l'étude complémentaire portant sur l'évaluation de nouvelles modalités d'interaction en réponses aux commentaires et jugements des participants. Les résultats présentés dans cette partie permettront d'affiner les résultats précédents et d'apporter des explications sur les préférences des participants, en faisant le lien entre le système et l'interaction.

### 5.1 Analyse statistique

L'évaluation de chaque modalité d'interaction (déplacement, navigation, rotation, redimensionnement) a été réalisée par des échelles de Likert à 5 modalités pour 5 critères distincts : l'efficacité globale, la rapidité, la précision, la simplicité et l'intuitivité de l'action. Nous allons nous appuyer sur des statistiques descriptives (moyenne, écart-type, médiane) pour évaluer ces 5 critères. Ensuite, nous opposerons les 4 modalités pour ces 5 critères. Pour réaliser cette étude, nous avons choisi de nous baser sur des tests non paramétriques pour échantillons appariés ; étant donné que nous évaluons 4 modalités, nous nous appuyerons sur le test de Friedman et si les résultats s'avèrent significatifs, sur le test de Wilcoxon pour chaque paire.

Concernant l'adéquation entre les systèmes et les modalités d'interaction, il a été demandé aux participants de choisir le système (Gants, Cam3D) qui d'une part permettait d'atteindre la meilleure précision et d'autre part qui offrait la meilleure facilité d'utilisation, pour chaque interaction possible (déplacement, navigation, redimensionnement, rotation). Les participants pouvaient également marquer une absence de préférence pour l'un ou l'autre système (« Egal »). Comme pour les préférences globales, nous réaliserons une étude statistique descriptive basée sur les comparaisons d'effectifs, illustrée de verbalisations issues des questionnaires finaux.

### 5.2 Évaluation des modalités d'interaction

Pour une meilleure lisibilité des résultats, nous avons réparti les réponses en 3 groupes plutôt que de considérer les 5 modalités des réponses aux échelles de Likert pour chaque critère : « Bon – Très bon », « Moyen », « Mauvais – Très mauvais ». Les effectifs sont présentés sous forme d'histogrammes empilés et commentés.

### 5.2.1 Efficacité

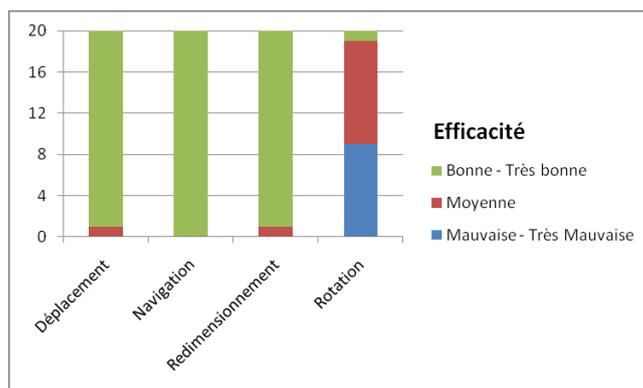


Figure 47 : Histogramme empilé de la répartition des participants selon leur évaluation de l'efficacité pour les 4 modalités d'interaction

L'efficacité du déplacement a été jugée bonne ou très bonne par 19 des 20 participants (95 %). L'efficacité de la navigation a été jugée bonne ou très bonne par l'ensemble des participants (100 %). L'efficacité du redimensionnement a été jugée bonne ou très bonne par 19 des 20 participants (95 %). Seul un participant a jugé bonne l'efficacité de la rotation (5 %), la moitié des participants l'ont jugée moyenne et 9 participants l'ont jugé mauvaise ou très mauvaise (45 %).

### 5.2.2 Rapidité

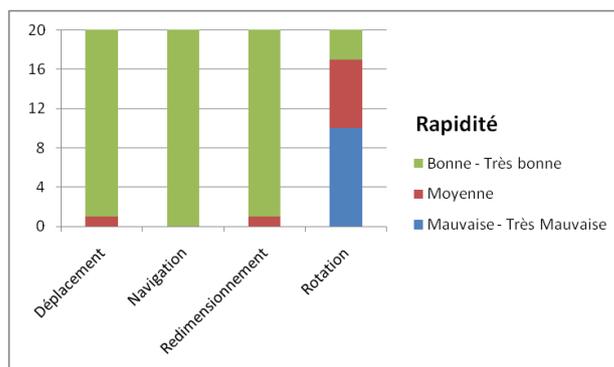


Figure 48 : Histogramme empilé de la répartition des participants selon leur évaluation de la rapidité pour les 4 modalités d'interaction

La rapidité du déplacement a été jugée bonne ou très bonne par 19 des 20 participants (95 %). La rapidité de la navigation a été jugée bonne ou très bonne par l'ensemble des participants (100 %). La rapidité du redimensionnement a été jugée bonne ou très bonne par 19 des 20 participants (95 %). Seuls 3 participants ont jugé bonne la rapidité de la rotation (15 %), 7 participants l'ont jugée moyenne (35%) et la moitié participants l'ont jugé mauvaise ou très mauvaise (50 %).

### 5.2.3 Précision

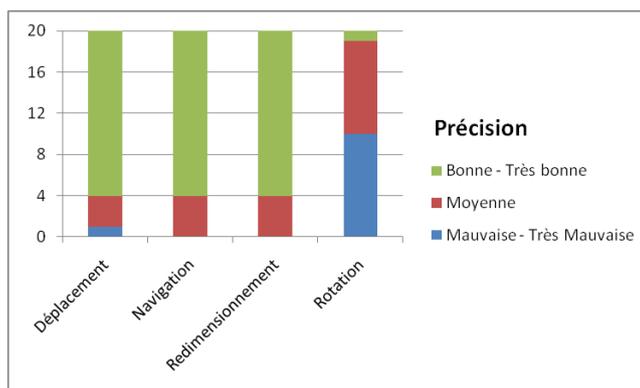


Figure 49 : Histogramme empilé de la répartition des participants selon leur évaluation de la précision pour les 4 modalités d'interaction

La précision du déplacement a été jugée bonne ou très bonne par 16 des 20 participants (80 %). La précision de la navigation a été jugée bonne ou très bonne par 16 des 20 participants (80 %). La précision du redimensionnement a été jugée bonne ou très bonne par 19 des 20 participants (95 %). Seul un participant a jugé bonne la précision de la rotation (5 %), 9 participants l'ont jugée moyenne (45%) et la moitié des participants l'ont jugée mauvaise ou très mauvaise (50 %).

### 5.2.4 Simplicité

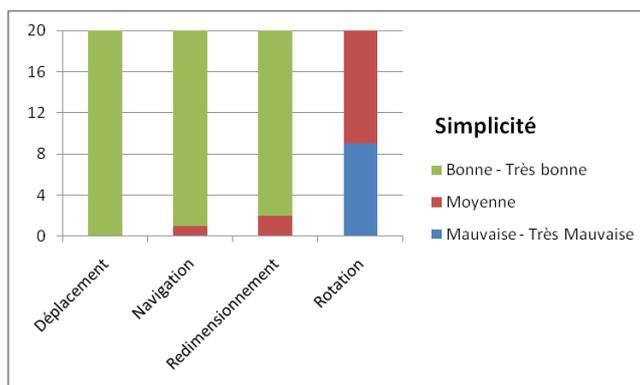


Figure 50 : Histogramme empilé de la répartition des participants selon leur évaluation de la simplicité pour les 4 modalités d'interaction

La simplicité du déplacement a été jugée bonne ou très bonne par l'ensemble des participants (100 %). La simplicité de la navigation a été jugée bonne ou très bonne par 19 des 20 participants (95 %). La simplicité du redimensionnement a été jugée bonne ou très bonne par 18 des 20 participants (90 %). 11 participants ont jugé moyenne la simplicité de la rotation (55 %) et 9 participants l'ont jugée mauvaise ou très mauvaise (45 %).

### 5.2.5 Intuitivité

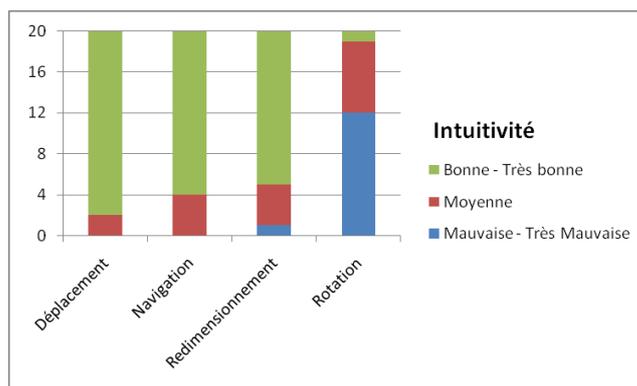


Figure 51 : Histogramme empilé de la répartition des participants selon leur évaluation de l'intuitivité pour les 4 modalités d'interaction

L'intuitivité du déplacement a été jugée bonne ou très bonne par 18 des 20 participants (90 %). L'intuitivité de la navigation a été jugée bonne ou très bonne par 16 des 20 participants (80 %). L'intuitivité du redimensionnement a été jugée bonne ou très bonne par 15 des 20 participants (75 %). Seul un participant a jugé bonne l'intuitivité de la rotation (5 %), 7 participants l'ont jugée moyenne (35%) et 12 participants l'ont jugée mauvaise ou très mauvaise (60 %).

### 5.2.6 Synthèse

Pour synthétiser ces résultats, nous avons codé les réponses aux échelles de Likert par des entiers compris entre 1 et 5 (1 = appréciation la plus mauvaise ; 5 = appréciation la plus élevée), pour chaque critère et chaque modalité d'interaction. Les résultats présentés dans sur la Figure 52 correspondent aux moyennes des participants.

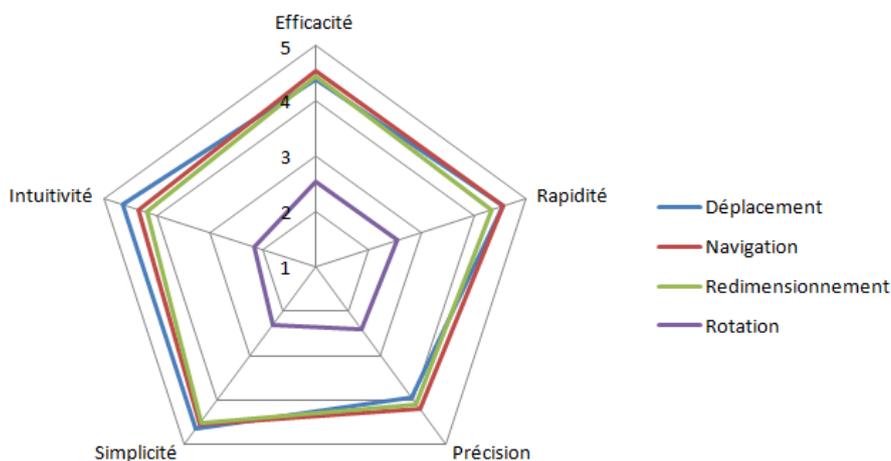


Figure 52 : représentation des scores moyens de chaque critère pour les 5 modalités d'interaction

La faiblesse de la rotation comparée aux autres modalités d'interaction est confirmée par les tests statistiques, synthétisés dans le Tableau 30.

Tableau 30 : Oppositions entre les modalités d'interaction pour chaque critère

	Test de Friedman	Test de Wilcoxon paire à paire		
		Rotation vs Navigation	Rotation vs Navigation	Rotation vs Redimensionnement
<b>Efficacité</b>	Khi2 = 41,525 p < 0,001	Z = 3,881 p < 0,001	Z = 3,897 p < 0,001	Z = 3,886 p < 0,001
<b>Rapidité</b>	Khi2 = 42,765 p < 0,001	Z = 3,863 p < 0,001	Z = 3,880 p < 0,001	Z = 3,778 p < 0,001
<b>Précision</b>	Khi2 = 36,264 p < 0,001	Z = 3,789 p < 0,001	Z = 3,831 p < 0,001	Z = 3,695 p < 0,001
<b>Simplicité</b>	Khi2 = 43,733 p < 0,001	Z = 3,953 p < 0,001	Z = 3,973 p < 0,001	Z = 3,854 p < 0,001
<b>Intuitivité</b>	Khi2 = 42,775 p < 0,001	Z = 3,852 p < 0,001	Z = 3,858 p < 0,001	Z = 3,689 p < 0,001

Il apparaît que la rotation est significativement inférieure aux 3 autres modalités d'interaction, que ce soit en termes d'efficacité, de rapidité, de précision, de simplicité ou d'intuitivité. De plus, le déplacement est jugé plus intuitif que la navigation ( $Z = 1,897$  ;  $p = 0,058$ ) et que le redimensionnement ( $Z = 2,310$  ;  $p = 0,021$ ).

Deux raisons principales justifient la faiblesse de la rotation selon les participants :

- L'aspect « inhabituel » de l'action à réaliser : « *Difficulté à l'apprentissage initial, car rare utilisation des trois dimensions en vie courante* » (Sujet 9). Cette raison a été mentionnée par 6 autres participants.
- La difficulté d'associer un mouvement à une rotation particulière : « *Difficile de savoir quel mouvement permet de tourner autour de quel axe* » (Sujet 20). Cette raison a été mentionnée par 4 autres participants.

Seul un participant a trouvé cette modalité intuitive (sujet 8).

Le déplacement a été jugé efficace et intuitif par l'ensemble des 20 participants : « *Très intuitif, simple et efficace* » (Sujet 13). Les commentaires négatifs sont davantage généraux et concernent la stabilité des systèmes et la qualité de la reconnaissance de l'état ouvert / fermé de la main.

De la même façon, le redimensionnement a été évalué de manière positive par la majeure partie des participants, 16 d'entre eux le jugeant « *intuitif, rapide et précis* » (sujet 3). Un participant a tout de même mentionné qu'il était « *difficile de se dire qu'il faut d'abord positionner les deux mains avant* ».

Enfin, la navigation a également été jugée « *très intuitive et très rapide* » (sujet 15) par 12 participants ; le « *ressenti ludique* » (sujet 14) a été mis en avant par 4 participants. Pourtant, les participants ont relevé que cette interaction était trop sensible (2 participants), « *peu précise sur de petites distances* » (sujet 14) et enfin que la « *vitesse [était] parfois trop élevée* » (sujet 4), cela étant mentionné par 5 participants.

Le fait que le déplacement ait été jugé plus intuitif que les autres modalités d'interaction peut s'expliquer par le fait que c'est la seule modalité d'interaction qui n'a pas recours à une métaphore d'interaction (joystick pour la navigation, étirer/comprimer pour le redimensionnement) ou une adaptation (décomposition x, y, z pour la rotation). Le mouvement du déplacement est identique à celui à exercer dans un environnement réel.

### 5.3 Évaluation des modalités d'interaction

Le Tableau 31 présente les effectifs de participants ayant choisi chaque possibilité pour chaque critère et chaque modalité d'interaction. Ces résultats sont représentés sur la Figure 53.

Tableau 31 : Effectifs de participants selon la préférence accordée pour modalité d'interaction et chaque critère

		Gants	Cam3D	Egal
<b>Déplacement</b>	Précision	2	10	8
	Facilité d'utilisation	4	9	7
<b>Navigation</b>	Précision	2	2	16
	Facilité d'utilisation	1	2	17
<b>Redimensionnement</b>	Précision	1	6	13
	Facilité d'utilisation	3	6	11
<b>Rotation</b>	Précision	2	10	8
	Facilité d'utilisation	6	9	5

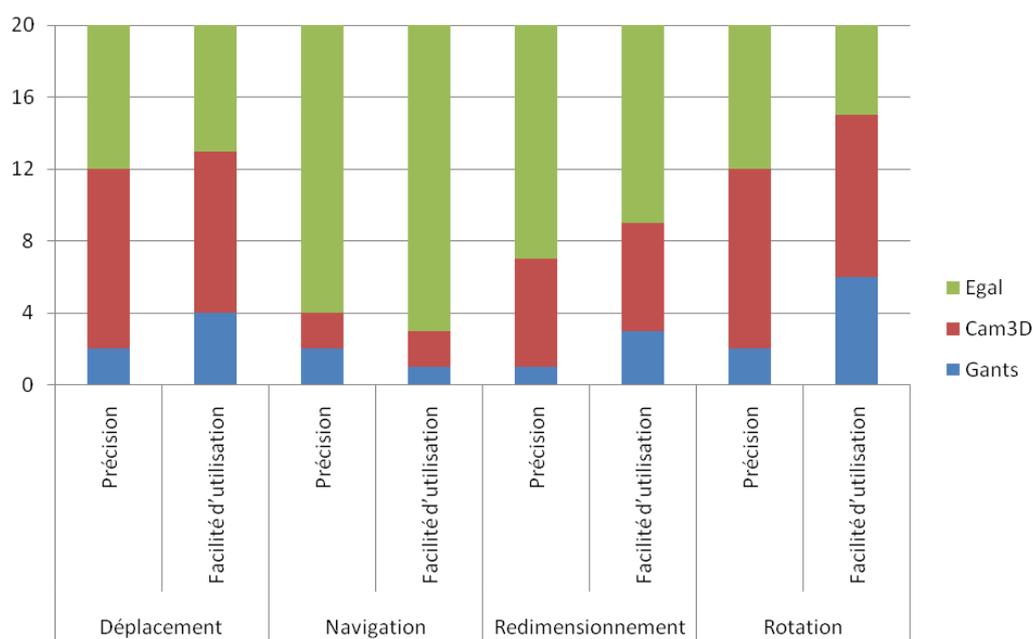


Figure 53 : Histogramme empilé de la répartition des participants selon le choix du participant pour chaque critère étudié et chaque modalité d'interaction

Plus de 75 % des participants n'indiquent pas de préférence pour l'un ou l'autre système concernant la précision et la facilité d'utilisation de la navigation (respectivement 80 % et 85 %).

Plus de 50 % des participants n'indiquent pas de préférence pour l'un ou l'autre système concernant la précision et la facilité d'utilisation du redimensionnement (respectivement 65 % et 55 %).

Les participants ont une préférence plus marquée pour la Cam3D concernant la précision et la facilité d'utilisation du déplacement (respectivement 50 % et 45%), ainsi que pour la rotation (respectivement 50 % et 45%).

De manière générale, la détection de l'état « ouvert » ou « fermé » de la main est un problème rencontré avec la Cam3D, et non avec les Gants3D. Cela entache la facilité d'utilisation et la précision :

« Les gants offrent une meilleure détection de l'ouverture / fermeture » (sujet 15).

A contrario, la Cam3D ne connaît pas les problèmes de tremblement dus à la sensibilité aux perturbations électromagnétiques du système de tracking de la position des gants. Le volume de travail se trouve également réduit. Cela gêne la précision et la facilité d'utilisation :

« *Trop de tremblements avec les gants, beaucoup moins précis* » (sujet 13), mentionné par 7 participants  
« *Avec les gants, la précision était moins bonne si on s'éloigne du capteur* » (sujet 10), mentionné par 2 participants.

Deux raisons peuvent expliquer que le déplacement et la rotation sont jugés supérieurs avec la Cam3D : l'amplitude des déplacements à effectuer et la précision nécessaire. En effet, les Gants3D induisent des tremblements et une perte de précision lorsque l'on s'éloigne de l'antenne. Dans le cas de la navigation, l'amplitude des déplacements est très limitée du fait de la métaphore utilisée (joystick). Dans le cas du redimensionnement, la possibilité de débrayer et de recommencer plusieurs fois l'agrandissement, fait que l'amplitude des gestes peut être petite et que les mains peuvent rester proches de l'antenne. Les tremblements peuvent être, ainsi, minimisés et apporter une meilleure précision. Cela n'est pas le cas du déplacement, qui nécessite des mouvements amples avec une précision importante en fin de geste. Ni de la rotation, qui demande également une forte précision en fin de geste, de plus, bien que le débrayage soit possible, il est moins aisé qu'avec le redimensionnement.

Ces résultats viennent affiner les résultats précédents concernant les préférences des participants au regard des critères de la précision et de l'aisance d'utilisation (voir partie « préférences globales »). En effet, 11 participants préféraient la Cam3D pour la précision contre seulement 6 qui préféraient les Gants. Cette préférence peut donc s'expliquer par la supériorité de la Cam3D pour les tâches de rotation et de déplacement.

Nous pouvons formuler le même commentaire pour l'aisance d'utilisation, attribuée à la Cam3D par 9 participants, contre 5 pour les Gants. Cela est là encore motivé par une supériorité de la rotation et du déplacement avec les Cam3D, les autres interactions étant équivalentes entre les 2 systèmes.

## 5.4 Synthèse

Dans un premier temps, nous avons évalué les modalités d'interaction, indépendamment des systèmes comparés. Il apparaît clairement que la rotation est inférieure aux autres modalités d'interaction, du fait de la complexité des actions à réaliser. La rotation est, avec le déplacement, la modalité d'interaction qui demande de conjuguer les 3 axes de l'espace. Pour autant, la complexité est loin d'être identique : pour le déplacement, la transposition entre le réel et le virtuel est immédiat, étant donné que le geste a été directement reproduit. Par contre, bien que faire tourner un objet sur lui-même nous paraisse anodin dans la vie de tous les jours, notre stratégie est totalement liée à la forme de l'objet et à sa taille et ce geste fait bien souvent intervenir les deux mains dans un enchaînement assez complexe de prises. Il n'est pas possible de reproduire cette complexité dans le virtuel sans retour d'effort. Il existe de nombreuses métaphores d'interaction tendant à simplifier ce problème, mais elles ne sont pas adaptées (e.g., le globe terrestre). L'interaction que nous avons choisie, correspondant à la décomposition et à l'isolation suivant les 3 axes de rotation, car il nous est difficile de réaliser un mouvement purement rectiligne. Ainsi, les rotations sont souvent effectuées selon 2 ou 3 axes, alors que les participants préféreraient modifier un axe à la fois. Se pose également le problème de l'assimilation de l'association mouvement / axe de rotation, mais cette assimilation est assez rapide d'après les participants. Il apparaît donc assez clairement que la méthode choisie pour la rotation reste encore perfectible. De manière générale, nous pouvons donc affirmer que la rotation n'est pas à la hauteur des 3 autres modalités d'interaction, cela s'expliquant par la complexité de reproduire ce mouvement en virtuel et

par le fait que le système que nous avons créé ne permet pas encore de connaître l'orientation des mains.

Dans un second temps, nous avons évalué l'adéquation entre les systèmes et les modalités d'interaction. Nous avons observé des préférences plus marquées pour la Cam3D concernant la précision et la facilité d'utilisation de la rotation et du déplacement. Cette observation est justifiée par la plus grande sensibilité de ces interactions aux tremblements rencontrés avec les Gants3D. Ce problème se rencontre effectivement dès lors que l'on s'éloigne de l'antenne, c'est-à-dire pour des mouvements de grande amplitude. L'amplitude de ceux-ci étant limitée pour le redimensionnement et la navigation. Bien que nous n'ayons pas observé de différences significatives dans les résultats de performances, nous pouvons remarquer que certains problèmes inhérents au matériel ont suscité de nombreux commentaires des participants et des avis assez tranchés. La Cam3D recueille alors davantage de suffrages que les Gants3D.

## 6 Suggestions d'améliorations

Les suggestions d'amélioration portent sur les remarques et les critiques d'ordre général ou à une modalité d'interaction en particulier.

De manière générale, les participants ont mis en avant un « *long temps d'attente pour la sélection de l'objet* » (sujet 15) et suggèrent donc de diminuer les temps de validation, sans proposer de solution concrète.

L'instabilité du système, particulièrement avec les Gants3D, rendent difficiles les tâches demandant de la précision, comme par exemple le déplacement. Plusieurs solutions sont proposées par les participants :

« *Varié la vitesse [de déplacement] en fonction de la distance* » (sujet 5)

« *Utiliser la seconde main pour valider la fin de déplacement* » (sujet 7)

Ces solutions pourraient s'appliquer également au redimensionnement et à la rotation.

Concernant la rotation, les avis sont partagés : si 6 participants ont apprécié « *l'accès aux 3 axes sans menu supplémentaire* » (sujet 1), 12 participants ont connu des difficultés, principalement concernant « *l'apprentissage initial* » (sujet 9) et la difficulté d'associer le mouvement réel aux rotations effectuées par les objets. Une solution a été proposée par 7 participants :

« *Mettre une aide pour savoir quel mouvement correspond à quel axe* » (sujet 4).

Un autre participant (sujet 14) propose un « *modèle d'interaction basé sur le choix d'un seul axe avec verrouillage des deux autres* » ce qui semble aller à l'encontre de l'intérêt de l'accès immédiat aux 3 axes. Enfin, un participant suggère de « *gérer la rotation en tournant la main* » (sujet 13).

Le redimensionnement a été jugé positivement par l'ensemble des participants. Pourtant, certains ont également mentionné des difficultés lorsque, par exemple, ils avaient débuté le mouvement de diminution avec les mains initialement trop proches. Ils finissaient donc avec les 2 mains en contact et étaient donc obligés de « *débrayer* » puis de repositionner leurs mains et de recommencer. Nous avons relevé 2 suggestions pour pallier ce problème :

« *Une seule main gère l'échelle dans un axe (profondeur ?) et l'autre main sert de on/off pour l'action* » (sujet 14)

« Peut-être faire une échelle exponentielle c'est-à-dire que quand les deux mains sont collées l'objet continue à diminuer de taille pour éviter de devoir relâcher » (sujet 20)

La navigation a été majoritairement évaluée comme rapide et intuitive par les participants. La vitesse a été jugée « parfois trop élevée » (sujet 4) par 5 participants, et la trop grande sensibilité a été notée par 2 participants. Un participant propose de « varier la vitesse en fonction de l'éloignement de la trajectoire idéale » (sujet 5), un autre de « réduire la proportionnalité entre déplacement main / déplacement virtuel » (sujet 16).

Ces différents commentaires et suggestions serviront de base à la construction de la troisième étude, étude complémentaire portant plus spécifiquement sur l'interaction.

## 7 Conclusions générales

### 7.1 Synthèse des résultats

Nous avons évalué les performances et les préférences subjectives ainsi que les commentaires des participants. Nous allons synthétiser dans cette partie l'ensemble des résultats présentés précédemment. Nous ne reviendrons pas sur les interprétations qui ont été formulées dans les parties correspondantes.

Les performances reposent sur l'étude des temps et de l'erreur de précision. Les temps mesurés correspondent aux temps d'expérimentation, aux temps d'exécution de chaque tâche et aux temps de manipulation différenciés selon chaque type d'interaction. L'erreur de précision est calculée pour chaque objet et chaque tâche, et différenciée selon chaque interaction le cas échéant. Elle correspond à un pourcentage de décalage entre l'état final de l'objet et l'état idéal. Nous n'avons pas noté de différence significative entre les deux systèmes pour le temps total d'expérimentation et le temps total de manipulation. L'analyse des données différenciées sur chacune des 9 tâches permet de déceler quelques différences significatives, mais celles-ci restent peu nombreuses au regard de l'ensemble des variables étudiées. Nous observons néanmoins une tendance, qu'il conviendrait d'étudier plus en détails pour l'infirmier ou la confirmer : les Gants3D semblent être à même de minimiser le temps d'exécution pour les actions de redimensionnement.

Nous avons également mené des analyses complémentaires portant sur trois points : l'étude de la présence éventuelle de différences de précision selon les axes pour le déplacement et la rotation, l'étude de l'effet d'apprentissage éventuel et enfin du ratio « précision / temps ». Il existe des différences de précision selon les axes et le type de manipulation à effectuer, mais nous avons pu remarquer qu'elles sont avant tout liées à la tâche à réaliser et à l'objet à manipuler (présence de point de repère visuel par exemple). De manière globale, nous observons des précisions plus faibles pour les axes Z et Y, pour lesquels les mouvements des participants se trouvent être les plus restreints ou contraints. De manière générale, sur l'ensemble des variables étudiées, nous n'avons pas observé d'effet d'apprentissage. Pour autant, nous avons observé que dans certains cas, les résultats obtenus avec la Cam3D pouvaient être meilleurs lorsque ce système était testé en second. Enfin, le ratio « précision / temps », pouvant être vu comme un compromis entre la précision à atteindre et le temps mis pour réaliser la tâche, n'est pas significativement différent entre les 2 systèmes, mis à part pour quelques variables isolées.

Les réponses des participants au questionnaire final nous ont permis d'étudier l'acceptabilité des 2 systèmes comparés, les préférences globales des participants, les modalités d'interactions et

l'adéquation entre les systèmes et les modalités d'interaction. L'acceptabilité est globalement meilleure avec la Cam3D, avec des différences significatives en faveur de ce système pour le confort, la liberté de mouvements et la maniabilité. Les Gants 3D se révélant seulement plus fiables, en raison de l'absence d'erreur de reconnaissance de l'état de la main. La Cam3D offre également un meilleur sentiment d'immersion que les Gants3D, d'après les participants. L'analyse de l'évaluation des modalités d'interaction a fait apparaître une faiblesse de la rotation comparées aux autres modalités, provenant de la difficulté à appréhender les gestes à effectuer sur les 3 axes de rotation. La rotation d'un objet sur lui-même, qui apparaît triviale dans la vie réel est très difficile à mettre en œuvre dans un environnement virtuel sans retour haptique et donc sans l'aide des propriétés géométriques de l'objet. Les commentaires des participants nous ont fourni certaines pistes d'amélioration. Enfin, concernant l'adéquation entre les systèmes et les modalités d'interaction, les participants ont manifesté des préférences plus marquées pour la Cam3D concernant la précision et la facilité d'utilisation de la rotation et du déplacement, en raison de l'absence des tremblements qui se manifestent avec les Gants3D lorsque le capteur électromagnétique est trop distant de l'antenne.

Une dernière partie de cette étude a été axée sur l'analyse des améliorations proposées par les participants. Celles-ci portent majoritairement sur des réponses aux problèmes d'instabilité et de difficulté de compréhension de la rotation.

## 7.2 Conclusion

Dans ce chapitre, nous avons présenté une expérimentation menée auprès de 20 participants, tous utilisateurs d'ordinateurs et pour la grande majorité possédant un niveau d'expertise avancé en réalité virtuelle, visant à démontrer l'intérêt d'un système de capture de mouvement de la main sans marqueur. Le système est ici entendu comme un système complet composé du périphérique (Kinect), des algorithmes de traitement et de détection ainsi que des modalités d'interaction développées. L'objectif sous-jacent et secondaire était d'évaluer ces dernières pour être en mesure *in fine* de disposer d'un système efficace, efficient et confortable.

Nous avons donc fait le choix d'opposer l'ensemble « Camera 3D + algorithmes de détection de la position des mains et de leur état » à une solution commerciale existante et fonctionnellement équivalente, les gants de données (pour détecter l'état des mains) associés aux capteurs électromagnétiques (pour connaître les positions des mains). Afin d'obtenir des résultats comparables, nous avons utilisé les mêmes modalités d'interaction entre les deux systèmes. Après nous être basé sur une tâche simple de sélection dans la première étude, nous avons choisi de considérer ici l'ensemble des tâches de manipulation d'un objet non déformable (déplacement, rotation et redimensionnement) ainsi que la tâche de navigation dans un environnement virtuel. C'est ce qui justifie par ailleurs l'usage d'un périphérique plus évolué que la souris.

Notre objectif principal et notre objectif secondaire nous ont amené à formuler les hypothèses de travail suivantes :

- H1 : le système Cam3D permet d'obtenir une meilleure précision dans les tâches de manipulation que les Gants3D, du fait de l'absence de port de matériel, donc par l'absence d'entrave des mouvements et de l'absence de poids de l'équipement ;
- H2 : le système Cam3D permet d'obtenir des temps d'exécution au moins similaires au système Gants3D, du fait d'une détection des mouvements aussi précise et aussi fiable ;

- H3 : le système Cam3D doit recueillir un meilleur niveau d'acceptabilité globale que les Gants3D, en raison d'un meilleur confort (pas d'équipement, pas de poids à porter/supporter), d'une meilleure efficacité (majoritairement en termes de précision et de stabilité ressentie) et d'une meilleure efficacité ;
- H4 : les préférences des participants sont en faveur de la Cam3D, du fait du plus haut degré d'immersion supposé, ainsi que d'une plus grande aisance d'utilisation.

Ces hypothèses nous ont amené à mesurer d'une part la performance des participants pour les 9 tâches à réaliser et différents critères globaux (comme les temps) ou relatifs à chaque tâche (temps, précision, ratio « précision / temps). D'autre part, nous avons évalué les préférences subjectives des participants vis-à-vis de l'expérimentation et confronté leurs commentaires et suggestions d'améliorations au travers d'un questionnaire post-passation.

Il y a un lien entre les hypothèses H1 et H3. En effet, l'absence de port de matériel et par contrecoup de lien filaire entre les périphériques et l'ordinateur, devait avoir une conséquence sur les performances et les préférences. Sur les performances, puisque l'absence de poids et d'entrave aurait pu permettre de minimiser la fatigue et maximiser la précision. Sur les préférences, parce que les participants auraient pu ressentir un meilleur confort d'utilisation et une meilleure efficacité.

Les résultats montrent que les performances sont similaires entre les deux systèmes étudiés, pour la précision. Cela infirme l'hypothèse H1, car nous pensions que l'absence de port de matériel (et donc l'absence de poids) était de nature à minimiser la fatigue et donc l'imprécision. Deux explications peuvent expliquer ce résultat : la première pourrait être que le temps d'expérimentation est insuffisant pour induire une fatigue susceptible d'influer sur les résultats ; la seconde est que la fatigue induite par le port du matériel est négligeable par rapport à celle induite par la position des bras durant l'expérimentation (sans support de repos). Les temps d'exécution, pris de manière globale ou différenciés selon les tâches, sont également similaires, ce qui valide l'hypothèse H2. Les commentaires des participants permettent d'apporter un éclairage supplémentaire à ces explications, tout en apportant un léger bémol à H2. En effet, les participants ont évoqué un problème de fiabilité dans la reconnaissance de l'état de la main avec la Cam3D, et un problème de tremblements de l'avatar avec les Gants3D dès que la main était trop loin de l'antenne (tracking magnétique). Ces deux points négatifs, inhérents aux solutions techniques retenues, ont pu lisser les résultats de performances. Ainsi, si la détection des mouvements proprement dite (déplacements dans l'espace) avec la Cam3D est d'une précision équivalente à celle obtenue par les Gants3D, la fiabilité de la reconnaissance de la main fermée – main ouverte reste inférieure. Ce problème pourrait être contourné par l'ajout de redondance dans les caméras, afin de s'assurer que la main est toujours orientée correctement. Une autre solution, plus simple à mettre en œuvre car logicielle, consisterait en la reconnaissance de deux points au lieu d'un au niveau de la main (par exemple le poignet et un point terminal sur la main). Pour le moment, seul le poignet est détecté par l'algorithme d'OpenNI et ce point bouge selon l'état de la main. Si ce point était détecté de façon robuste, cela permettrait à l'algorithme que nous avons développé d'être plus efficace. Le kit de développement fourni par Microsoft, disponible depuis juillet, permettrait d'obtenir ces informations. Les perspectives d'améliorations sont donc importantes, et permettent d'envisager à court terme la connaissance de la position de chaque doigt de manière précise, rendant possible une interaction plus riche avec des dictionnaires plus évolués qu'à l'heure actuelle.

Les préférences des participants sont majoritairement en faveur de la Cam3D, principalement pour des critères de confort, de liberté de mouvement et de maniabilité. Le sentiment d'immersion est équivalent entre les deux systèmes pour la moitié des participants, mais supérieur avec la Cam3D pour 8

des 10 autres. De même, l'aisance d'utilisation est ressentie de manière supérieure avec la Cam3D par 9 participants, quand seulement 5 évoquent les Gants3D pour ce critère. Ces différents résultats tendent donc à valider les hypothèses H3 et H4.

A l'heure actuelle, avec le système dont nous disposons, nous sommes en mesure de fournir une alternative aux systèmes classiques tels les gants et les capteurs électromagnétiques pour la manipulation d'objets en réalité virtuelle. L'avantage majeur reste bien entendu le coût, largement moins inférieur à un équipement de réalité virtuelle ainsi que les potentialités d'une caméra 3D, qui dépassent largement le cadre des tâches de sélection, manipulation et navigation. Notre système peut ainsi remplacer les équipements classiques tout en apportant une meilleure « utilité-valeur » (Loup-Escande et al., 2011). La prise en compte des améliorations régulières des librairies de détection du « squelette » des participants permettra l'amélioration continue de notre système tout en étendant progressivement ses potentialités, même si aujourd'hui les interactions proposées permettent de couvrir les besoins les plus courants en environnements virtuels.

Ainsi, si l'amélioration des algorithmes de détection des mouvements des mains ainsi que de leur état est dans une certaine mesure tributaire des mises à jour du kit du développement, les modalités d'interaction peuvent d'ores et déjà être affinées pour tenir compte des commentaires des participants, tout en s'inspirant de leurs suggestions. C'est l'objet de l'étude complémentaire présentée dans le chapitre suivant.

## Chapitre 5

# Vers des préconisations pour le choix d'une interaction : une étude complémentaire

---



## 1 Retour sur l'étude 2

### 1.1 Analyse des commentaires et suggestions des participants concernant l'interaction

Suite à la seconde expérimentation présentée dans le chapitre précédent, nous avons analysé les questionnaires afin de déterminer les idées ou suggestions apparaissant comme récurrentes. Certaines remarques étaient communes à plusieurs types de manipulation, comme le manque de précision en fin de positionnement, de redimensionnement ou de rotation.

En ce qui concerne le positionnement, la modalité a été qualifiée de « *très intuitive, simple et efficace* » car elle utilise la « *même gestuelle que dans la réalité (prendre, déplacer, relâcher)* ». Enfin le « *déplacement est très efficace et rapide* ». D'un autre côté, les points négatifs concernent « *la difficulté de positionner un objet à un endroit voulu (un peu instable)* » ou encore la « *sensibilité lors de l'arrêt du déplacement* ». Cela est dû à l'algorithme utilisé par *OpenNI*, qui permet d'appliquer le squelette sur l'utilisateur. Ce squelette, bien qu'assez précis dans l'ensemble, permet de connaître la position des articulations, mais a toutefois l'inconvénient de déplacer légèrement le centre de la main selon que celle-ci est ouverte ou fermée. Puisque les modalités d'interaction développées prennent en compte l'état des mains des participants et qu'il a été décidé que les actions ne peuvent opérer que lorsque les mains de l'utilisateur sont fermées, il est donc normal que la phase de transition entre main ouverte et main fermée produise une légère modification en fin de manipulation. Cela est frustrant pour l'utilisateur qui positionne (ou oriente ou redimensionne) correctement un objet dans l'espace, mais le voit se déplacer légèrement lors de l'ouverture de la main. Nous avons donc décidé de proposer une solution à ce problème en implémentant différemment le « verrouillage » pour le positionnement, le redimensionnement et l'orientation.

Pour ce qui est du redimensionnement, peu de commentaires ont été donnés, et la modalité a été globalement bien perçue et maîtrisée par les participants. Parmi les points positifs, la modalité a été qualifiée de « *simple, intuitive, rapide et précise* » par certains et permet « *un bon contrôle des deux mains* ». En outre, elle « *permet une précision importante et permet de répéter sans limite* ». D'un autre côté, deux personnes ont formulé la même remarque portant sur le fait qu'il est « *difficile de se dire qu'il faut d'abord positionner les deux mains avant.* » En effet pour le redimensionnement, lorsque l'utilisateur ferme ses mains, l'écart entre les deux sert de référence si cet écart augmente, l'objet s'agrandit et si l'écart diminue, l'objet rétrécit. Par conséquent, si un utilisateur ferme ses deux mains trop proches l'une de l'autre, et qu'il souhaite diminuer l'objet, il ne pourra pas avoir une très grande amplitude. Il faut donc, avant de fermer les deux mains, penser à l'action à réaliser (augmenter ou diminuer la taille de l'objet) et prévoir en conséquence l'écart entre les mains. Nous avons donc développé une nouvelle modalité qui prend en compte cette remarque et permette à l'utilisateur de pouvoir débiter le redimensionnement sans connaissance obligatoire préalable de l'action qu'il souhaite exécuter.

Enfin, en ce qui concerne la rotation, les avis ont été plus mitigés. Parmi les points positifs, six personnes trouvent intéressant « *l'accès aux 3 axes sans menu supplémentaire* » (l'utilisateur peut travailler sur les 3 axes en même temps). De plus, l'« *utilisation de la translation plutôt que la rotation de la main [est] moins fatigante* ». Trois personnes ont estimé que la modalité « *permet de gagner en rapidité une fois l'outil pris en main* », enfin elle est considérée comme « *intuitive : on prend l'objet au centre avec la main droite et on le tourne avec la main gauche* ». D'un autre côté, les remarques négatives portaient sur le fait qu'il est « *difficile de décomposer le geste selon les 3 axes* ». Sept personnes ont remarqué une « *difficulté à l'apprentissage initial, car l'utilisation des trois dimensions en vie courante est rare* ». Enfin,

cinq personnes ont trouvé qu'il est « *difficile de savoir quel mouvement permet de tourner autour de quel axe* ». Encore une fois, ces remarques ont été prises en compte pour proposer une autre modalité d'interaction pour la rotation.

## 1.2 Suggestions prises en compte et modifications implémentées

Cette étude complémentaire vise à comparer deux modalités différentes pour une même action, ici portant sur le déplacement, le redimensionnement et la rotation d'objets. Nous avons comparé les modalités utilisées dans l'expérimentation 2 avec une nouvelle implémentation de chacune d'elles, issues de la prise en compte des commentaires et remarques des participants. Dans le cadre de cette étude, nous avons décidé d'utiliser notre système (Cam 3D) car l'étude porte uniquement sur les modalités d'interactions qui ne dépendent donc pas du système utilisé. L'expérimentation se décompose en 3 phases, elles-mêmes subdivisées en deux parties. Chaque phase correspond à une interaction : déplacement, redimensionnement, rotation. Chaque partie correspond à une modalité d'interaction : celle issue de l'expérimentation 2 (modalité 1) et celle nouvellement implémentée (modalité 2). Les tâches sont numérotées et nommées de la façon suivante :

- T1-1 : correspond à la tâche 1, déplacement de 5 cubes selon la modalité 1 ;
- T1-2 : correspond à la tâche 2, déplacement de 5 cubes selon la modalité 2 ;
- T2-1 : correspond à la tâche 3, redimensionnement de 5 cubes selon la modalité 1 ;
- T2-2 : correspond à la tâche 4, redimensionnement de 5 cubes selon la modalité 2 ;
- T3-1 : correspond à la tâche 5, orientation de 5 cubes selon la modalité 1 ;
- T3-2 : correspond à la tâche 6, orientation de 5 cubes selon la modalité 2 ;

### 1.2.1 Modalités proposées pour le déplacement

La première modalité (reprise de l'expérimentation 2) consiste, pour l'utilisateur, à positionner sa main dominante sur l'objet. Une fois l'avatar de la main positionné sur celui-ci, il doit fermer la main (comme s'il souhaitait le saisir). L'objet s'attache alors à l'avatar et il suffit à l'utilisateur de déplacer sa main à l'endroit voulu pour positionner l'objet. Lorsque celui-ci est correctement placé, il peut ouvrir la main, l'objet est relâché.

La deuxième modalité proposée fait appel à la main secondaire de l'utilisateur, qui joue alors le rôle d'« interrupteur ». La main dominante effectue exactement la même chose que dans la modalité 1, cependant le déplacement de l'objet s'effectue uniquement si la main secondaire est fermée (interrupteur en position fermée). Si la main secondaire est ouverte, le déplacement ne se fait pas même si la main dominante se déplace. De cette façon le positionnement devrait être beaucoup plus précis en évitant le léger déplacement final. L'utilisateur peut ainsi positionner l'objet précisément avec sa main dominante et puis ouvrir l'« interrupteur » c'est-à-dire sa main secondaire et enfin sa main dominante. Ainsi, l'objet n'est plus sensible au léger déplacement de la main dominante lors de l'ouverture.

### 1.2.2 Modalités proposées pour le redimensionnement

La première modalité (reprise de l'expérimentation 2) consiste, pour l'utilisateur, à positionner sa main dominante sur l'objet puis la fermer (comme s'il souhaitait le saisir), et enfin fermer sa main secondaire. L'écart entre les deux mains devient la distance de référence. Si l'utilisateur écarte ses mains, l'objet s'agrandit, si l'écart diminue, l'objet rétrécit. Lorsque l'objet est correctement redimensionné, l'utilisateur ouvre ses deux mains.

La deuxième modalité proposée consiste, pour l'utilisateur, à utiliser sa main secondaire en guise d'interrupteur et à effectuer le redimensionnement uniquement avec la main dominante. Si la main secondaire est fermée, le redimensionnement peut s'effectuer, autrement non. Si le redimensionnement est autorisé, l'utilisateur doit alors positionner sa main dominante sur l'objet puis la fermer. La position de la main dominante lors de la fermeture devient le point « neutre ». Si l'utilisateur déplace sa main vers le haut, l'objet s'agrandit, si la main se déplace vers le bas, l'objet rétrécit. Pour expliquer le fonctionnement à l'utilisateur, nous avons utilisé l'analogie à un potentiomètre qu'il déplacerait verticalement pour régler la taille de l'objet. Là encore, les décalages qui pouvaient se produire lors de l'ouverture des mains sont supprimés grâce au rôle d'interrupteur de la main secondaire qui n'a par conséquent pas d'incidence directe sur la taille de l'objet.

### *1.2.3 Modalités proposées pour la rotation*

La première modalité (reprise de l'expérimentation 2) consiste, pour l'utilisateur, à positionner sa main dominante sur l'objet puis la fermer, comme s'il souhaitait le saisir. L'avatar se positionne alors au centre de l'objet et un gizmo (une aide visuelle) apparaît permettant de montrer sur quel axe travaille l'utilisateur lorsqu'il déplace sa main dans une direction particulière. Lorsque l'utilisateur ferme sa main secondaire, la position devient la référence. Un déplacement de la main secondaire vers la gauche va induire une rotation vers la gauche de l'objet, il en est de même lorsque la main se déplace vers la droite. Lorsque la main se dirige vers le haut l'objet tourne dans le sens des aiguilles d'une montre et lorsque la main descend, l'objet tourne dans le sens contraire. Enfin si la main avance (en profondeur) l'objet tourne vers l'avant, si elle va en arrière, l'objet tourne dans l'autre sens. Lorsque l'objet est correctement orienté, l'utilisateur ouvre en premier sa main dominante puis sa main secondaire.

La deuxième modalité proposée consiste, pour l'utilisateur, à positionner et à fermer sa main dominante sur l'objet. Puis il doit fermer sa main secondaire pour autoriser les rotations. La main dominante gère uniquement les axes de rotations X et Z, c'est-à-dire que la main se déplace dans le même plan que l'écran et la main secondaire gère l'axe de rotation Y et se déplace, quant à elle, d'avant en arrière. La main dominante ne gère donc plus à elle seule la rotation selon les 3 axes à la fois, mais l'action est partagée entre les deux mains. Lorsque l'objet est correctement orienté, l'utilisateur ouvre les mains. Bien que cognitivement, la modalité soit certainement plus complexe, la répartition des axes sur les deux mains, doit permettre de mieux décomposer les mouvements. Cela demande tout de même une certaine coordination des mains, contrairement à la modalité 1.

La Figure 54 illustre le fonctionnement de la rotation avec la modalité 1 : les boules rouges représentent la main de l'utilisateur. Ainsi si la main secondaire se déplace sur l'axe X, c'est comme si l'utilisateur enroulait ou déroulait un fil autour de l'axe Y, provoquant ainsi la rotation de l'objet selon cet axe. Pour un déplacement de la main selon l'axe Y, la rotation de l'objet se fera selon l'axe Z et un déplacement de la main selon l'axe Z entraînera une rotation de l'objet selon l'axe X.

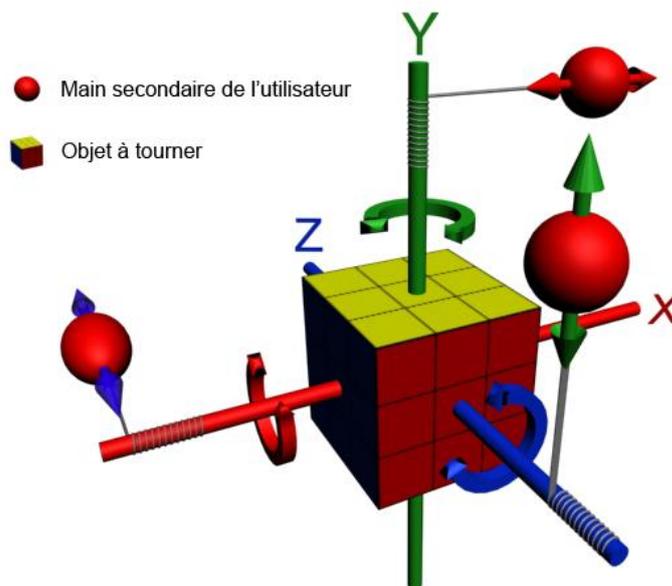


Figure 54 : fonctionnement de la rotation avec la modalité 1

## 2 Méthodologie

### 2.1 Présentation générale

L'objectif de cette étude complémentaire est d'évaluer les modalités d'interaction (déplacement, redimensionnement et rotation) de la seconde étude comparativement à une version améliorée de celles-ci, implémentées d'après les commentaires et suggestions d'amélioration des participants. Ces nouvelles modalités d'interaction sont adaptées à la manipulation d'objet en 3D et visent à être intuitives, efficaces et rapides pour l'utilisateur, tout minimisant la fatigue et augmentant l'intuitivité et la précision. Ce dernier critère est pour nous primordial au regard des commentaires formulés par les participants de l'expérimentation 2 sur le manque général de précision pour les 3 modalités étudiées. Nous avons fait le choix de ne pas étudier la navigation en environnement virtuel, étant donné que cette modalité d'interaction n'a été que peu critiquée et les améliorations que nous aurions pu apporter n'auraient pas nécessité une nouvelle évaluation. Les interactions étudiées ici sont de nature à rendre l'utilisateur acteur dans l'environnement virtuel, et non uniquement spectateur. Par conséquent, si nous réussissons à proposer des modalités d'interaction efficaces, intuitives et rapides, tout en minimisant les efforts de l'utilisateur, cela permettra certainement d'améliorer indirectement le système que nous avons créé en particulier du point de vue de l'acceptation, mais peut être également au niveau du ressenti de la précision permise par les modalités développées.

### 2.2 Terminologie

Voici les différents acronymes et termes que nous allons utiliser dans la suite de ce chapitre :

- Environnement virtuel (EV) : environnement 3D dans lequel les actions sont effectuées ;
- Système Cam3D : système utilisant une caméra 3D Kinect comme interface transparente pour récupérer les actions et les mouvements effectués par l'utilisateur ;
- Main dominante : main de prédilection de l'utilisateur. La main droite sera la main dominante pour une personne droitrière ;
- Main secondaire : main qui n'est pas la main dominante de l'utilisateur. La main gauche sera la main secondaire pour une personne droitrière ;

- Modalité 1 : modalité d'interaction reprise de la seconde étude ;
- Modalité 1 : modalité d'interaction implémentée dans le cadre de cette étude ;
- Tâche X-Y :
  - X correspond à l'interaction (1 = déplacement ; 2 = redimensionnement ; 3 = rotation) ;
  - Y correspond à la modalité (1 = modalité 1 ; 2 = modalité 2) ;
- G1 : groupe des utilisateurs ayant réalisé l'expérimentation 2 ;
- G2 : groupe des utilisateurs n'ayant jamais utilisé notre système de capture de mouvements des mains ni les modalités précédemment développées.

## 2.3 Hypothèses

Etant donné le caractère complémentaire de notre étude, nos hypothèses correspondent à notre objectif principal. En effet, étant donné que nous proposons des améliorations aux modalités d'interaction précédemment implémentées, nous attendons d'obtenir de meilleurs résultats, autant en termes de performances qu'au niveau de l'acceptabilité et des préférences subjectives. Comme nous l'avons mentionné précédemment, nous attendons plus particulièrement une amélioration de précision sur les 3 modalités nouvellement implémentées. Cela nous conduit à formuler les deux hypothèses suivantes :

- Hypothèse A : les 3 modalités d'interaction proposées permettront d'accomplir les tâches de l'expérimentation avec de meilleures performances, en termes de précision et de temps d'exécution.
- Hypothèse B : l'acceptabilité sera meilleure avec les nouvelles modalités que nous proposons. En particulier, le ressenti en termes de précision sera nettement meilleur.

Ces hypothèses, si elles sont confirmées, permettront de conclure sur la supériorité des nouvelles modalités d'interaction que nous proposons en réponse aux commentaires et suggestions des participants à la seconde expérimentation. Nous pourrions ainsi les inclure dans notre système, en remplacement des précédentes.

## 2.4 Tâches de manipulation

Les tâches de manipulation correspondent à celles définies dans la partie 1.4 du chapitre précédent. Il s'agit des actions de :

- Déplacement, c'est-à-dire plus précisément le positionnement d'un objet au sein d'un environnement virtuel en 3 dimensions.
- Redimensionnement, c'est-à-dire la modification du facteur d'échelle d'un objet au sein d'un E.V. Ce redimensionnement est uniforme selon les 3 axes ;
- Rotation, c'est-à-dire l'orientation d'un objet dans un E.V. selon les 3 dimensions.

## 2.5 Environnement virtuel : conception

Nous avons conçu un environnement virtuel qui est le même pour l'ensemble des tâches à réaliser. Il reprend celui utilisé pour l'expérimentation 1, à savoir un cube évidé et quadrillé avec une face en moins, la caméra virtuelle étant positionnée face à celle-ci. La durée d'expérience étant plus courte que celle de l'expérimentation 2, et les tâches à réaliser étant plus simples, nous n'avons pas jugé utile de concevoir des scènes « originales ».

La scène comprend donc un objet qu'il faut faire correspondre au mieux à un modèle, comme pour l'expérience précédente. Pour aider l'utilisateur, une représentation fantôme (en transparence) est affichée : celle-ci correspond au positionnement (ou orientation ou taille) idéal. L'utilisateur doit faire correspondre au mieux les objets à manipuler avec leurs représentations sous forme de « ghost ». L'objet à manipuler est le même pour toutes les interactions proposées et pour chaque modalité. Il s'agit d'un cube dont chaque face a une couleur différente, à la manière d'un *Rubik's Cube*. Dans le cas des tâches de rotation, l'utilisateur peut donc prendre comme point de repère les couleurs de chaque face pour se repérer plus facilement. La Figure 55 représente la scène et l'objet utilisé pour chaque tâche, ici dans le cas d'une tâche de déplacement, de redimensionnement et de rotation.

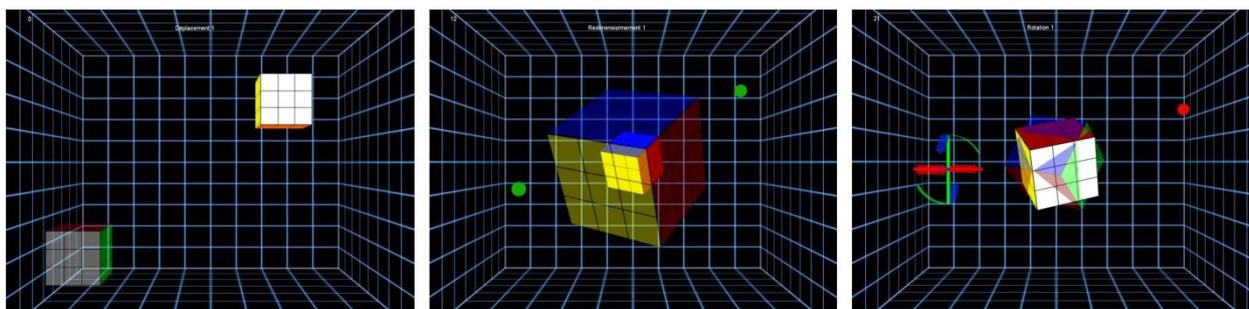


Figure 55 : Scène et actions de déplacement, redimensionnement et rotation

Une tâche consiste à exécuter 5 fois de suite une même action. Par exemple la tâche T1-1 consiste à déplacer 5 cubes selon la première modalité. Pour cela, un cube positionné initialement en bas de l'écran et centré, doit être déplacé à un autre endroit de l'espace de jeu. Chaque participant doit ainsi effectuer 30 sous-tâches (5 par type d'interaction et par modalité).

Le jeu est affiché en perspective, utilisant le principe de la stéréoscopie active, afin d'aider le participant à mieux évaluer les distances et les orientations des objets.

Pour chaque interaction, les participants doivent effectuer les 5 mêmes actions entre les 2 modalités. Nous avons décidé de garder les mêmes conditions (position ou orientation ou dimension) afin d'obtenir des données comparables entre les deux modalités proposées. Comme dans le cas des tâches simples de l'expérience précédente, nous estimons que cela ne biaise pas les résultats dans le sens où les participants n'élaborent pas de stratégie pour atteindre le but demandé. Nous avons tout de même appliqué un contre balancement de l'ordre de présentation des modalités pour assurer la validité de nos résultats.

## 2.6 Implémentation et développements logiciels

L'ordinateur utilisé est le même que précédemment à savoir un appareil de marque *Alienware*, équipé d'un processeur *Intel® Core™ i7 CPU930* cadencé à 2.80GHz avec 6 Giga-octets de RAM et une carte graphique de type *NVidia GeForce GTX480* ainsi qu'un écran *Samsung 120Hz (3D Vision Ready)*. Notre système se base toujours sur la caméra 3D *Kinect* de *Microsoft*.

Nous n'avons pas modifié l'algorithme permettant de détecter la position des mains dans l'espace ainsi que leur état, ouvertes ou fermées. La partie « client » de notre application n'a ainsi pas été modifiée. Nous avons travaillé selon deux axes :

- Modélisation de l'objet à manipuler, insertion dans l'environnement et modélisation de l'aide visuelle pour la rotation avec la modalité 1 : utilisation du logiciel *Autodesk 3DSMax* ;

- Implémentation des nouvelles modalités d'interaction et conception de l'expérimentation à réaliser : utilisation du logiciel *Virtools* de *Dassault Systèmes*.

Enfin, le programme d'extraction des données pertinentes des traces d'interaction a été réalisé en C++. Les questionnaires d'identification et post-passation ont quant à eux été conçus avec *Adobe LiveCycle Form*.

### 3 Protocole expérimental

#### 3.1 Généralités

Pour cette expérimentation, nous avons repris de nombreux éléments du protocole expérimental de la deuxième étude, en particulier concernant les procédures de recueil ainsi que les sources de données des préférences subjectives et commentaires des participants. Nous conservons également le principe d'une évaluation par les participants des deux modalités proposées pour les différentes actions. Enfin, dans cette étude, nous avons décidé de réaliser 2 groupes différents : le premier est composé de personnes ayant réalisé l'expérimentation 2 auparavant et le groupe 2 est composé de personnes n'ayant jamais utilisé ni notre système de capture de mouvements des mains ni les modalités associées. Certains mécanismes seront mis en place pour assurer de ce fait la validité interne de nos résultats.

#### 3.2 Choix des participants

L'expérimentation a impliqué 30 participants volontaires, 7 femmes et 23 hommes, âgés entre 14 et 43 ans. Nous avons initialement 31 participants (8 femmes et 23 hommes) mais nous en avons exclu un de notre expérimentation en raison de problèmes techniques rencontrés, qui auraient pu apporter un biais dans nos résultats. Tous les utilisateurs disposaient de connaissances variées dans le domaine de la réalité virtuelle et possédaient un niveau d'étude BAC+3 minimum, mis à part un participant. La plupart des participants étaient des étudiants en cours de préparation du Master « Ingénierie du Virtuelle et Innovation » ou des membres de l'équipe « Presence & innovation » du « Laboratoire des Arts et Métiers ParisTech d'Angers » (Doctorants, Ingénieurs de Recherche). L'âge moyen des participants était de 25,3 ans (écart-type : 5,5). Les effectifs relatifs des femmes (moyenne d'âge = 23,3 ; E.T. = 2,7) et des hommes (moyenne = 25,9 ; E.T. = 6,0), même s'ils n'étaient pas équivalents (7 contre 23), nous ont tout de même permis de contrebalancer l'ordre de passation suivant le sexe. Etant donné que nous avons souhaité prendre en compte la participation ou non à la seconde expérimentation et que nous avons contrebalancé l'ordre de présentation des modalités, le plan expérimental est quelque peu complexe. Il est représenté sur la Figure 56.

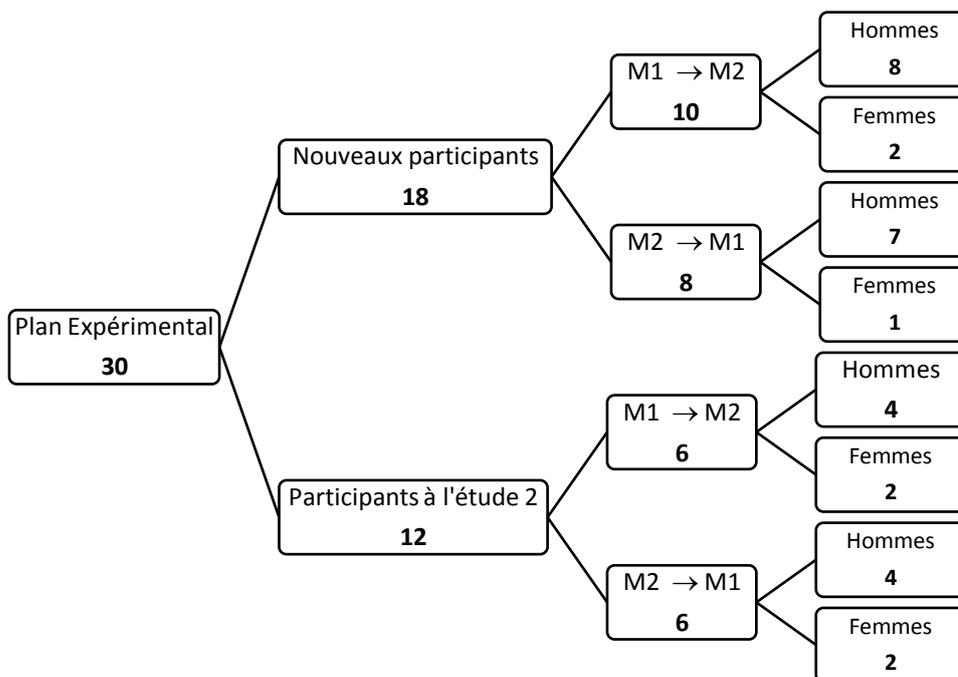


Figure 56 : plan expérimental

Le sexe, comme l'expertise en réalité virtuelle, seront des variables dont l'impact sera évalué dans l'étude statistique conduite dans ce chapitre.

### 3.3 Conditions initiales et consignes

Chaque participant commence la phase d'expérimentation par la réalisation des 10 sous-tâches de déplacement, puis les 10 sous-tâches de redimensionnement et enfin les 10 sous-tâches de rotation. Seul l'ordre de présentation des modalités diffère d'un participant sur l'autre. Les conditions initiales sont identiques pour tous les participants. Contrairement aux deux expérimentations précédentes, il n'y a pas de phase d'apprentissage différenciée : le participant commence l'expérimentation directement. Par contre, sur les 5 sous-tâches par interaction et par modalité, la première est considérée comme une tâche d'entraînement : elle permet au participant de se familiariser avec la modalité d'interaction et n'est de ce fait pas prise en compte dans l'étude des performances des participants.

Chaque participant se trouve dans les mêmes conditions que les deux expérimentations précédentes : il est assis face à l'écran, la *Kinect* étant positionnée au dessus de celui-ci.

La consigne donnée était de faire correspondre l'objet et le modèle, pour chacune des 30 sous-tâches, en essayant d'obtenir une précision maximale. Pour éviter au maximum toute variabilité dans les indications données aux sujets, une fiche descriptive a été mise en place pour expliquer en quoi consiste l'expérimentation d'une part et d'autre part pour expliquer comment fonctionnent les modalités d'interaction afin de réaliser les différentes tâches demandées. Cette fiche est expliquée oralement à chacun des participants et leur est ensuite laissée pour qu'ils puissent s'y référer à tout moment pendant la manipulation, si nécessaire (cf. Annexe C). L'expérimentateur expliquait la phase de calibration, ajustait les paramètres de détection, puis, au début de chaque tâche, rappelait la modalité à utiliser. Il restait à côté du participant, en retrait, et pouvait également être consulté le cas échéant. L'expérimentateur était également présent en cas de problème technique (e.g., problème de reconnaissance de la *Kinect*).

### 3.4 Déroulement de l'expérimentation

Le participant commence par remplir le questionnaire d'identification (format *pdf* réalisé avec *Adobe LiveCycle Form*), nous permettant de recueillir les informations générales le concernant (âge, sexe, latéralisation, niveau d'étude) et les informations spécifiques nous permettant d'établir son niveau d'expertise en réalité virtuelle.

Les participants effectuent alors les 30 sous-tâches, d'abord pour le déplacement, 5 avec une modalité et 5 avec l'autre (selon l'ordre de présentation) et de la même façon pour le redimensionnement et la rotation. L'ordre de présentation des interactions (déplacement, redimensionnement et rotation) est le même pour l'ensemble des participants. Pour chacune des tâches, le participant doit indiquer à l'expérimentateur lorsqu'il estime avoir atteint un niveau de précision suffisant et souhaite passer à la tâche suivante. C'est alors l'expérimentateur qui lance la tâche suivante.

A l'issue de la réalisation des 30 sous-tâches, l'expérimentation prend fin et la collecte des informations s'arrête. Il n'est pas possible de revenir en arrière. Le participant doit remplir un questionnaire final (également réalisé sous la forme d'un formulaire *pdf*) permettant de recueillir ses impressions, commentaires et jugements subjectifs. La Figure 57 présente un participant passant cette expérimentation.

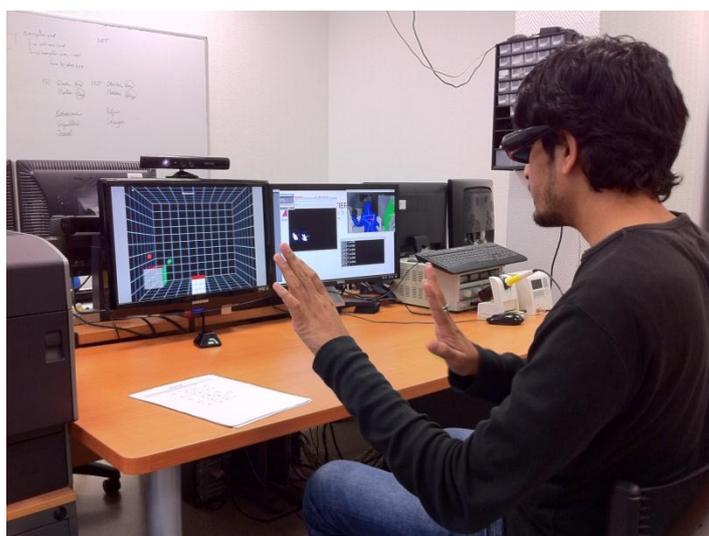


Figure 57 : Participant passant l'expérimentation 3 - tâche de déplacement

### 3.5 Données recueillies, variables, mesures et méthodes d'analyse

La comparaison des deux modalités pour chaque interaction se fait aussi bien de façon quantitative que qualitative. Nous évaluons d'une part les performances des participants pour chacune des modalités proposées et, d'autre part, les commentaires et les évaluations des participants vis-à-vis de ces modalités.

Comme pour la seconde expérimentation, l'ensemble des interactions des participants réalisées au travers des différentes modalités ont été enregistrées : les actions des participants, l'état (ouvert, fermé) des mains des participants, ainsi que la position, l'orientation et la taille des objets. Toutes ces données ont été automatiquement marquées, datées et enregistrées. Ces données nous permettront d'extraire les variables qui rendront possible l'analyse des performances.

Les données permettant l'identification des participants et l'évaluation de l'expertise en réalité virtuelle sont obtenues à partir du questionnaire initial, qui contient des questions ouvertes ainsi que des questions fermées et des échelles de Likert. Le questionnaire utilisé est le même que celui utilisé pour l'expérimentation 2.

Les commentaires et jugement subjectifs sont recueillis dans le questionnaire post-expérimentation. Celui-ci contient des questions ouvertes, des questions fermées ainsi que des questions sous la forme d'échelle de Likert. Les données recueillies concernent l'évaluation comparative de deux modalités pour chacune des actions demandées ainsi que les remarques générales quant à l'expérimentation et aux suggestions d'améliorations éventuelles.

Les mesures suivantes ont été prises en compte et ont servi de base à l'étude des performances des participants :

- Le temps d'exécution moyen pour chacune des 6 tâches à réaliser (en secondes). Pour chaque tâche, ce temps correspond à la moyenne du temps mis pour exécuter chacune des 4 dernières sous-tâches (la première étant considérée comme un entraînement).
- L'erreur moyenne de précision par tâche. Elle est calculée comme un pourcentage d'écart par rapport à une référence selon chaque axe, puis moyenné sur les 3 axes. Enfin, les valeurs d'erreur de précision sont moyennées pour les 4 dernières sous-tâches de chaque tâche.
- Le ratio précision / temps d'action pour chaque tâche. Ce ratio met en relation la précision atteinte avec le temps de manipulation du participant et illustre le compromis à atteindre entre précision et temps d'exécution. Là encore, seules les 4 dernières sous-tâches de chaque tâche sont considérées.

L'expertise en Réalité Virtuelle a été calculée d'après les données du questionnaire initial, sur la même base que pour l'expérimentation 2. Sur les 30 participants, 17 sont qualifiés d'expert en réalité virtuelle et 13 ont un niveau qualifié de moyen. Aucun participant n'est considéré comme débutant.

L'étude des préférences subjectives et commentaires des participants est basée sur l'analyse comparative des 2 modalités proposées pour chaque interaction (déplacement, redimensionnement, rotation). Pour cela, nous avons demandé aux participants d'évaluer chaque modalité d'interaction selon des échelles de Likert à 5 niveaux pour les critères suivants :

- la facilité d'assimilation ;
- la simplicité ;
- l'intuitivité ;
- la précision ;
- la fatigue engendrée.

Les réponses aux échelles de Likert sont traitées comme des variables numériques, chaque modalité étant codée par un nombre (1 = plus mauvais ; 5 = meilleur). D'autre part, les participants ont pu également donner leurs impressions grâce à des questions ouvertes : points positifs / négatifs et suggestions éventuelles d'amélioration pour chaque modalité. Ces commentaires nous serviront à illustrer les résultats présentés dans les parties suivantes.

### 3.6 Validité interne et externe

Pour assurer la validité interne et externe, nous avons eu recours aux mêmes techniques que dans l'expérimentation précédente. Les conditions d'expérimentation étaient identiques. Nous avons eu recours à l'appariement compte tenu de l'effectif des participants, relativement restreint. Cela permet de contrôler les facteurs parasites. Ce choix impose de fait le recours au contre-balancement, pour éviter l'éventuel effet d'apprentissage. Ainsi, l'ordre de présentation des modalités change un individu sur deux. La moitié des participants du groupe G1 et G2 commencent l'expérimentation par les modalités 1 (dans l'ordre suivant : T1-1, T1-2, T2-1, T2-2, T3-1, T3-2) tandis que la deuxième moitié des participants des groupes commencent par les modalités 2 (dans l'ordre T1-2, T1-1, T2-2, T2-1, T3-2, T3-1). Les deux groupes G1 et G2 nous permettront de vérifier s'il existe un effet d'apprentissage et si celui-ci a une incidence sur les résultats (le groupe G1 étant composé de personnes ayant déjà utilisé une modalité de déplacement, de redimensionnement et de rotation sur les deux proposées dans cette expérimentation). Nous pourrions aussi étudier si le fait d'avoir déjà utilisé une partie des modalités auparavant est de nature à influencer les performances et les préférences.

### 3.7 Tests préliminaires et tests statistiques

L'étude statistique repose sur la même démarche que les études précédentes.

Le test de Kolmogorov-Smirnov à un échantillon nous permet de vérifier la normalité des distributions des valeurs prises par chaque variable et de ce fait indique la nature des tests statistiques à réaliser. Concernant les mesures de performances, l'ensemble des variables étudiées suit une loi normale, nous utiliserons donc le test T de Student pour échantillons appariés afin de comparer les 2 modalités proposées, pour chaque interaction. Concernant les préférences et les réponses aux échelles de Likert (codées de « 1 » à « 5 »), en raison de la non-normalité d'une majorité de variables, nous aurons recours à un test de Wilcoxon pour échantillons indépendants.

Afin d'effectuer une inférence sur la modalité préférée pour chaque interaction, nous avons calculé des intervalles de confiance selon la méthode de Wilson (cf. chapitre 3 – partie 4.1.1).

L'étude de l'influence du sexe, du niveau d'expertise, de l'ordre de passation et de la participation ou non à l'expérimentation sur les performances s'appuie sur le test T de Student pour échantillons indépendants. Concernant les préférences et les réponses aux échelles de Likert, nous avons utilisé le test non-paramétrique équivalent, c'est-à-dire le test de Mann-Whitney. Par contre, nous ne pourrions pas utiliser d'intervalles de confiance dans le cas de l'influence éventuelle de ces facteurs sur le choix de la modalité préférée pour chaque interaction, les effectifs étant trop faibles ( $n < 30$ ).

Nous avons adopté les règles suivantes concernant la significativité des tests statistiques :

- $p \leq 0,05$  : test significatif
- $0,05 < p \leq 0,1$  : tendance
- $p > 0,1$  : test non significatif

#### 4 Influence de la participation à la seconde expérimentation

Dans un premier temps, nous allons évaluer si le fait d'avoir participé à la seconde expérimentation est de nature à influencer les commentaires, préférences et performances des participants.

Des différences significatives ou des tendances sont observées sur seulement 2 variables, en faveur des participants ayant effectué l'expérimentation 2 : le temps de déplacement avec l'ancienne modalité ( $T = 1,728$  ;  $p = 0,095$ ) et l'erreur de redimensionnement avec l'ancienne modalité ( $T = 2,116$  ;  $p = 0,043$ ).

Tableau 32 : Différences significatives de performances pour chaque modalité selon la participation à l'expérimentation 2

	Temps Déplacement Modalité 1		Erreur Echelle Modalité 1	
	Etude 2	Nouveau	Etude 2	Nouveau
<b>Moyenne</b>	<b>25,1</b>	36,4	<b>1,9</b>	2,7
<b>E.-T.</b>	<b>8,7</b>	21,4	<b>0,96</b>	1,12
<b>T-Test</b>	1,728		2,116	
<b>p</b>	0,095		<b>0,043</b>	

Ces résultats sont trop peu nombreux au regard de l'ensemble des variables testées pour conclure à une quelconque influence de la participation à la seconde expérimentation sur les performances des participants.

Concernant les critères de difficulté d'assimilation, de simplicité, d'intuitivité, de précision et de fatigue, seule une variable est significative et on observe deux tendances. D'une part, les participants ayant participé à l'étude 2 jugent moins difficile l'assimilation du déplacement avec la nouvelle modalité ( $Z = 2,096$  ;  $p = 0,036$ ) et jugent moins importante la fatigue de déplacement avec la nouvelle modalité ( $Z = 1,731$  ;  $p = 0,083$ ). D'autre part, les participants n'ayant pas participé à la seconde expérimentation jugent meilleure la simplicité de la « nouvelle » rotation.

Tableau 33 : Différences significatives de préférences pour chaque modalité selon la participation à l'expérimentation 2

	Difficulté assim. Déplacement M2		Fatigue Déplacement M1		Simplicité Rotation M2	
	Etude 2	Nouveau	Etude 2	Nouveau	Etude 2	Nouveau
<b>Moyenne</b>	<b>4,25</b>	3,59	<b>4,17</b>	3,72	2,67	<b>3,22</b>
<b>E.-T.</b>	<b>0,754</b>	0,795	<b>0,577</b>	0,752	0,888	<b>0,732</b>
<b>Mann-Whitney</b>	2,096		1,731		1,715	
<b>p</b>	<b>0,036</b>		0,083		0,086	

Les participants n'ayant pas participé à l'expérimentation 2 préfèrent majoritairement la nouvelle rotation (13 sur 18).

Les participants ayant participé à l'expérimentation 2 préfèrent majoritairement la nouvelle rotation (10 sur 12), le nouveau déplacement (8 sur 12) mais l'ancien redimensionnement (9 sur 12).

Comme pour les performances, le caractère partiel des résultats significatifs concernant les préférences ne nous permet pas de conclure à une influence de la participation à la seconde expérimentation sur celles-ci. Nous étudierons donc les performances et des préférences des participants en les considérant comme une population unique.

## 5 Étude des performances et préférences subjectives des participants concernant les interactions

Dans cette partie, nous allons directement confronter les performances et les préférences et commentaires des participants, pour chaque interaction. Après la présentation des tests statistiques nous présenterons, pour chaque interaction, les mesures de performances significatives et l'évaluation comparée selon les modalités des réponses aux questions concernant la facilité d'assimilation, la simplicité, l'intuitivité, la précision et la fatigue engendrée. Cette évaluation sera complétée des préférences globales pour chaque modalité et sera illustrée de commentaires des participants. L'étude des performances et des préférences selon les facteurs de sexe et d'expertise en réalité virtuelle sera détaillée dans la partie suivante.

### 5.1 Le déplacement

Concernant les performances, c'est-à-dire les mesures du temps moyen d'exécution, des erreurs moyennes de précision et du ratio « temps / précision », nous n'observons pas de différences significatives ni même de tendance en faveur de l'une au l'autre modalité.

En termes de préférences, la nouvelle modalité est significativement jugée plus précise ( $Z = 8,865$  ;  $p < 0,001$ ). La modalité de l'expérimentation 2 (modalité 1) est par contre jugée plus intuitive ( $Z = 2,805$  ;  $p = 0,005$ ) et moins fatigante ( $Z = 2,668$  ;  $p = 0,008$ ). Il n'y a pas de différences concernant la facilité d'assimilation ou la simplicité. Le Tableau 34 présente les différences significatives.

Tableau 34 : Différences significatives entre modalité 1 et modalité 2 pour le déplacement

	Intuitivité		Précision		Fatigue	
	Modalité 1	Modalité 2	Modalité 1	Modalité 2	Modalité 1	Modalité 2
<b>Moyenne</b>	<b>4,33</b>	3,77	3,23	<b>4,43</b>	<b>3,90</b>	3,53
<b>Ecart-type</b>	<b>0,711</b>	0,679	0,898	<b>0,626</b>	<b>0,712</b>	0,937
<b>Wilcoxon</b>	2,805		4,048		2,668	
<b>p</b>	<b>0,005</b>		<b>&lt; 0,001</b>		<b>0,008</b>	

La Figure 58 représente les différences entre les 2 modalités selon les 5 critères étudiés. Les critères pour lesquels les différences sont significatives sont encadrés en noir.

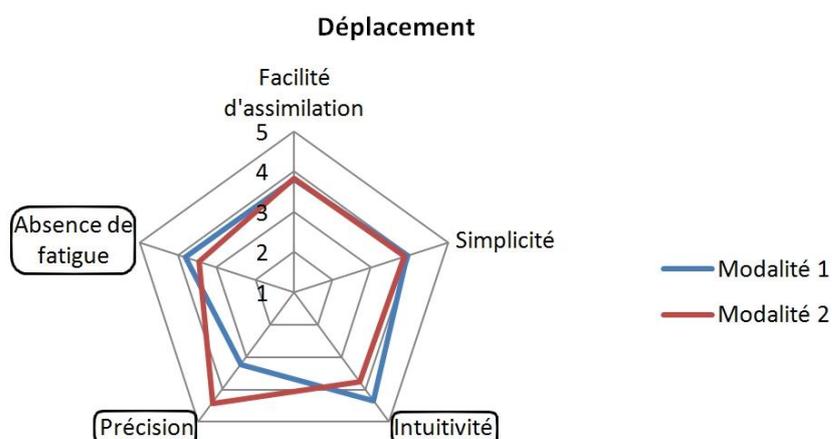


Figure 58 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées

Enfin, de manière globale, Le « nouveau » déplacement a été préféré par 19 des 30 participants, soit 63,3% (intervalle de confiance : 45,5% - 78,1%).

Les participants qui ont donné leur préférence à la modalité 2 expliquent leur choix par « *la meilleure précision* » (sujet 3) du fait de « *l'absence de déplacement lors de l'ouverture de la main* » (sujet 19). Les participants préférant la modalité issue de l'expérimentation 2 motivent leur choix par le caractère « *plus intuitif* » (sujet 6) de ce déplacement et parce qu'il « *paraît plus naturel de manipuler les objets avec une seule main, ce qui se rapproche plus d'une manipulation réelle* » (sujet 8). Ces justifications se retrouvent à l'identique dans les points négatifs et positifs de chaque modalité proposée, indépendamment de la préférence avancée. Ces résultats confirment les différences significatives entre les modalités, pour l'intuitivité, en faveur de la modalité 1, et la précision, en faveur de la modalité 2. La fatigue, plus importante, est une conséquence de « *l'obligation de coordonner les deux mains* » (sujet 15).

Peu de participants ont proposé des améliorations, celles-ci étant avant tout prospectives et moins opérationnelles, comme par exemple le « *verrouillage par la parole* » (sujet 28) ou « *combiner la modalité 2 sur une seule main avec la reconnaissance des doigts* » (sujet 11).

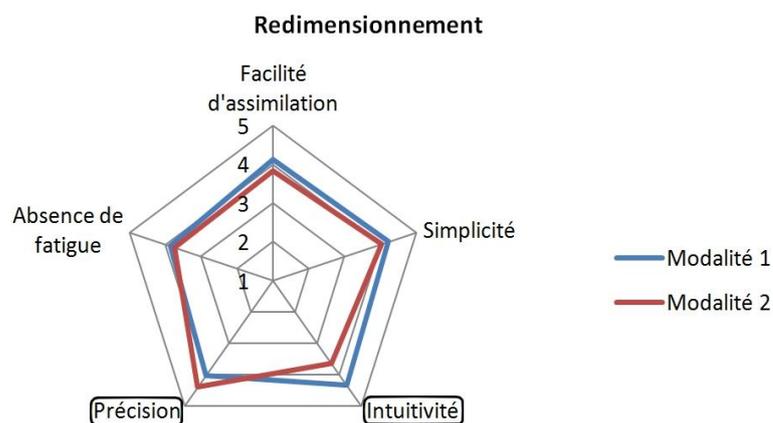
## 5.2 Le redimensionnement

Concernant les performances, c'est-à-dire les mesures du temps moyen d'exécution, des erreurs moyennes de précision et du ratio « temps / précision », nous n'observons pas de différences significatives ni même de tendance en faveur de l'une au l'autre modalité pour le redimensionnement.

La nouvelle modalité est significativement jugée plus précise par les participants ( $Z = 2,592$  ;  $p = 0,017$ ). La modalité de l'expérimentation 2 est par contre jugée plus intuitive ( $Z = 3,155$  ;  $p = 0,002$ ). Les valeurs moyennes sont présentées dans le Tableau 35. Il n'y a pas de différences significatives concernant la facilité d'assimilation, la simplicité et l'absence de fatigue. La Figure 59 représente, quant à elle, les différences entre les 2 modalités selon les 5 critères étudiés. Les critères pour lesquels les différences sont significatives sont encadrés en noir.

Tableau 35 : Différences significatives entre modalité 1 et modalité 2 pour le redimensionnement

	Intuitivité		Précision	
	Modalité 1	Modalité 2	Modalité 1	Modalité 2
<b>Moyenne</b>	<b>4,43</b>	3,67	4,07	<b>4,43</b>
<b>Ecart-type</b>	<b>0,711</b>	0,758	0,828	<b>0,568</b>
<b>Wilcoxon</b>	3,155		2,392	
<b>p</b>	<b>0,002</b>		<b>0,017</b>	



**Figure 59 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées**

Le « nouveau » redimensionnement a été préféré par 10 des 30 participants, soit 33,3% (intervalle de confiance : 19,2% - 51,2%).

Comme précédemment pour le déplacement, les préférences des participants pour la première modalité s'expliquent par son côté « *plus intuitif* » (sujet 12) et « *plus simple d'utilisation* » (sujet 13). Certains participants ont fait l'analogie avec ce qui se fait pour les interfaces tactiles, parlant d'une « *métaphore proche du multitouch* » (sujet 13). Seuls 5 participants ont évoqué un problème de précision, l'objet « *bougeant quand on ouvre la main* » (sujet 17), ce faible nombre semble surprenant étant donné que la précision est jugée significativement moins bonne avec cette modalité. De plus, un seul participant a évoqué le problème de la réduction de taille « *lorsque les mains arrivent en contact* ».

Le redimensionnement proposé par la seconde modalité est appréciée avant tout car elle apporte « *plus de précision pour de petits objets* » (sujet 20) tout en étant jugée « *facile à assimiler et à réaliser* » (sujet 5) par certains participants. Le redimensionnement, que ce soit pour la modalité 1 ou la modalité 2, n'a pas beaucoup suscité de commentaires, ce qui est cohérent avec les résultats de l'étude précédente.

Les améliorations proposées concernent principalement la vitesse de redimensionnement, comme par exemple l'idée de « *dérivée la position du bras en vitesse de redimensionnement* » (sujet 17). Sur les 30 participants, 21 n'ont pas proposé d'améliorations.

### 5.3 La rotation

La nouvelle modalité implémentée pour la rotation permet de minimiser significativement l'erreur d'orientation ( $T = 8,865$  ;  $p < 0,001$ ), le ratio « précision / temps » ( $T = 2,974$  ;  $p = 0,006$ ) et nous observons une forte tendance concernant le temps d'exécution des tâches de rotation ( $T = 1,950$  ;  $p = 0,061$ ) en faveur de la nouvelle modalité de rotation. Les données sont synthétisées dans le Tableau 36.

**Tableau 36 : Différences significatives entre modalité 1 et modalité 2 pour la rotation**

	Temps (s)		Erreur de précision (%)		Ratio temps/précision	
	Modalité 1	Modalité 2	Modalité 1	Modalité 2	Modalité 1	Modalité 2
<b>Moyenne</b>	63,76	<b>56,0</b>	6,47	<b>1,44</b>	1,61	<b>2,06</b>
<b>Ecart-type</b>	20,72	<b>23,27</b>	3,45	<b>0,70</b>	0,52	<b>0,81</b>
<b>T-Test</b>	1,950		<b>8,865</b>		<b>2,974</b>	
<b>p</b>	0,061		<b>&lt; 0,001</b>		<b>0,006</b>	

La nouvelle modalité implémentée pour la rotation est significativement jugée plus précise ( $Z = 3,346$  ;  $p = 0,001$ ). On note également une tendance en faveur de cette modalité pour la fatigue ( $Z = 1,667$  ;  $p = 0,096$ ). Les valeurs moyennes sont présentées dans le Figure 37. Il n’y pas de différences significatives concernant la facilité d’assimilation, la simplicité et l’intuitivité.

Tableau 37 : Différences significatives entre modalité 1 et modalité 2 pour la rotation

	Précision		Fatigue	
	Modalité 1	Modalité 2	Modalité 1	Modalité 2
<b>Moyenne</b>	2,87	<b>3,53</b>	2,97	<b>3,13</b>
<b>Ecart-type</b>	0,900	<b>0,681</b>	0,890	<b>0,819</b>
<b>Wilcoxon</b>	3,346		1,677	
<b>p</b>	<b>0,001</b>		0,096	

La Figure 60 représente quant à elle les différences entre les 2 modalités selon les 5 critères étudiés.

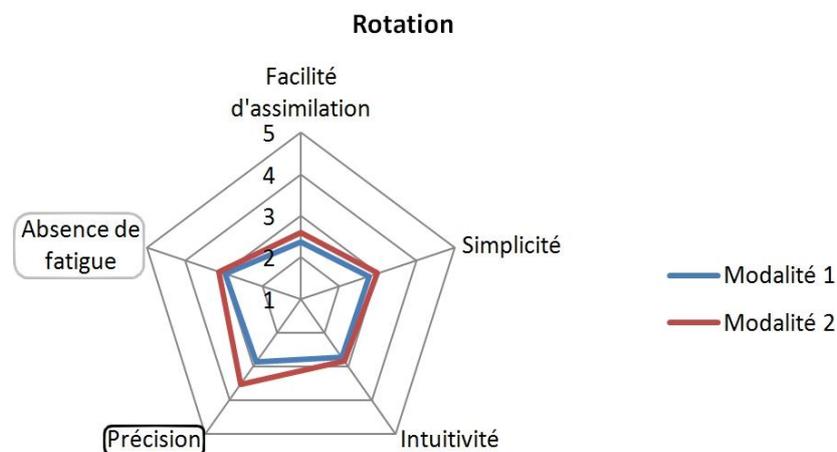


Figure 60 : Représentation des scores moyens de chaque critère pour les 2 modalités étudiées. Les variables pour lesquelles les différences sont significatives sont encadrées en noir(en gris pour les tendances)

La « nouvelle » rotation a été préférée par 23 des 30 participants, soit 76,7% (intervalle de confiance : 59,1% - 88,2%).

Il y a une réelle opposition entre les personnes qui préfèrent la modalité 1 et celles qui préfèrent la modalité 2, puisque c’est principalement le critère de l’intuitivité qui sert de justification. En effet, la modalité 2 est appréciée, car « l’utilisation des deux mains rend la rotation plus simple » (sujet 4) et « qu’au moins un axe peut facilement être bloqué » (sujet 31) ; cette modalité est jugée « plus précise » (sujet 29) et « moins fatigante » (sujet 6). Ces participants jugent trop « difficile de gérer les trois axes en même temps » avec la modalité 1 (sujet 2). De façon similaire, les participants ayant préféré la rotation issue de l’expérimentation 1 le justifient par « un travail cognitif moins important, les rotations s’effectuant d’une seule main » (sujet 7) et par la « difficulté de dissocier les axes de rotation » (sujet 17). Si les préférences sont majoritairement en faveur de la modalité 2, les participants ont tout de même évoqué la nécessité d’une « plus grande coordination des deux mains » (sujet 16). La difficulté d’assimilation reste une problématique commune aux deux modalités : s’il peut être « difficile de trouver les axes de rotations des objets en fonction des mouvements » (sujet 1) avec la modalité 1, il est également « difficile de différencier les axes » (sujet 29) avec la modalité 2.

## 5.4 Synthèse des résultats

La modalité 2, implémentée d'après les commentaires des participants à l'expérimentation 2, permet d'obtenir de meilleures performances pour la rotation, tant au niveau du temps moyen d'exécution, que de la précision et du ratio « précision / temps ». Les performances sont équivalentes entre les deux systèmes pour le déplacement et le redimensionnement.

Une majorité de participants donnent leurs préférences aux nouvelles modalités implémentées pour le déplacement et la rotation mais à la modalité issue de l'expérimentation 2 (modalité 1) pour le redimensionnement.

La précision ressentie se révèle significativement supérieure pour l'ensemble des nouvelles modalités d'interaction, mais cela se fait au détriment de l'intuitivité, jugée meilleure pour le déplacement et le redimensionnement. C'est le fait de devoir coordonner les mouvements des deux mains qui explique la faiblesse de l'intuitivité avec la nouvelle modalité. Ce phénomène était moins prononcé pour la rotation, car la première modalité implémentée avait déjà un certain niveau de complexité. Le critère de la fatigue a ici été jugé plutôt du point de vue cognitif que du point de vue physique, la plupart des commentaires rejoignant ceux formulés à propos de l'intuitivité.

Lors de la seconde expérimentation, c'est avec le déplacement que le manque de précision avait été jugé le plus gênant. Nous retrouvons une cohérence dans les commentaires des participants de cette étude complémentaire, puisque c'est l'amélioration de la précision avec la modalité 2 qui a été majoritairement évoquée pour le déplacement. De même, pour la rotation, on retrouve la difficulté de gérer les 3 axes en même temps et avec une seule main, qui avait été largement évoquée lors de l'étude précédente, et qui justifie en creux la préférence plus marquée pour la rotation nouvellement implémentée. Enfin, le redimensionnement, qui avait fait l'objet de peu de commentaires dans l'étude précédente, n'apporte pas de réelle amélioration selon les participants.

## 6 Influence de critères interindividuels sur les performances et les préférences

Cette partie présente l'étude des performances et des préférences selon les facteurs de sexe, d'expertise en réalité virtuelle.

### 6.1 Le sexe

L'erreur de rotation, pour les modalités 1 et 2, est significativement plus faible avec les hommes (resp.  $T = 2,560$  ;  $p = 0,016$  et  $T = 3,384$  ;  $p = 0,002$ ). Le temps de rotation avec la nouvelle version est significativement inférieur pour les hommes ( $T = 2,448$  ;  $p = 0,021$ ). Nous avons une forte tendance concernant l'erreur de déplacement avec l'ancienne version ( $T = 1,918$  ;  $p = 0,065$ ) en faveur des hommes, ainsi qu'une tendance pour le ratio « temps / précision » pour la rotation avec la nouvelle version ( $T = 1,795$  ;  $p = 0,085$ ).

Pour l'ensemble des variables étudiées, nous ne notons pas d'effet du sexe sur les mesures de difficulté d'assimilation, de simplicité, d'intuitivité, de précision et de fatigue pour les 2 implémentations des modalités d'interaction.

Une majorité d'hommes préfèrent le « nouveau » déplacement (15 sur 23) et la « nouvelle » rotation (18/23) mais le redimensionnement issu de l'expérimentation 2 (16 sur 23). Les femmes semblent avoir une préférence pour la « nouvelle » rotation (5 sur 7).

## 6.2 L'expertise en RV

Le niveau d'expertise a une importance significative avec les modalités d'interaction issues de l'expérimentation 2 (modalité 1), avec de meilleurs résultats pour les participants qualifiés d'experts en réalité virtuelle, pour le temps de rotation ( $T = 2,057$  ;  $p = 0,049$ ), l'erreur de déplacement ( $T = 2,897$  ;  $p = 0,007$ ), l'erreur de redimensionnement ( $T = 2,029$  ;  $p = 0,052$ ) (forte tendance pour la variable en question), l'erreur de rotation ( $T = 2,122$  ;  $p = 0,043$ ), le ratio « temps / précision » pour le déplacement ( $T = 2,088$  ;  $p = 0,046$ ) et le ratio « temps / précision » pour la rotation ( $T = 2,212$  ;  $p = 0,035$ ). Les données sont synthétisées dans le Tableau 38.

Tableau 38 : Différences significatives de performances pour chaque modalité selon l'expertise en RV

	Tps_Rot_M1		Err_Dep_M1		Err_Sca_M1		Err_Rot_M1		Rat_Dep_M1		Rat_Rot_M1	
	Moy.	Exp.										
<b>Moyenne</b>	72,2	57,3	18,0	12,1	2,8	2,0	7,9	5,4	2,7	3,8	1,4	1,8
<b>E.-T.</b>	24,0	15,6	6,3	4,9	1,2	0,9	2,8	3,5	1,3	1,5	0,4	0,6
<b>T-Test</b>	2,057		2,897		2,029		2,122		2,088		2,212	
<b>p</b>	0,049		0,007		0,052		0,043		0,046		0,035	

On observe des différences significatives pour l'expertise pour seulement 2 variables : les participants avec un niveau d'expertise qualifié de moyen donnent une meilleure évaluation de l'intuitivité du « nouveau » déplacement ( $Z = 2,374$  ;  $p = 0,018$ ) et de la simplicité de la « nouvelle » rotation ( $Z = 2,253$  ;  $p = 0,024$ ). On observe enfin une tendance en faveur de l'intuitivité de la « nouvelle » rotation selon les participants dont l'expertise est qualifiée de moyenne ( $Z = 1,936$  ;  $p = 0,053$ ).

Tableau 39 : Différences significatives de préférences pour chaque modalité selon l'expertise en RV

	Intuitivité Déplacement M2		Simplicité Rotation M2		Intuitivité Rotation M2	
	Moy.	Exp.	Moy.	Exp.	Moy.	Exp.
<b>Moyenne</b>	4,08	3,53	3,38	2,71	3,15	2,59
<b>E.-T.</b>	0,49	0,72	0,77	0,77	0,69	0,80
<b>Mann-Whitney</b>	2,374		2,253		1,936	
<b>p</b>	0,018		0,024		0,053	

Les participants avec un niveau moyen en réalité virtuelle préfèrent en majorité la « nouvelle » rotation (10 sur 13) et le « nouveau » déplacement (10 sur 13). Les participants qualifiés d'experts préfèrent en majorité le redimensionnement issu de l'expérimentation 2 (14 sur 17) mais préfèrent la « nouvelle » rotation (13 sur 17).

## 6.3 Synthèse

Le sexe des participants n'a pas d'influence sur les préférences et les évaluations subjectives des participants. Ce constat est moins catégorique pour les performances, et en particulier pour la rotation. En effet, nous observons que les hommes obtiennent de meilleurs résultats avec la nouvelle modalité implémentée pour cette interaction, aussi bien en termes de temps d'exécution que d'erreur de précision ou de ratio « précision / temps ».

Nous constatons que l'expertise en réalité virtuelle est de nature à influencer les performances et les préférences des participants. Pour les premières, nous remarquons que seules, sont impactées, les performances liées aux interactions issues de l'expérimentation 2 : les participants qualifiés d'experts obtiennent de meilleurs résultats que ceux dont le niveau est qualifié de moyen. Cela semble indiquer

que les modalités d'interaction étaient plus compliquées à appréhender, cette différence étant « lissée » avec les nouvelles modalités d'interaction. Seuls les participants avec un niveau d'expertise qualifiés de moyen sont sensibles au caractère plus intuitif du nouveau déplacement ou de la simplicité et de l'intuitivité offertes par la rotation de la modalité 2. La nouvelle rotation fait l'unanimité quel que soit le niveau d'expertise en réalité virtuelle.

## 7 Discussion et préconisations

A la suite de la seconde expérimentation, nous avons analysé les commentaires des participants relatifs aux modalités d'interaction. Cela nous a conduit à implémenter de nouvelles modalités d'interaction pour le déplacement, le redimensionnement et la rotation afin d'être en mesure de proposer, dans notre système, une interaction permettant de réaliser les tâches de manipulation en environnement virtuel de façon performante avec un niveau d'acceptabilité maximal. Nous avons opposé, dans le cadre de cette étude complémentaire, les modalités d'interactions développées dans l'expérimentation 2 aux modalités d'interactions implémentées suite aux commentaires des participants. Nous souhaitons démontrer la supériorité des « nouvelles » modalités d'interaction tant en termes de performances (temps d'exécution, erreur de précision et ratio « précision / temps ») qu'en termes de préférences, aussi bien générales qu'étudiées sous l'angle de la facilité d'assimilation, la simplicité, l'intuitivité, la précision et la fatigue engendrée. L'étude a impliqué 30 participants et a porté sur la réalisation de 30 sous-tâches simples, soit 5 par interaction et par modalité.

Nos résultats permettent de valider les deux hypothèses que nous avons émises, tout du moins en partie concernant l'hypothèse A. En effet, si les performances sont équivalentes entre les deux modalités pour le déplacement et le redimensionnement, elles sont significativement en faveur de la nouvelle modalité pour la rotation, que ce soit en termes de temps d'exécution, d'erreur de précision ou de ratio « précision / temps ». Ce résultat est important car c'est la rotation qui avait cristallisé le plus de critiques lors de l'expérimentation 2. Nous avons ainsi apporté une réponse satisfaisante concernant cette interaction, puisqu'elle recueille également la majeure partie des préférences des participants. Ce qui est également le cas du déplacement. Le redimensionnement, par contre, ne semble pas apporter de différence significative, ce qui peut s'expliquer par le fait qu'elle était déjà peu critiquée lors de la seconde expérimentation.

Le problème de précision dû à l'ouverture de la main et au recalage de la position terminale du squelette sur la main a disparu. La précision ressentie, bien qu'elle ne soit effective que pour la rotation, est significativement meilleure avec la nouvelle modalité, pour toutes les interactions. Ce problème, qui était surtout gênant pour le déplacement, explique la préférence des participants pour la seconde modalité. De la même façon, la possibilité de décomposer la rotation sur les deux mains rend, pour la majorité des participants, la rotation plus efficace.

Par contre, le fait d'utiliser deux mains pour réaliser chaque modalité rend les interactions moins intuitives. La précision a été obtenue au détriment de la simplicité d'utilisation même si l'apprentissage du fonctionnement des interactions reste très rapide. Dans le cas du déplacement et du redimensionnement, le rôle de la main secondaire reste mineur, avec un simple rôle de « verrou », ce qui minimise la complexité de l'interaction. Dans le cas de la rotation, chaque main joue un rôle majeur, ce qui peut induire une certaine complexité d'action chez certaines personnes. Même si celles-ci sont minoritaires dans le cadre de notre étude, il convient de prendre en compte cette population. En particulier, nous avons remarqué que les femmes obtenaient de moins bonnes performances avec la nouvelle modalité. Nous n'avons pas à l'heure actuelle d'explication concrète à formuler concernant ce

résultat. Le fait que seule l'obligation d'utiliser deux mains et donc de coordonner ses mouvements change entre les 2 modalités suggère qu'il faille chercher dans cette direction et dans la littérature traitant des capacités visuo-spatiales et des travaux relatives aux tâches simples de coordination des mouvements en environnements réels.

A l'issue de cette étude, nous formulons 2 préconisations quant au choix des modalités d'interaction à intégrer à notre système :

- En fonction de la précision demandée, privilégier le déplacement et le redimensionnement de l'étude 2 ou ceux implémentés dans le cadre de cette étude. En effet, pour des tâches de positionnement ou de redimensionnement dans lesquelles les mouvements pourraient être assistés (par exemple, magnétisme de la position à atteindre), l'utilisation de la seconde main comme « verrou » n'est pas obligatoire. Par contre, dans le cas d'un déplacement ou d'un redimensionnement libre, ou demandant une grande précision, nous avons pu montrer que la nouvelle modalité implémentée permettait d'apporter une réponse efficace, a minima au niveau du ressenti du participant.
- Concernant la rotation, nous privilégions la nouvelle modalité d'interaction, car elle permet de meilleures performances et recueille également de meilleures préférences de la part des participants. Cette interaction est tout à fait adaptée à l'orientation libre d'un objet au sein d'un environnement virtuel.

Ainsi, si un concepteur d'application intégrant de la manipulation d'objets 3D en environnement virtuel souhaite utiliser notre système mais ne connaît pas *a priori* la nature précise des tâches à effectuer, nous lui conseillons d'implémenter, dans l'état actuel de nos développements, le déplacement proposé dans cette étude, car il est plus « polyvalent ». Nous recommandons également l'implémentation de la rotation basée sur la coordination des deux mains. Le choix du redimensionnement est laissé libre, les deux étant équivalents mais les participants ayant une préférence marquée pour celui issue de l'étude 2. Par contre, si les tâches de manipulation ne demandent pas une précision importante, le déplacement issu de l'étude 2 sera tout à fait adapté.

## Chapitre 6

# Conclusions et perspectives

---



## 1 Conclusion

Dans la vie de tous les jours, nous utilisons nos mains pour réaliser toutes sortes de tâches sans même y prêter attention. Elles couvrent près de 70% des capacités motrices de l'homme et la richesse fonctionnelle et la précision qu'elles offrent en font des outils pour le moment inégalés et très difficiles à reproduire. De nombreuses recherches, en particulier dans le domaine de la robotique, tentent de recréer cette complexité. Cependant, aucun robot n'est capable, à l'heure actuelle, d'égaliser à la fois l'habileté, la précision et la finesse du toucher que permet une main. Des prototypes très évolués s'en approchent et peuvent surpasser la main dans des cas bien précis (force, rapidité, etc.) mais ne peuvent cependant effectuer toutes les actions qu'une main humaine est capable de faire et ce avec la même aisance et la même efficacité (Namiki et al., 2009; Senoo et al., 2009; Shadow\_Robot\_Company, 2003). Pour fonctionner, ces robots tentent de reproduire les mouvements humains le plus fidèlement possible. Pour cela, il est possible d'enregistrer un mouvement ou encore de l'exécuter en temps réel afin que le robot le reproduise. Il est donc primordial, dans un premier temps, de pouvoir capturer les mouvements des mains avec différents capteurs et outils.

Dans le domaine de la réalité virtuelle, le problème se pose également car, pour interagir avec un environnement virtuel, nous utilisons principalement nos mains. Il faut donc capturer les gestes de l'utilisateur pour les retranscrire virtuellement et déclencher certaines actions en fonction d'eux. Les outils standards comme le clavier et la souris fonctionnent parfaitement pour des applications 2D (de type bureautique). De même, certains périphériques un peu plus évolués comme les joysticks ou les souris 3D sont utilisés dans les applications ludiques (e.g. les jeux vidéo) ou dans les applications à portée professionnelle (e.g. conception, modélisation, réalité virtuelle). Cependant, ces outils ne suffisent plus lorsqu'il s'agit de manipuler des objets en 3 dimensions de façon efficace, intuitive et rapide dans un environnement virtuel et sont inutilisables pour des actions à l'échelle 1:1. La démocratisation rapide des outils de réalité virtuelle, et en particulier la diffusion des affichages permettant une 3D stéréoscopique (par exemple, dans les applications à domicile comme les jeux ou les vidéos) rend nécessaire le développement de moyens d'interaction adaptés à la 3D temps réel. En effet, aucun périphérique ne répond actuellement aux critères de fiabilité, de simplicité d'utilisation, de coût réduit et de transparence pour l'utilisateur. Ceci est particulièrement vrai pour les gants de données qui sont les plus utilisés mais qui ne remplissent pas, et loin s'en faut, l'ensemble de ces conditions. La notion d'immersion revêt par ailleurs un intérêt particulier dans le cadre de la réalité virtuelle où l'absence de métaphore intermédiaire pourrait simplifier la conception d'interfaces (Fuchs & Richir, 2006). De plus, il est parfois nécessaire de pouvoir capturer les gestes de l'utilisateur sans lui imposer la présence de capteurs sur les mains. Il faut par conséquent utiliser des systèmes transparents. Si la transparence peut renforcer le sentiment d'immersion, elle peut être une contrainte forte lorsqu'il n'est pas possible d'installer du matériel sur les participants (comme par exemple pour des personnes autistes). C'est pourquoi nous pensons qu'il est aujourd'hui nécessaire de porter nos efforts vers une solution qui serait plus simple à mettre en œuvre, moins coûteuse et plus facilement adoptée par les utilisateurs issus du grand-public.

Puisque les interactions en environnement virtuel s'appuient majoritairement sur la main, et tendent par cela à se rapprocher de la réalité, une solution, a priori évidente, serait de transposer aux interfaces 3D, les mouvements naturels que nous réalisons dans la vie réelle. De cette manière, nous pourrions interagir avec l'environnement virtuel tout en utilisant les interactions manuelles (ou bi-manuelles) du monde réel.

La question que nous nous sommes posée était, par conséquent, de savoir comment nous pouvions interagir en environnement virtuel, avec la même simplicité que nous le faisons dans le réel, sans avoir pour autant recours à des interfaces intrusives ?

L'état de l'art que nous avons réalisé montre le nombre impressionnant de techniques existantes permettant de capturer plus ou moins finement les mouvements des mains. Plusieurs solutions commerciales sont fiables et relativement efficaces mais possèdent, en contrepartie, des inconvénients importants comme le prix, le poids ou encore le port de matériel. Ce dernier aspect est contraignant pour l'utilisateur et rend ces solutions intrusives. Des solutions reposant sur l'utilisation de caméras et d'algorithmes de traitements d'images permettent de palier, en partie, ces problèmes, mais ne permettent pas (pour la plupart d'entre elles) de déterminer de façon précise la position des mains dans l'espace. Une solution qui semble prometteuse est d'utiliser des caméras 3D qui sont capables de réaliser des cartes de distances d'une scène réelle. De cette façon, il est possible de connaître la position des mains dans l'espace tout en gardant les avantages des solutions reposant sur le traitement d'images. Un système combinant ces deux atouts permettrait de capturer les mouvements de l'utilisateur de façon précise sans que celui-ci ne soit gêné, ni dans ses mouvements, ni par le poids du matériel ou par les câbles. De plus, un tel système permettrait de rendre totalement transparent le périphérique pour l'utilisateur, ce qui tendrait à améliorer le sentiment d'immersion de celui-ci dans l'environnement.

Le principal verrou concerne la capture de mouvements sans marqueurs des mains qui demeure aujourd'hui peu performante avec les technologies de caméras standards. De plus, dans de nombreux cas, la zone de travail est très restreinte. Cependant, le traitement d'images par ordinateur, associé à de nouveaux périphériques comme les caméras 3D (e.g. *Microsoft Kinect*) et à une puissance de calcul (GPU et CPU) suffisante pour le traitement du flux de données, est suffisamment mature pour permettre une interaction temps réel offrant les avantages énoncés précédemment. Il reste néanmoins à attester l'intérêt de ce type de technologie pour l'interaction temps réel en environnement virtuel, et à en explorer les potentialités.

Nos travaux reposent sur plusieurs postulats. D'une part, la conception d'une technologie ayant pour objectif de faire « disparaître » l'interface afin de permettre un mode d'interaction « naturel » aux utilisateurs est souvent présentée comme l'aboutissement final des interfaces de manipulation directe. Fuchs et Moreau (Fuchs & Moreau, 2003) parlent à ce propos d'interfaces comportementales, ce qui implique que les recherches doivent se tourner vers la conception de systèmes « visant à exploiter un comportement humain, naturel et sans acquis préalable ». De même, comme le suggèrent Fuchs et Mathieu, dans le volume 2 du *Traité de la Réalité Virtuelle* (Fuchs & Mathieu, 2003), de manière idéale, les réponses motrices (dans le cadre de la capture de mouvements) doivent être transmises sans support matériel entre l'homme et la machine. Enfin, d'après (Winkler et al., 2007), le fait de n'avoir aucun matériel à porter et de rendre le système transparent pour l'utilisateur est de nature à améliorer le sentiment d'immersion et de présence ce qui est primordial dans le domaine de la réalité virtuelle.

Le système que nous avons créé tente d'apporter, en partie, des réponses à ces questions qui sont cruciales dans le domaine de la réalité virtuelle. En lien avec cet objectif, nous avons défini cinq hypothèses générales. Elles affirment, d'une part, que les performances obtenues avec notre système, en termes de temps d'exécution, sont équivalentes à celles obtenues avec une souris (pour une tâche de sélection) – hypothèse A – et équivalentes, en termes de précision et de temps de réalisation (pour des tâches de manipulation et de navigation), à celles obtenues avec des gants de données couplés à des capteurs magnétiques - hypothèse C. D'autre part, nous affirmons que l'acceptabilité de notre système

est meilleure, en termes d'utilisabilité, d'utilité et de sentiment d'immersion qu'une souris (hypothèse B) ou qu'un système à gants de données couplé à des capteurs magnétiques (hypothèse D). Enfin, notre dernière hypothèse concerne plus précisément les modalités d'interaction développées et affirme que ces dernières sont adaptées à notre système et permettent de réaliser de manière simple, confortable efficace et efficiente les tâches de sélection, manipulation et navigation (hypothèse E). Pour vérifier ces hypothèses, nous avons développé deux études principales et une étude complémentaire.

La première version de notre système utilisait une caméra 3D de type *SwissRanger SR4000* et permettait de suivre les mouvements d'une main et de connaître sa position et son orientation (dans une certaine mesure) dans une zone de travail de 30 x 30 x 30cm. Cette caméra nous a permis de concevoir un système transparent, capturant les mouvements de la main en temps réel à un taux de rafraîchissement de 30fps. La première étude est basée sur l'évaluation comparative de notre système et d'une souris pour une tâche simple de sélection 3D en environnement virtuel. Elle s'appuie sur la première version de nos développements et implique 71 participants. Pour cette expérimentation, nous avons développé un environnement virtuel servant de support à une tâche de sélection d'objets (10 cubes à sélectionner par ordre croissant) et nous avons comparé les performances obtenues avec notre système et ceux obtenus avec la souris. D'autre part, nous avons évalué l'acceptabilité de notre système vis-à-vis de cette dernière.

Les résultats montrent, d'une part, que les performances avec la Cam3D ne sont pas sensiblement meilleures que celles obtenues avec la souris. Ce premier résultat est à mettre en perspective avec le temps d'expérimentation très court, qui constitue un désavantage pour la Cam3D, la souris bénéficiant d'un niveau de pratique supérieur, malgré la phase de prise en main au début de l'expérimentation. Ce premier résultat confirme notre hypothèse. Nous avons également montré que le sexe n'influence pas les performances obtenues avec la cam3D, contrairement à l'expertise en réalité virtuelle et à la taille de la main. Nos résultats indiquent que les personnes expertes et initiées obtiennent de meilleurs résultats que les personnes novices. De la même manière, les personnes avec des mains qualifiées de grande taille ou de taille moyenne achèvent plus rapidement la tâche que les personnes avec de petites mains. D'autre part, la perception de l'acceptabilité est divisée : la Cam3D est jugée plus utile et apporte un meilleur sentiment d'immersion mais l'utilisabilité est par contre jugée moindre en raison de la fatigue ressentie comme supérieure. C'est la position de l'utilisateur, assis et bras tendu sans point d'appui, qui justifie ce jugement. Par contre, le sexe ou l'expertise en réalité virtuelle n'ont pas influencé l'acceptabilité perçue par les participants.

Cette première étude a mis en valeur les limites de la première version de notre système, car les opérations bi-manuelles n'étaient pas possibles et la zone de tracking était relativement faible. Ces limitations étaient principalement dues à la faible résolution autorisée par la caméra *SwissRanger* (176 x 144 pixels). Nous avons, dans un second temps, amélioré notre système en utilisant un autre type de caméra 3D : la *Kinect* de *Microsoft* sortie un an après nos premiers développements. Celle-ci nous a permis de procéder à de nombreuses améliorations : l'utilisateur peut désormais utiliser ses 2 mains pour réaliser les interactions demandées, plusieurs utilisateurs peuvent travailler de façon collaborative (même si cela n'est pas l'objet de notre étude) et un tracking complet du corps est possible. De plus, nous avons développé d'autres interactions s'ajoutant à la simple sélection d'objets de la première version, comme le déplacement, le redimensionnement et la rotation mais également la navigation dans l'environnement virtuel. Cela a imposé l'implémentation d'un menu permettant à l'utilisateur de choisir l'action qu'il souhaite réaliser.

Cette deuxième version a été comparée à un système plus sophistiqué et davantage en adéquation avec les nouvelles actions proposées que la souris. Ce système est composé de deux gants de données et de capteurs magnétiques qui sont très souvent utilisés en réalité virtuelle pour effectuer des interactions naturelles. L'étude 2 permet de comparer de manière quantitative et qualitative ces deux solutions pour des tâches de manipulation (positionnement, redimensionnement, orientation) et de navigation.

Les résultats de l'étude 2 (qui impliquait 20 participants) montrent que les performances sont similaires entre les deux systèmes étudiés, en ce qui concerne la précision. Les participants ont évoqué un problème de fiabilité dans la reconnaissance de l'état de la main avec la Cam3D, et un problème de tremblements de l'avatar avec les Gants3D dès que la main s'éloignait trop (environ 30 cm) de l'antenne (tracking magnétique). Ces deux points négatifs, inhérents aux solutions techniques retenues, ont pu lisser les résultats des performances. Ainsi, si la détection des mouvements proprement dite (déplacements dans l'espace) avec la Cam3D est d'une précision équivalente à celle obtenue par les Gants3D, la fiabilité de la reconnaissance de la main fermée – main ouverte reste inférieure. Les préférences des participants sont majoritairement en faveur de la Cam3D, principalement pour des critères de confort, de liberté de mouvement et de maniabilité. Le sentiment d'immersion est équivalent entre les deux systèmes pour la moitié des participants, mais supérieur avec la Cam3D pour 8 des 10 autres. De même, l'aisance d'utilisation est ressentie de manière supérieure avec la Cam3D par 9 participants, quand seulement 5 évoquent les Gants3D pour ce critère.

Enfin, l'étude 3 a porté uniquement sur les modalités d'interaction. Nous avons développé, en fonction des remarques, suggestions et commentaires obtenus à partir des questionnaires post-passation de l'expérimentation 2, de nouvelles interactions pour le déplacement, la rotation et le redimensionnement d'objets. Le périphérique qui a été utilisé est le même que dans l'expérimentation 2 (le système que nous avons créé). L'objet de cette étude complémentaire (qui a impliqué 30 participants) est de comparer les nouvelles modalités d'interaction développées avec les « anciennes », issues de la seconde expérimentation. Cette comparaison est effectuée de manière quantitative en étudiant les performances obtenues pour chacune d'elle et de manière qualitative en étudiant les préférences des utilisateurs sur différents critères comme la simplicité d'utilisation, l'intuitivité, la facilité d'assimilation, la précision.

Si les performances sont équivalentes entre les deux modalités pour le déplacement et le redimensionnement, elles sont significativement en faveur de la nouvelle modalité pour la rotation, que ce soit en termes de temps d'exécution, d'erreur de précision ou de ratio « précision / temps ». Ce résultat est important, car c'est la rotation qui avait cristallisé le plus de critiques lors de l'expérimentation 2. Nous avons ainsi apporté une réponse satisfaisante concernant cette interaction, puisqu'elle recueille également la majeure partie des préférences des participants. Ce qui est également le cas du déplacement. Le redimensionnement, par contre, ne semble pas apporter de différence significative, ce qui peut s'expliquer par le fait qu'elle était déjà peu critiquée lors de la seconde expérimentation. Le problème de précision dû à l'ouverture de la main et au recalage de la position terminale du squelette sur la main a disparu. La précision ressentie, bien qu'elle ne soit effective que pour la rotation, est significativement meilleure avec la nouvelle modalité, pour toutes les interactions. Ce problème, qui était surtout gênant pour le déplacement, explique la préférence des participants pour la seconde modalité. De la même façon, la possibilité de décomposer la rotation sur les deux mains rend, pour la majorité des participants, la rotation plus efficace. Par contre, le fait d'utiliser deux mains pour réaliser chaque modalité rend les interactions moins intuitives. La précision a été obtenue au détriment de la simplicité d'utilisation même si l'apprentissage du fonctionnement des interactions reste très rapide.

Ces résultats montrent, d'une part, que de façon globale, le système que nous avons mis au point permet d'ores et déjà d'obtenir des résultats au moins aussi bons que des systèmes couramment utilisés comme une souris classique ou encore des gants de données couplés à des capteurs magnétiques. Cela confirme les hypothèses A et C. Et nous remarquons, d'autre part, que notre système ne nécessite pas d'apprentissage particulier et que les utilisateurs se l'approprient très rapidement. Les résultats obtenus montrent, par ailleurs, une faiblesse importante au niveau de la rotation d'objet qui a été cependant améliorée dans notre 3<sup>ème</sup> étude. Il sera nécessaire d'encore améliorer notre système pour déterminer l'orientation des mains et ainsi faciliter cette action. Enfin, même si la détection de l'état des mains (ouverte ou fermée) et la précision du système peuvent être améliorés, l'acceptabilité du système est déjà très bonne et surpasse même celle des autres périphériques précédemment cités. Le dernier avantage de notre système réside dans son coût, très faible vis-à-vis des solutions commerciales existantes. Le risque demeure que l'engouement actuel pour les caméras time-of-flight, renforcé par le succès de la Microsoft Kinect, ne soit qu'un effet de mode et d'entraînement. S'il n'apporte pas de réelle plus value à l'utilisateur le système pourra être utilisable sans qu'il puisse, pour autant, être qualifié d'utile (Loup-Escande et al., 2011). Il n'en reste pas moins que ces résultats sont très encourageants, et nous incitent fortement à poursuivre nos efforts dans cette direction qui semble très prometteuse.

## 2 Perspectives

Bien que le système que nous avons mis au point soit opérationnel tel quel et permette d'ores et déjà de réaliser des tâches de manipulation d'objets et de navigation en environnement virtuel, il reste néanmoins perfectible sur plusieurs aspects.

La principale amélioration à apporter concerne l'algorithme de détection de l'état des mains (ouvertes ou fermées), qui pourrait être nettement amélioré et permettre une reconnaissance plus fiable et plus robuste. La limitation actuelle est directement liée à l'orientation des mains. Lorsque celles-ci sont faces à la caméra, cela ne pose pas de problèmes particuliers pour déterminer leur état. En revanche, lorsque l'utilisateur effectue les différentes tâches, il arrive qu'il soit attentif uniquement à l'action qu'il est en train de réaliser et il peut arriver que ses mains ne soient plus tout à fait face à la caméra, mais légèrement tournées sur les côtés. Cela pose, dans ce cas, problème étant donné que notre algorithme repose sur les contours de la main pour déterminer son état. Si la main est légèrement tournée, sa forme peut alors être considérée comme une main fermée alors qu'en réalité elle est ouverte et vice versa. Une deuxième caméra pourrait être placée au dessus de la scène afin de lever les ambiguïtés qui peuvent se poser pour ces cas précis. D'autre part, si l'utilisateur sort ses mains du champ de la caméra, il peut arriver que le tracking se perde momentanément, ce qui peut être problématique s'il est en train de manipuler un objet (particulièrement pour le déplacement). Il est cependant possible de corriger rapidement ce problème en gardant en mémoire les dernières valeurs de positions connues ainsi que l'état des mains juste avant la perte du tracking, ce dernier reprenant lorsque les mains sont à nouveau détectées.

Une meilleure détection de l'état des mains permettra au système d'être plus robuste et de fournir de meilleures réponses aux actions des utilisateurs.

Une seconde amélioration souhaitable concerne cette fois l'algorithme fourni par *OpenNI* et qui permet d'appliquer un squelette sur une personne. D'une part, notre système, du fait de l'utilisation de l'algorithme d'*OpenNI*, requiert une phase d'initialisation dans laquelle l'utilisateur doit lever ses bras au niveau de la tête pour former la lettre phi ( $\Psi$ ) afin d'être reconnu par le système comme étant l'utilisateur principal. D'autre part, le squelette appliqué sur la personne, bien que fiable et robuste (en

particulier si la personne sort totalement du champ de vision de la caméra et revient) possède tout de même quelques faiblesses. En effet, l'algorithme permet de déterminer les positions des mains, cependant, ces positions (représentées par des points) sont situées au centre de celles-ci. Par conséquent, lorsque l'utilisateur ouvre ou ferme la main, un léger décalage se produit. Cela pose des problèmes si l'on souhaite réaliser des manipulations de précision comme c'est le cas lors du positionnement, du redimensionnement et de l'orientation d'objets. Ces deux points devraient être très rapidement améliorés car le dernier algorithme fourni par le SDK de Microsoft permet maintenant de ne plus avoir de phase d'initialisation, le squelette est appliqué directement sur la personne sans que celle-ci soit obligée d'adopter une position particulière. De plus, le nombre de points du squelette a été augmenté et permet désormais de connaître les positions des poignets mais également de l'extrémité des mains. Le point représentant le poignet ne dépend plus, en théorie, de l'état (ouvert ou fermée) des mains, ce qui devrait nous permettre d'avoir un système beaucoup plus robuste et permettant une meilleure précision pour les manipulations requérant de la finesse.

N'étant pas capable de déterminer l'orientation des mains, notre système permet de connaître uniquement les positions et l'état de celles-ci. Cela nous a, cependant, permis de développer des modalités d'interactions se basant uniquement sur ces informations. Les résultats obtenus pour les tâches impliquant la rotation, bien que dans l'ensemble relativement proches de ce qui était demandé, montrent que notre système, même s'il est fonctionnel, ne permet pas une orientation simple, intuitive et directe d'un objet. L'acceptation de cette modalité a été moyenne et bien inférieure aux autres modalités proposées. La prochaine version de notre système devra prendre en compte ce constat et être en mesure de fournir de façon fiable l'orientation précise des mains de l'utilisateur. De cette façon, la rotation d'objet devrait être grandement simplifiée et l'interaction proposée devrait pouvoir se rapprocher d'une manipulation d'objet réelle.

Le dernier point qui mériterait d'être amélioré concerne la navigation et l'observation dans l'environnement virtuel. Pour le moment, la modalité de navigation qui a été développée ne permet de se déplacer dans l'environnement virtuel que par des mouvements de translation (gauche - droite, haut - bas, avant - arrière). Nous n'avons pas développé de modalités pour orienter la caméra (la vision de l'utilisateur) ni même pour tourner la caméra sur elle-même. Il n'est, par conséquent, pas possible de regarder ou d'orienter la caméra où l'on veut (derrière soi par exemple, en penchant la tête vers la gauche). Ce point peut être également rapidement corrigé. Une solution simple consisterait à réaliser les déplacements en translation avec la main dominante (ce que l'on fait déjà) et orienter la caméra avec l'autre main. La connaissance de l'orientation des mains simplifierait cette tâche puisque, dans ce cas, l'utilisateur agirait comme s'il tenait la caméra et ainsi les mouvements de celle-ci seraient copiés à l'identique sur ceux de la main, aussi bien en vitesse, en déplacement qu'en orientation.

Pour déterminer si les mains sont ouvertes ou fermées, nous nous sommes basés sur un algorithme qui détermine, à partir des contours des mains, le nombre de doigts visibles. On considère qu'un doigt est levé si la longueur de ce dernier dépasse une certaine taille. De même, si le nombre de doigts levés est supérieur à 2, on considère que la main est ouverte. D'autres solutions pourraient être envisageables et seraient à même de fournir d'excellents résultats comme par exemple l'usage de réseaux de neurones, ou de chaînes de Markov cachées. De même, l'usage de *template matching* pourrait être une solution efficace (c'est d'ailleurs ce principe qui est à la base de l'application du squelette sur une personne dans la *Kinect*). Plus le nombre d'images dans la base de données est grand et plus la détection est efficace. L'apprentissage est relativement long, car il faut apprendre au système à déterminer si une main est ouverte ou fermée à partir de très nombreux exemples. Cependant, une fois l'apprentissage effectué, il

n'est pas nécessaire de le refaire. Cette solution sera envisagée, dans les versions futures de notre système.

Comme nous venons de le voir, le système que nous avons développé permet d'interagir avec les environnements virtuels de façon efficace et peut d'ores et déjà remplacer des solutions commerciales pour un coût moindre. Cependant, les applications potentielles vont bien au-delà du domaine de la réalité virtuelle, les domaines d'application visés étant nombreux. Il est par exemple possible d'imaginer d'utiliser ses gestes pour commander son ordinateur et naviguer dans les applications grâce à ces derniers. *Microsoft* travaille d'ailleurs déjà sur ce thème avec la *Kinect* et compte fournir des applications compatibles dans la prochaine version de son système d'exploitation. Il sera, par exemple, possible de choisir un film dans une liste, le sélectionner pour obtenir des informations, puis le regarder. Différents mouvements permettront de régler différents paramètres comme le son ou encore d'avancer rapidement la vidéo. L'utilisation d'une télécommande ne sera, par conséquent, plus nécessaire et les personnes pourront piloter leur ordinateur à distance, sans médium intermédiaire. Les usages peuvent également s'étendre au domaine de la santé et plus particulièrement de la rééducation, aussi bien des membres supérieurs (comme les bras et les mains) que des membres inférieurs grâce à la capture du corps entier. Une utilisation dans le domaine du handicap est également envisageable : un tel système pourrait, par exemple, faciliter l'interaction avec les enfants et les adultes polyhandicapés qui ont des difficultés lors de la réalisation de mouvements volontaires des membres supérieurs. Les activités actuelles, réalisées à base de commutateurs (Lancioni et al., 2011) pourraient être avantageusement remplacées par un système sans marqueurs, avec un langage relativement simple et adapté à chaque enfant. Un exemple dans lequel s'inscrivent directement nos travaux est le module *HandiCam'3D* faisant partie du projet *AccesSim*. Ce projet a pour objectif d'attirer l'attention et de sensibiliser les pouvoirs publics, les acteurs urbains (maires, architectes...) mais aussi la population, à tous les problèmes de déplacement, d'accessibilité et aux difficultés de motricité urbaine des handicapés moteur. La finalité du projet *AccesSim* est de réaliser un simulateur virtuel de fauteuil roulant en collaboration étroite avec le CEREMH (Centre de Ressources & d'Innovation Mobilité Handicap) et EDF R&D. Dans ce cadre, le module *HandiCam'3D* est un système de capture de mouvements qui permet de détecter les gestes de l'utilisateur et d'évaluer, en temps réel, les actions et les efforts réalisés, ils seront par ailleurs analysés et permettront de déclencher des interactions. En effet, lorsqu'un utilisateur de fauteuil roulant manuel se déplace, il s'aide des objets de l'environnement (bureaux, rampes, murs, etc.). Il est alors nécessaire, afin de reproduire au mieux les conditions réelles, de pouvoir capter les gestes qui sont effectués par l'utilisateur. Le système que nous avons créé peut être très rapidement intégré à ce module.

En dehors de la santé et du handicap, on peut par exemple imaginer des utilisations de notre système dans divers domaines comme en domotique où il serait possible de piloter différents appareils uniquement avec les gestes (baisser la lumière, descendre ou monter les volets électriques, etc.). Ou encore dans l'art, la musique où il pourrait être intéressant d'enregistrer les mouvements des doigts d'un virtuose. Enfin, il pourrait être intéressant d'utiliser, à terme, un système transparent de capture de mouvements dans le domaine du sport, pour évaluer un geste sportif par exemple, ou enregistrer un professionnel en train d'effectuer un mouvement en vue de l'analyser et de le reproduire. Bien entendu, la capture doit être améliorée et une détection plus fine permettant de déterminer l'orientation de chaque doigt doit être possible avant d'envisager de telles applications. On pourrait également envisager une utilisation d'un tel système, par le grand public, où un utilisateur lambda pourrait évaluer son geste par rapport à celui d'un professionnel. Un dernier domaine évident est tout ce qui se rapporte au monde des jeux vidéo, mais celui-ci est déjà naturellement très exploité par *Microsoft*. Ce ne sont que des exemples parmi d'autres mais l'étendue des applications possibles est très vaste.

Il est certain que nous trouverons de plus en plus ce genre d'applications dans notre quotidien, sans que nous en soyons forcément conscients, du fait de leur transparence. A terme, nous sommes persuadés que ces nouvelles façons d'interagir avec l'environnement (réel ou virtuel) feront partie intégrante de nos vies. Il reste néanmoins à améliorer les algorithmes de capture de mouvements, les caméras 3D utilisées (plus grandes résolutions, plus grand taux de rafraîchissement, meilleure précision, etc.) et également améliorer les modalités d'interactions en 3D. Ces perspectives enthousiasmantes nous incitent à poursuivre nos efforts et nos travaux dans cette voie passionnante et pleine de promesses.

# Annexes

---



## Annexe A

### Etude 1 – Fiche explicative pour la manipulation à la souris

#### Manipulation à l'aide de la souris

Lors de cette phase de l'expérimentation, vous allez manipuler une main virtuelle à l'aide d'un périphérique 2D : la souris.

Le but de l'expérimentation, est de faire disparaître par ordre croissant les cubes numérotés de 1 à 10 répartis dans l'espace 3D.

Pour ce faire, il vous faut bouger la souris normalement sur la table afin de positionner la main virtuelle dans le plan de l'écran en face de vous (figure 1). Pour ajuster la position en profondeur, il faut cliquer sur le bouton droit de la souris et la déplacer verticalement (figure 2).

Une fois la main virtuelle correctement positionnée sur le cube, un clic sur le bouton gauche de la souris le fera disparaître (figure 3). Une fois le cube disparu, passez au suivant. L'expérimentation se termine lorsque plus aucun cube n'apparaît à l'écran.

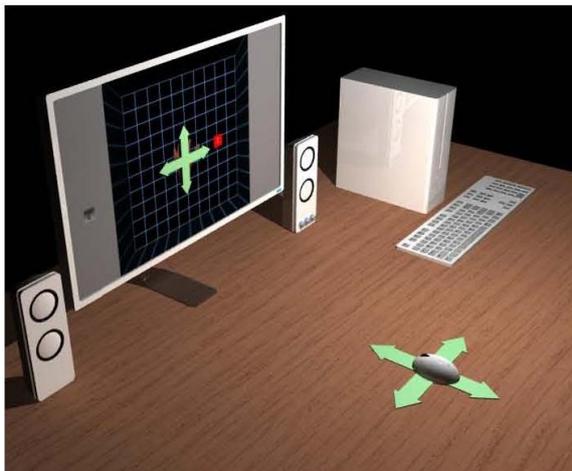


Figure 1

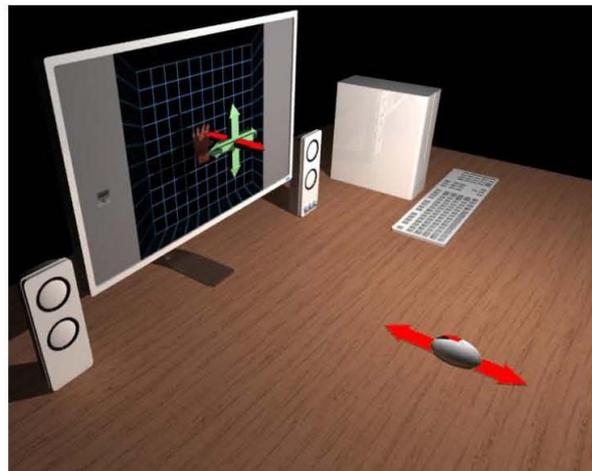
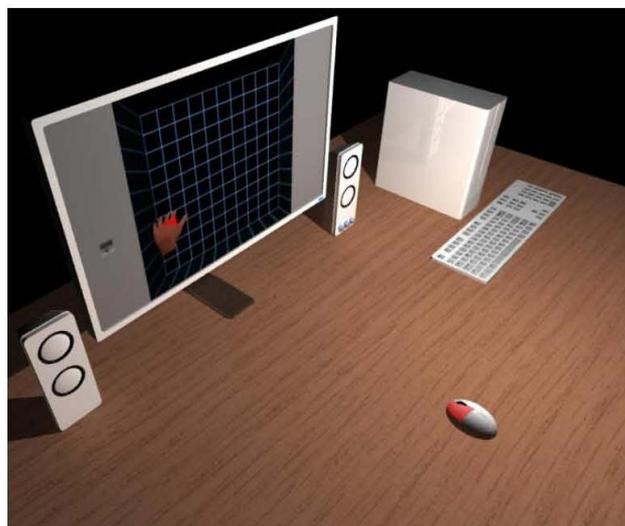


Figure 2



## Manipulation à l'aide du **Systeme 3D**

Lors de cette phase de l'expérimentation, vous allez manipuler une main virtuelle à l'aide d'un périphérique 3D transparent.

Le but de l'expérimentation, est de faire disparaître par ordre croissant les cubes numérotés de 1 à 10 répartis dans l'espace 3D.

Pour ce faire, il vous faut positionner votre main (bras tendu) dans l'espace juste devant vous, comme si vous vouliez toucher un objet réel. Il est important de garder la main bien ouverte (doigts écartés) lors de vos déplacements dans l'espace (Figure 1).

Une fois la main virtuelle correctement positionnée sur le cube, il vous suffit de fermer la main (avoir le poing fermé) comme si vous vouliez prendre l'objet, pour faire disparaître le cube (Figure 2). Une fois le cube disparu, passez au suivant. L'expérimentation se termine lorsque plus aucun cube n'apparaît à l'écran.

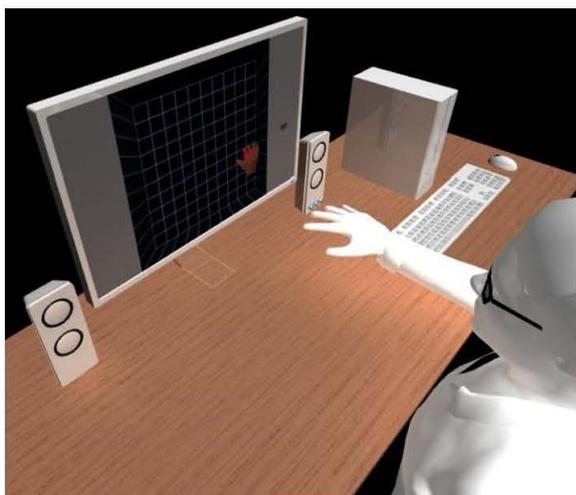


Figure 1

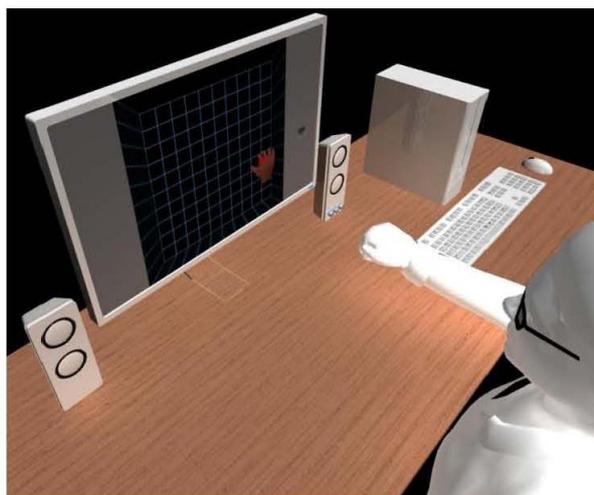


Figure 2

## Etude 1 – Questionnaire d'identification

### PRESENTATION DE L'EXPERIMENTATION

Nous vous invitons à participer à l'évaluation de deux types d'interface de capture de mouvements pour l'interaction 3D temps réel en environnement virtuel :

- Une interface « classique » : la souris
- Une interface « sans capteurs » : la caméra 3D

L'expérimentation vise à faire disparaître les cubes colorés et numérotés de 1 à 10, par ordre croissant. Pour ce faire, vous devez positionner l'avatar "main" sur les cubes, puis cliquer (au moyen de la souris) ou fermer la main (au moyen de la caméra 3D) pour les faire disparaître.

Dans un objectif scientifique, la trace de vos interactions ainsi que votre activité gestuelle (déplacement de la souris et de la main) seront enregistrées.

Cette expérimentation préservera votre anonymat conformément à la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Les informations sollicitées dans ce questionnaire nous serviront à préciser votre profil en tant qu'utilisateur de logiciels grand public. Elles resteront confidentielles.

Nous vous remercions d'avoir accepté de participer à cette expérimentation.

#### INFORMATIONS PASSATION

Numéro de passage

Date

#### INFORMATIONS PERSONNELLES

Nom

Tél

Prénom

E-mail

Age

Latéralisation

Gaucher

Droitier

Ambidextre

#### CONNAISSANCES GENERALES

Dernier diplôme obtenu (Niveau - spécialité; ex : master - MNRV) :

Diplôme en cours de préparation (Niveau - spécialité; ex : master - MNRV) :

## COMPETENCES INFORMATIQUES

### Avez-vous des connaissances en réalité virtuelle ?

Pas du tout      Excellentes connaissances

### Quels dispositifs avez-vous déjà utilisé et à quelle fréquence ?

	Jamais	Peu	Parfois	Souvent	Très souvent
Clavier	<input type="radio"/>				
Souris	<input type="radio"/>				
Souris 3D	<input type="radio"/>				
Ecran tactile	<input type="radio"/>				
Commande vocale	<input type="radio"/>				
Tablette graphique	<input type="radio"/>				
Joystick	<input type="radio"/>				
Manette de jeu	<input type="radio"/>				
Bras à retour d'effort	<input type="radio"/>				
Gant de données	<input type="radio"/>				

Autres, précisez :

### Quelles activités informatiques utilisant la 3D pratiquez vous?

	Jamais	Peu	Parfois	Souvent	Très souvent
Jeux vidéo à moteur 3D	<input type="radio"/>				
Logiciels de CAO (Autocad Architectural Desktop, sketchup)	<input type="radio"/>				
Logiciels de dessin, modeling (3DS max, Catia, Maya)	<input type="radio"/>				
Logiciels de création de scènes interactives et temps réel (Virtools, Unity, OpenScene Graph, Ogre)	<input type="radio"/>				

Autres, précisez :

Envoyer par messagerie

## Etude 1 – Questionnaire post-passation

### QUESTIONNAIRE POST-PASSATION

#### INFORMATIONS PASSATION

Numéro de passage

Date

*Pour simplifier, on désignera, dans le reste du document, le système de capture de mouvements de la main sans capteurs par : le système Cam3D.*

⇒ Par quel système, souris ou Cam 3D, avez vous commencé l'expérimentation ?

Souris  Cam 3D

#### IMPRESSIONS GENERALES

⇒ Etes vous resté concentré tout au long de l'expérimentation ?

Pas du tout      Tout le temps

⇒ Quel système, souris ou Cam 3D, vous a semblé le plus facile à utiliser ?

Souris  Cam 3D

Pourquoi?

⇒ Quel système, souris ou Cam 3D, vous a semblé le moins fatiguant ?

Souris  Cam 3D

Pourquoi?

⇒ Quel système, souris ou Cam 3D, vous a semblé le plus efficace, concernant :  
- La rapidité pour se positionner dans l'espace virtuel 3D ?

Souris  Cam 3D

Pourquoi?

- La précision du positionnement dans l'espace virtuel 3D ?

Souris  Cam 3D

Pourquoi?

## IMPRESSIONS GENERALES

⇒ La vue en perspective vous a-t-elle gênée ?

Oui  Non

Si oui, pourquoi ?

⇒ Avez-vous eu le sentiment d'être le plus immergé lors de l'utilisation de la souris ou du système Cam 3D ?

Souris  Cam 3D

Pourquoi ?

⇒ Quel dispositif, souris ou Cam3D, vous a semblé le plus approprié pour la tâche à réaliser ?

Souris  Cam 3D

Pourquoi ?

⇒ D'une manière générale, avec quel dispositif vous êtes vous sentis le plus à l'aise ?

Souris  Cam 3D

Pourquoi ?

⇒ D'une manière générale, quel dispositif vous a semblé le plus efficace ?

Souris  Cam 3D

Pourquoi ?

⇒ Avez-vous trouvé la tâche à réaliser plutôt :

Très facile  Facile  Sans difficulté particulière  Difficile  Très difficile  Infaisable

## IMPRESSIONS GENERALES

⇒ L'expérimentation vous a semblé :

Longue                      Courte  
Difficile                     Facile  
Ennuyeuse                     Attrayante

⇒ Avez-vous ressenti de la fatigue au niveau du bras ou de la main lors de l'utilisation de la souris :

Aucune                      Elevée

⇒ Avez-vous ressenti de la fatigue au niveau du bras ou de la main lors de l'utilisation de la Cam3D :

Aucune                      Elevée

⇒ Comment jugez-vous l'apport de la Cam3D pour les interactions en environnement 3D :

Aucun                      Elevé

Pourquoi ?

⇒ Quelles sont vos suggestions d'améliorations à apporter, concernant :

- Le système de capture de mouvements 3D transparent (Cam3D) :

- La vue perspective :

- Les fonctionnalités, actions ou gestes supplémentaires que vous aimeriez pouvoir effectuer :

- La fluidité de l'application (affichage ou réaction du système Cam 3D) :

- Autres :

## IMPRESSIONS GENERALES

⇒ En tant que concepteur de logiciels grand public, quelles applications envisageriez-vous pour ce nouveau type d'interface ?

⇒ Problèmes, difficultés et incidents rencontrés :  
(Les résumer brièvement en précisant quel périphérique était utilisé lors du problème rencontré)

MERCI POUR VOTRE CONTRIBUTION

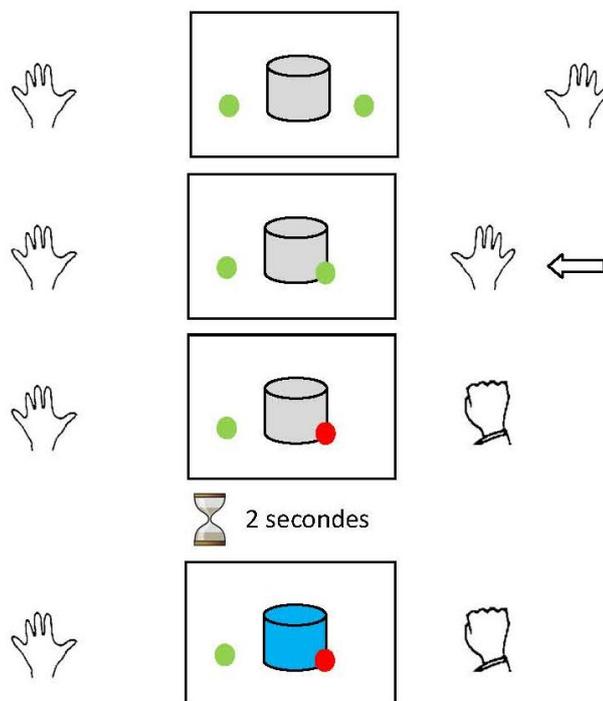
Envoyer par messagerie

## Annexe B

### Etude 2 – Fiche explicative sur la sélection / désélection

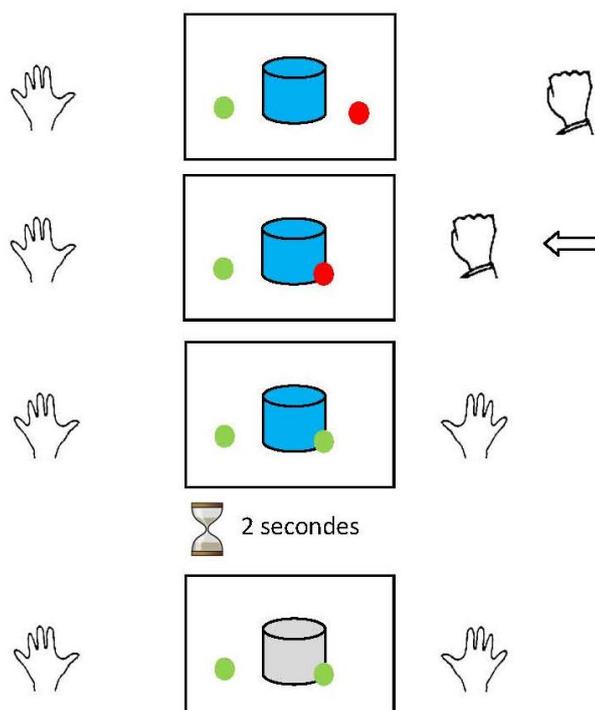
#### Sélection d'objet :

Déplacez votre main dominante sur l'objet à sélectionner, fermez votre main dessus pendant 2 secondes. Lorsque l'objet se met en bleu, il est alors sélectionné.



#### Désélection d'objet :

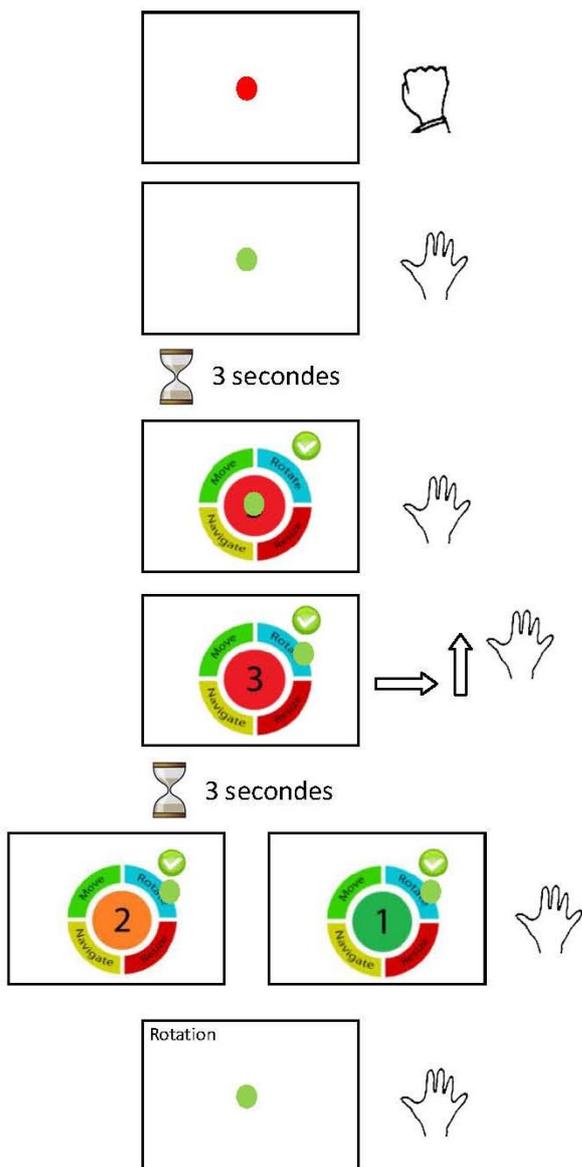
Déplacez votre main dominante sur l'objet à désélectionner, maintenez votre main ouverte dessus pendant 2 secondes. Lorsque l'objet reprend sa couleur d'origine, il est alors désélectionné.



## Etude 2 – Fiche explicative sur l'appel du menu et sélection d'une action

### Appel du menu

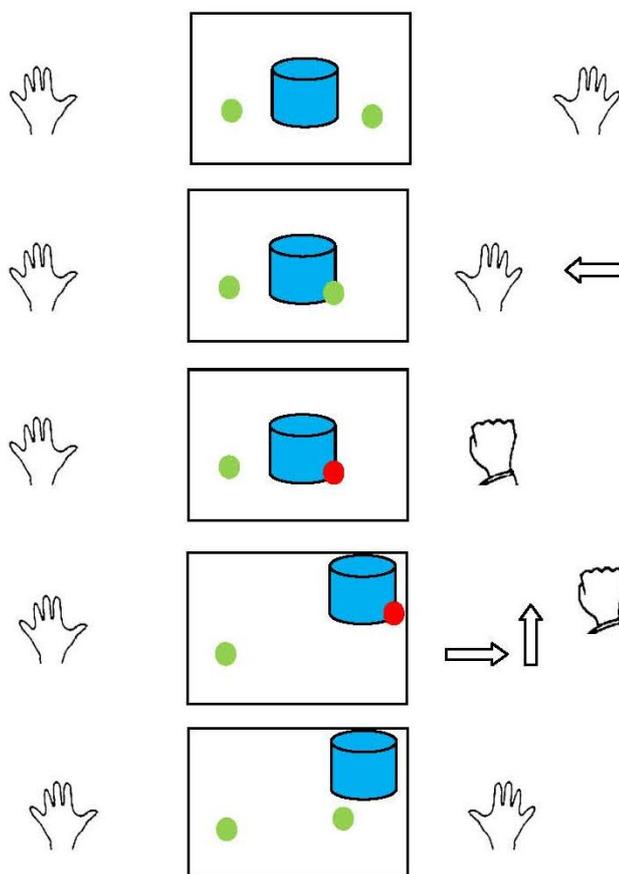
Fermez votre main dominante et ouvrez-la rapidement sans bouger de place. Maintenez votre main ouverte pendant 3 secondes et le menu apparaîtra. Déplacer votre main (toujours ouverte) sur un sous menu pendant 3 secondes. Un compte à rebours apparaît au centre du menu, une fois le compte à rebours à 0, votre sous menu est sélectionné. Un texte en haut à gauche de l'écran vous indiquera quel sous menu est actuellement sélectionné.



## Etude 2 – Fiche explicative sur le déplacement d'un objet

### Déplacer un objet :

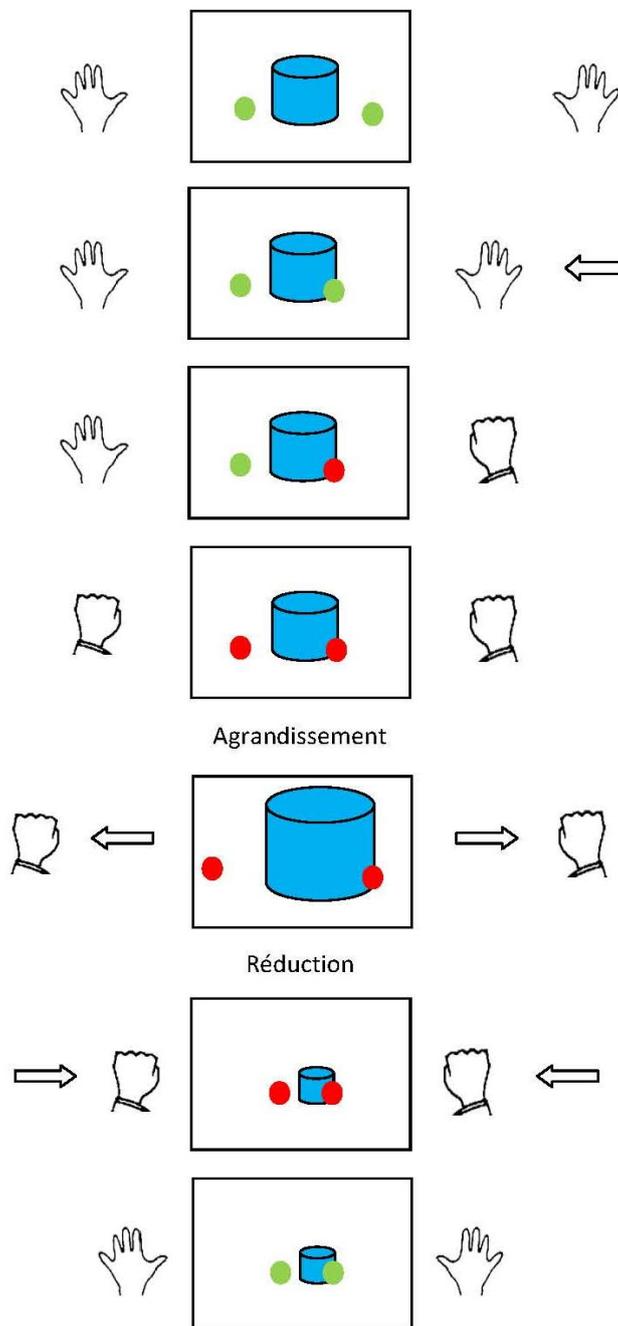
Une fois l'objet sélectionné, positionnez votre main dominante fermée dessus. Déplacez votre main dominante à l'endroit voulu tout en la maintenant fermée. Ouvrez votre main dominante lorsque l'objet est bien positionné. Pensez à désélectionner l'objet une fois la manipulation terminée.



## Etude 2 – Fiche explicative sur le redimensionnement d'un objet

### Redimensionner un objet :

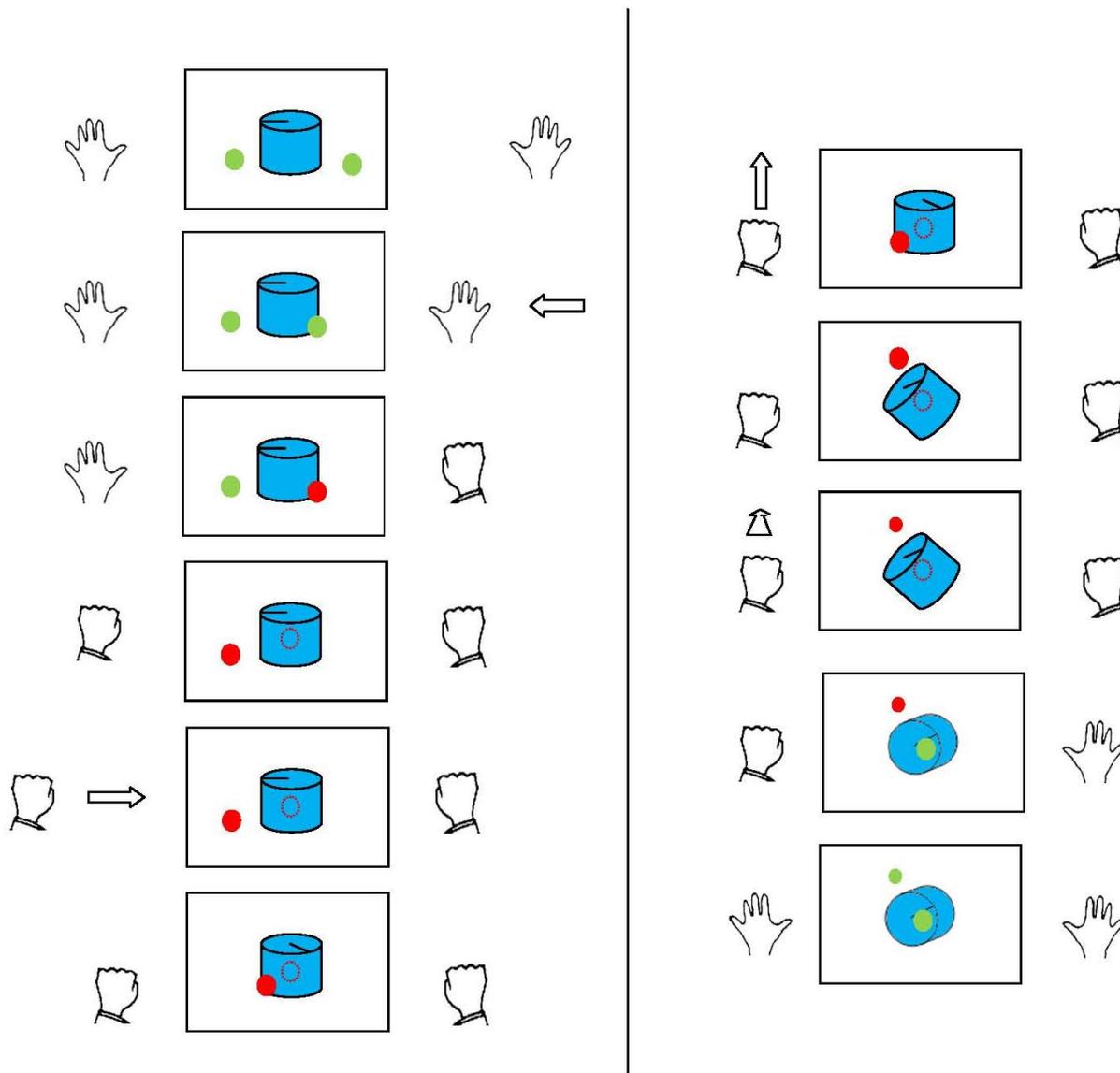
Une fois l'objet sélectionné, positionnez votre main dominante fermée dessus et fermez votre deuxième main pour autoriser le redimensionnement. Ecartez ou rapprochez vos mains pour redimensionner l'objet. Ecarter les mains permet d'agrandir l'objet, les rapprocher permet de le réduire. Une fois le redimensionnement effectué, ouvrez vos mains. Pensez à désélectionner l'objet une fois celui-ci bien positionné.



## Etude 2 – Fiche explicative sur la rotation d'un objet

### Rotation d'un objet :

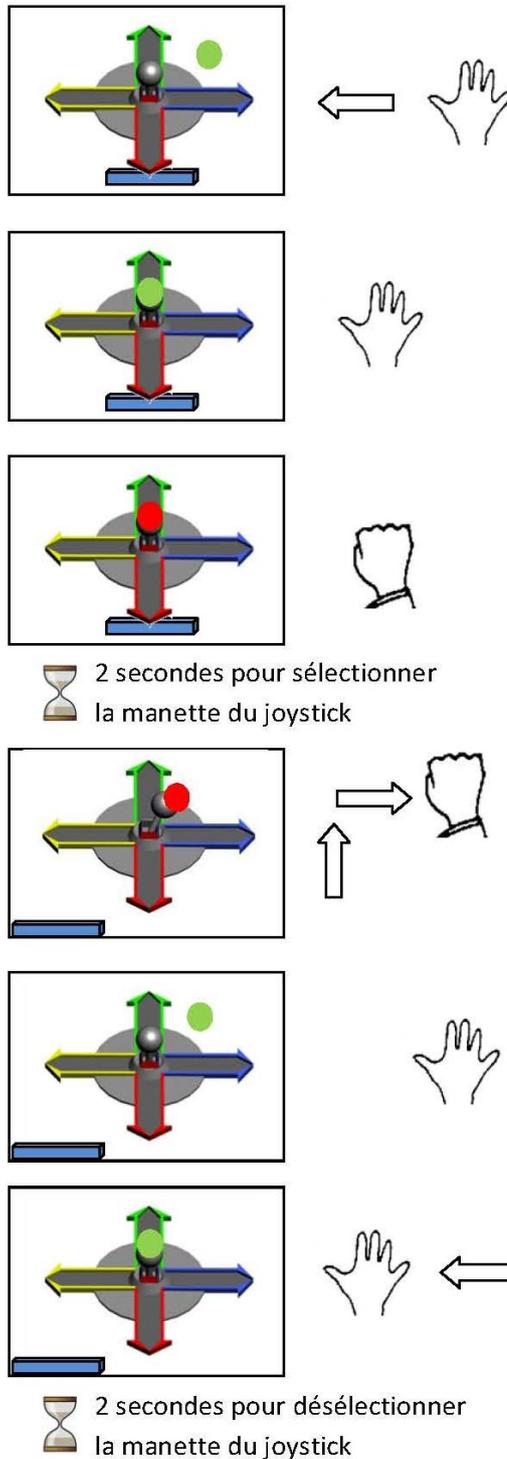
Une fois l'objet sélectionné, positionnez votre main dominante fermée dessus, elle disparaîtra et ira se positionner automatiquement au centre de l'objet. Fermez votre deuxième main pour autoriser la rotation. Déplacer votre deuxième main selon l'axe des X permettra à l'objet de tourner dans cette même direction. Il en est de même pour les axes Y et Z. Une fois la rotation effectuée, ouvrez votre main dominante d'abord puis la deuxième. Pensez à désélectionner l'objet une fois celui-ci bien orienté.



## Etude 2 – Fiche explicative sur la navigation en EV

### Navigation

Une fois le menu « navigation » sélectionné, un joystick apparaît. Pour naviguer fermez votre main dominante sur la poignée du joystick pendant 2 secondes afin de la sélectionner et déplacer la dans la direction vers laquelle vous souhaitez aller. Une fois au bon endroit ouvrez la main pour stopper la navigation. Plus l'amplitude du mouvement de votre main est grande, plus le déplacement est rapide. Plus l'amplitude est faible, plus vous vous déplacerez précisément. Pensez à désélectionner la manette du joystick une fois que vous êtes à l'endroit voulu.



## Consignes

Cette expérimentation a pour but de vous demander d'évaluer :

- 2 systèmes de capture de mouvements (Cam3D et Gants3D) ;
- les modalités d'interactions utilisées pour interagir avec l'environnement virtuel (sélection d'objet, déplacement, rotation, redimensionnement et navigation dans l'environnement virtuel).

Pour cela, la manipulation va se décomposer en 9 tâches que vous devrez réaliser dans l'ordre.

0°) Pour commencer la manipulation, placez votre main dominante ouverte (les doigts écartés) au centre de l'écran pendant 3 secondes. Votre main virtuelle doit se superposer à la main fixe dessinée à l'écran. Lorsque la main disparaît, l'expérimentation commence.

### TACHES 1 à 4 – environnement « Paysage »

#### Tâche 1

- 1°) Sélectionnez le **Puits** en haut à droite.
- 2°) Affichez le menu et choisissez **Move** (Déplacer) pour déplacer le puit
- 3°) Déplacez le puits pour le faire correspondre au mieux à sa représentation transparente en bas à gauche.
- 4°) Désélectionnez le puits

#### Tâche 2

- 5°) Sélectionnez le **Cheval**
- 6°) Affichez le menu et choisissez **Resize** (Redimensionner) pour redimensionner le cheval
- 7°) Redimensionnez le cheval pour le faire correspondre au mieux à sa représentation transparente.
- 8°) Désélectionnez le cheval.

#### Tâche 3

- 9°) Sélectionnez la **Maison**
- 10°) Affichez le menu et choisissez **Rotate** (Tourner) pour orienter la maison
- 11°) Orientez la maison pour la faire correspondre au mieux à sa représentation transparente.
- 12°) Désélectionnez la maison

#### Tâche 4

- 13°) Affichez le menu et choisissez **Navigate** (Naviguer) pour vous déplacer vers la **Scène 2**.
- 14°) Sélectionnez la manette du **Joystick** et déplacez vous vers la scène 2 en **suivant les flèches**.
- 15°) Désélectionnez la manette lorsque vous êtes correctement positionné sur la scène 2.

## TACHES 5 à 7 – environnement « Bureau »

### Tâche 5

- 1°) Sélectionnez le **Globe** lorsque vous êtes correctement positionné sur la scène 2.
- 2°) Affichez le menu et choisissez **Move** (Déplacer) pour déplacer le Globe.
- 3°) Déplacez le Globe à gauche pour le faire correspondre au mieux à sa représentation transparente à droite.
- 4°) Affichez le menu et choisissez **Resize** (Redimensionner) pour redimensionner le globe
- 5°) Redimensionnez le globe pour le faire correspondre au mieux à sa représentation transparente.
- 6°) **Réitérez les étapes 2 à 5 autant de fois que vous jugerez nécessaire.**
- 7°) Désélectionnez le globe.

### Tâche 6

- 8°) Sélectionnez l'**Horloge** à droite.
- 9°) Affichez le menu et choisissez **Move** (Déplacer) pour déplacer l'horloge.
- 10°) Déplacez l'horloge pour la faire correspondre au mieux à sa représentation transparente.
- 11°) Affichez le menu et choisissez **Rotate** (Tourner) pour orienter l'horloge.
- 12°) Orientez l'horloge pour la faire correspondre au mieux à sa représentation transparente.
- 13°) **Réitérez les étapes 9 à 12 autant de fois que vous jugerez nécessaire.**
- 14°) Désélectionnez l'horloge.

### Tâche 7

- 15°) Sélectionnez l'**Ordinateur portable**.
- 16°) Affichez le menu et choisissez **Rotate** (Tourner) pour orienter l'ordinateur portable
- 17°) Orientez l'ordinateur portable pour le faire correspondre au mieux à sa représentation transparente.
- 18°) Affichez le menu et choisissez **Resize** (Redimensionner) pour redimensionner l'ordinateur portable
- 19°) Redimensionnez l'ordinateur portable pour le faire correspondre au mieux à sa représentation transparente.
- 20°) **Réitérez les étapes 16 à 19 autant de fois que vous jugerez nécessaire.**
- 21°) Désélectionnez l'ordinateur portable.

## TACHES 8 ET 9 – environnement « Bateau »

### Tâche 8

- 1°) Affichez le menu et choisissez **Navigate** (Naviguer) pour vous déplacer vers la **scène 3**.
- 2°) Sélectionnez la manette du **Joystick** et déplacez vous vers la scène 3 en **suivant les flèches**.
- 3°) Désélectionnez la manette lorsque vous êtes correctement positionné sur la scène 3.

### Tâche 9

- 4°) Sélectionnez le **Bateau**.
- 5°) Affichez le menu et choisissez **Move** (Déplacer) pour déplacer le bateau.
- 6°) Déplacez le bateau à gauche pour le faire correspondre au mieux à sa représentation transparente à droite.
- 7°) Affichez le menu et choisissez **Rotate** (Tourner) pour orienter le bateau.
- 8°) Orientez le bateau pour le faire correspondre au mieux à sa représentation transparente.
- 9°) Affichez le menu et choisissez **Resize** (Redimensionner) pour redimensionner le bateau.
- 10°) Redimensionnez le bateau pour le faire correspondre au mieux à sa représentation transparente.
- 11°) **Réitérez les étapes 5 à 10 autant de fois que vous jugerez nécessaire.**
- 12°) Désélectionnez le bateau
- 13°) Affichez le menu et choisissez l'icône verte en haut à droite « **Check** » (Valider) pour terminer l'expérimentation.

## Etude 2 – Questionnaire d'identification

### PRESENTATION DE L'EXPERIMENTATION

Nous vous invitons à participer à l'évaluation de plusieurs modalités d'interaction permettant la manipulation d'objets 3D en environnement virtuel (E. V.) et ce en utilisant un système de capture de mouvements sans marqueurs.

L'expérimentation vise à effectuer différentes interactions en E. V. qui sont le déplacement, la rotation et le redimensionnement d'objets 3D. Toutes ces actions se regroupent sous le terme de manipulation d'objets et constituent une tâche très importante en réalité virtuelle. Pour cela, vous devrez utiliser vos 2 mains dans l'espace entre vous et l'écran et ouvrir ou fermer les mains pour effectuer les actions demandées. Le but est de comparer deux modalités d'interactions pour chacune des tâches précédemment citées.

Dans un objectif scientifique, la trace de vos interactions ainsi que votre activité gestuelle (déplacement des mains, ouverture, fermeture, etc.) seront enregistrées.

Cette expérimentation préservera votre anonymat conformément à la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Les informations sollicitées dans ce questionnaire nous serviront à préciser votre profil en tant qu'utilisateur de logiciels grand public. Elles resteront confidentielles.

Nous vous remercions d'avoir accepté de participer à cette expérimentation.

#### INFORMATIONS PASSATION

Numéro de passage

Date

#### INFORMATIONS PERSONNELLES

Nom

Tél

Prénom

E-mail

Age

Latéralisation  
 Gaucher  
 Droitier  
 Ambidextre

#### CONNAISSANCES GENERALES

Dernier diplôme obtenu (Niveau - spécialité; ex : master - MNRV) :

Diplôme en cours de préparation (Niveau - spécialité; ex : master - MNRV) :

## COMPETENCES INFORMATIQUES

### Avez-vous des connaissances en réalité virtuelle ?

Pas du tout      Excellentes connaissances

### Quels dispositifs de réalité virtuelle avez-vous déjà utilisés et à quelle fréquence ?

	Jamais	Peu	Parfois	Souvent	Très souvent
Souris 3D	<input type="radio"/>				
Joystick	<input type="radio"/>				
Manette de jeu	<input type="radio"/>				
Bras à retour d'effort	<input type="radio"/>				
Gant de données	<input type="radio"/>				

Autres, précisez :

### Quelles activités informatiques utilisant la 3D pratiquez vous?

	Jamais	Peu	Parfois	Souvent	Très souvent
Jeux vidéo à moteur 3D	<input type="radio"/>				
Logiciels de CAO (Autocad Architectural Desktop, sketchup)	<input type="radio"/>				
Logiciels de dessin, modeling (3DS max, Catia, Maya)	<input type="radio"/>				
Logiciels de création de scènes interactives et temps réel (Virttools, Unity, OpenScene Graph, Ogre)	<input type="radio"/>				

Autres, précisez :

### Avez vous déjà été confronté à des interactions bi-manuelles?

Oui  Non

Si oui, avec quel(s) périphérique(s) :

Et pour quelle(s) application(s) :

Envoyer par messagerie

## Etude 2 – Questionnaire post-passation

### QUESTIONNAIRE POST-PASSATION

#### INFORMATIONS PASSATION

Numéro de passage

Date

*Pour simplifier, on désignera, dans le reste du document, le système de capture de mouvements de la main sans capteurs par : "Cam3D" et le système gants + capteurs magnétiques par "Gants".*

#### IMPRESSIONS GENERALES

⇒ Par quel système, Gants ou Cam 3D, avez vous commencé l'expérimentation ?

Gants  Cam 3D

⇒ Etes vous resté concentré tout au long de l'expérimentation ?

Pas du tout      Tout le temps

⇒ La vue en perspective vous a-t-elle gênée ?

Oui  Non

Si oui, pourquoi ?

⇒ Avez-vous trouvé la tâche à réaliser plutôt :

Très facile  Facile  Sans difficulté particulière  Difficile  Très difficile  Infaisable

⇒ L'expérimentation vous a semblé :

Très longue      Très courte

Très difficile      Très facile

Très ennuyeuse      Très attractive

⇒ Comment évaluez vous le **degrés de difficulté d'assimilation** des modalités d'interaction pour :

	Très difficile	Difficile	Normal	Facile	Très facile
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

## QUESTIONS SUR LES MODALITES D'INTERACTION

⇒ Comment évaluez-vous (de très mauvaise à très bonne) l'**efficacité globale** pour :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **rapidité** pour :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **précision** pour :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **simplicité** pour :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) l'**intuitivité de l'action** pour :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacer un objet :	<input type="radio"/>				
Tourner un objet :	<input type="radio"/>				
Redimensionner un objet :	<input type="radio"/>				
Naviguer dans l'Environnement virtuel :	<input type="radio"/>				

## QUESTIONS SUR LES MODALITES D'INTERACTION

### DEPLACEMENT :

⇒ Quels sont, selon vous, les points positifs ?

⇒ Quels sont, selon vous, les points négatifs ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction des déplacements ?

### ROTATION :

⇒ Quels sont, selon vous, les points positifs ?

⇒ Quels sont, selon vous, les points négatifs ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction des rotations ?

### REDIMENSIONNEMENT :

⇒ Quels sont, selon vous, les points positifs ?

⇒ Quels sont, selon vous, les points négatifs ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction du redimensionnement ?

### NAVIGATION :

⇒ Quels sont, selon vous, les points positifs ?

⇒ Quels sont, selon vous, les points négatifs ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction de la navigation ?

## QUESTIONS SUR LES GANTS

⇒ Comment jugez-vous le système gants + capteur concernant les critères suivants :

Poids :	Très lourd <input type="radio"/>	Lourd <input type="radio"/>	Moyen <input type="radio"/>	Léger <input type="radio"/>	Très léger <input type="radio"/>
Maniabilité :	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>
Liberté de mouvement :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Fatigue engendrée :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Facilité d'utilisation :	Très difficile <input type="radio"/>	Difficile <input type="radio"/>	Normale <input type="radio"/>	Facile <input type="radio"/>	Très facile <input type="radio"/>
Confort d'utilisation :	Très mauvais <input type="radio"/>	Mauvais <input type="radio"/>	Moyen <input type="radio"/>	Bon <input type="radio"/>	Très bon <input type="radio"/>
Efficacité globale :	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>
Précision :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Moyenne <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Stabilité :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Fiabilité de la reconnaissance : (ouverture / fermeture de la main)	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>

### QUESTIONS SUR LA CAM3D

⇒ Comment jugez-vous le système Cam3D concernant les critères suivants :

Maniabilité :	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>
Liberté de mouvement :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Fatigue engendrée :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Facilité d'utilisation :	Très difficile <input type="radio"/>	Difficile <input type="radio"/>	Normale <input type="radio"/>	Facile <input type="radio"/>	Très facile <input type="radio"/>
Confort d'utilisation :	Très mauvais <input type="radio"/>	Mauvais <input type="radio"/>	Moyen <input type="radio"/>	Bon <input type="radio"/>	Très bon <input type="radio"/>
Efficacité globale :	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>
Précision :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Moyenne <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Stabilité :	Très faible <input type="radio"/>	Faible <input type="radio"/>	Correcte <input type="radio"/>	Grande <input type="radio"/>	Très grande <input type="radio"/>
Fiabilité de la reconnaissance : (ouverture / fermeture de la main)	Très mauvaise <input type="radio"/>	Mauvaise <input type="radio"/>	Moyenne <input type="radio"/>	Bonne <input type="radio"/>	Très bonne <input type="radio"/>

COMPARAISON ENTRE LES 2 SYSTEMES

- |   | Gants                 | Cam3D                 | Egal                  |
|---|-----------------------|-----------------------|-----------------------|
| ⇒ De manière générale, quel système avez-vous préféré?                      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ De manière générale, quel système vous a semblé le plus approprié?        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ Avec quel système vous êtes vous senti le plus immergé ?                  | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ De manière générale, quel système vous a semblé le plus précis ?          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ De manière générale, quel système avez-vous préféré utiliser ?            | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ Avec quel système vous êtes vous senti le plus à l'aise ?                 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| ⇒ Quel système vous a semblé le plus facile à utiliser pour le déplacement? | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Justifiez :

- ⇒ Quel système vous a semblé le plus facile à utiliser pour la rotation?

Justifiez :

- ⇒ Quel système vous a semblé le plus facile à utiliser pour le redimensionnement?

Justifiez :

- ⇒ Quel système vous a semblé le plus facile à utiliser pour la navigation?

Justifiez :

- ⇒ Quel système vous a semblé le plus précis pour le déplacement?

Justifiez :

- ⇒ Quel système vous a semblé le plus précis pour la rotation?

Justifiez :

- ⇒ Quel système vous a semblé le plus précis pour le redimensionnement?

Justifiez :

- ⇒ Quel système vous a semblé le plus précis pour la navigation?

Justifiez :

MERCI POUR VOTRE CONTRIBUTION

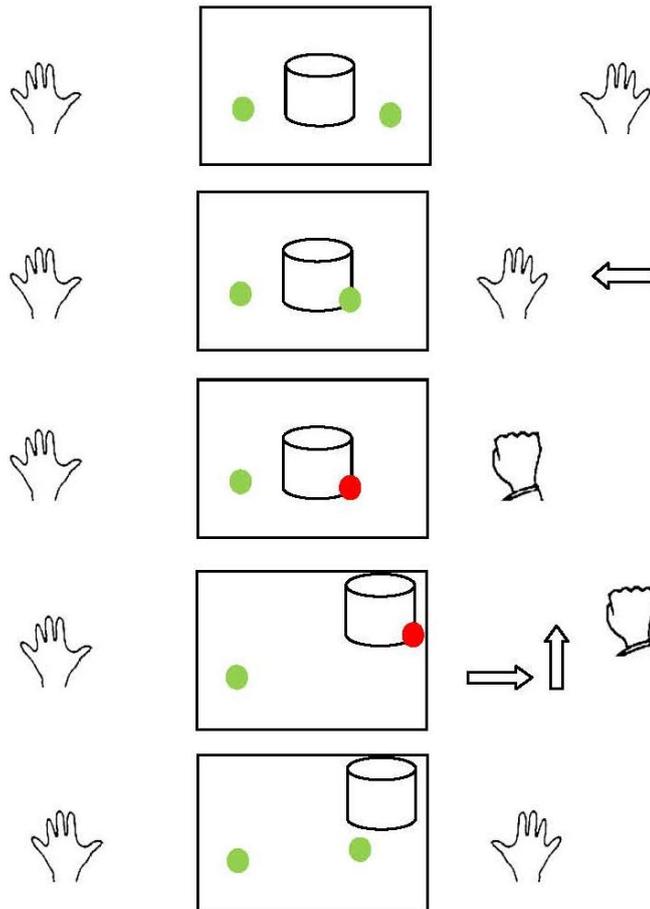
Envoyer par messagerie

## Annexe C

### Etude 3 – Fiche explicative sur la modalité 1 du déplacement

#### Déplacement d'objet 1 :

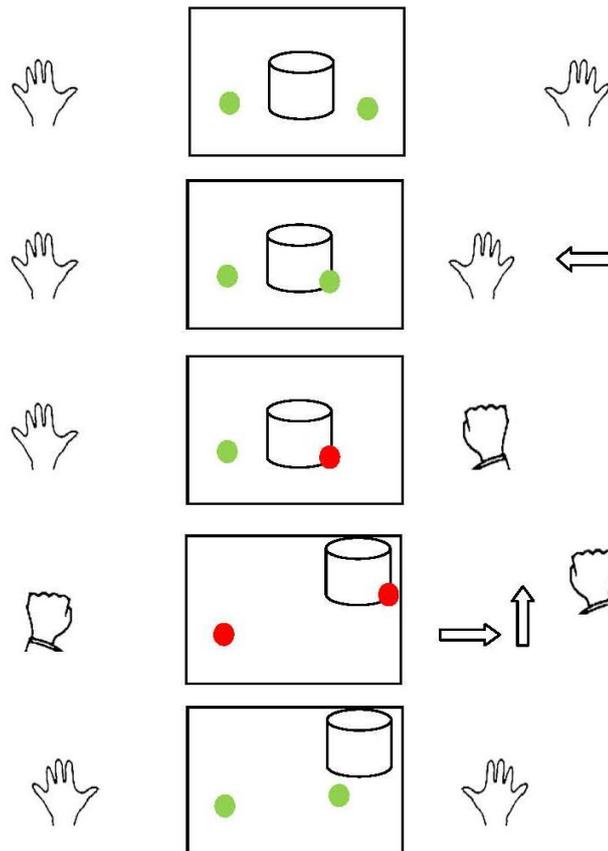
Positionnez votre main dominante fermée sur l'objet. Déplacez votre main dominante à l'endroit voulu tout en la maintenant fermée. Ouvrez votre main dominante lorsque l'objet est bien positionné.



### Etude 3 – Fiche explicative sur la modalité 2 du déplacement

#### Déplacement d'objet 2 :

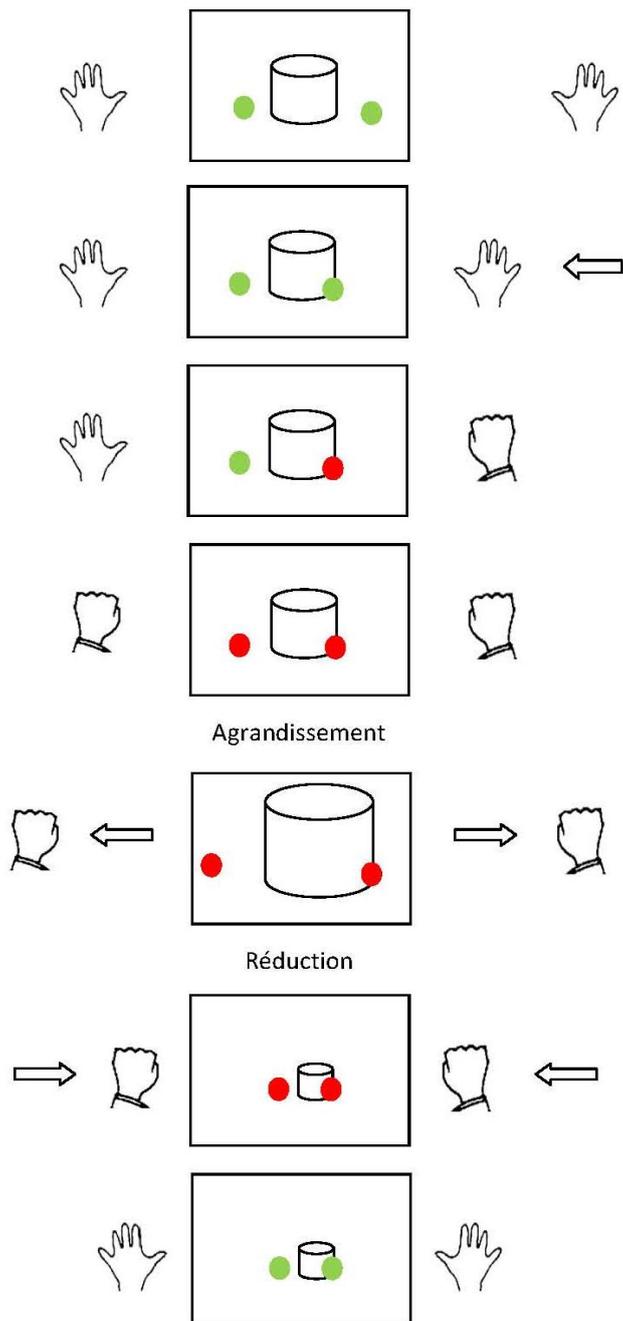
Positionnez votre main dominante fermée dessus. Fermez votre main secondaire pour autoriser le déplacement. Déplacez votre main dominante à l'endroit voulu tout en maintenant vos 2 mains fermées. Ouvrez votre main secondaire en premier lorsque l'objet est bien positionné. Une fois que l'objet est à la bonne place, vous pouvez ouvrir votre main dominante. Si votre main secondaire est ouverte, le déplacement n'est pas autorisé.



### Etude 3 – Fiche explicative sur la modalité 1 du redimensionnement

#### Redimensionner un objet 1 :

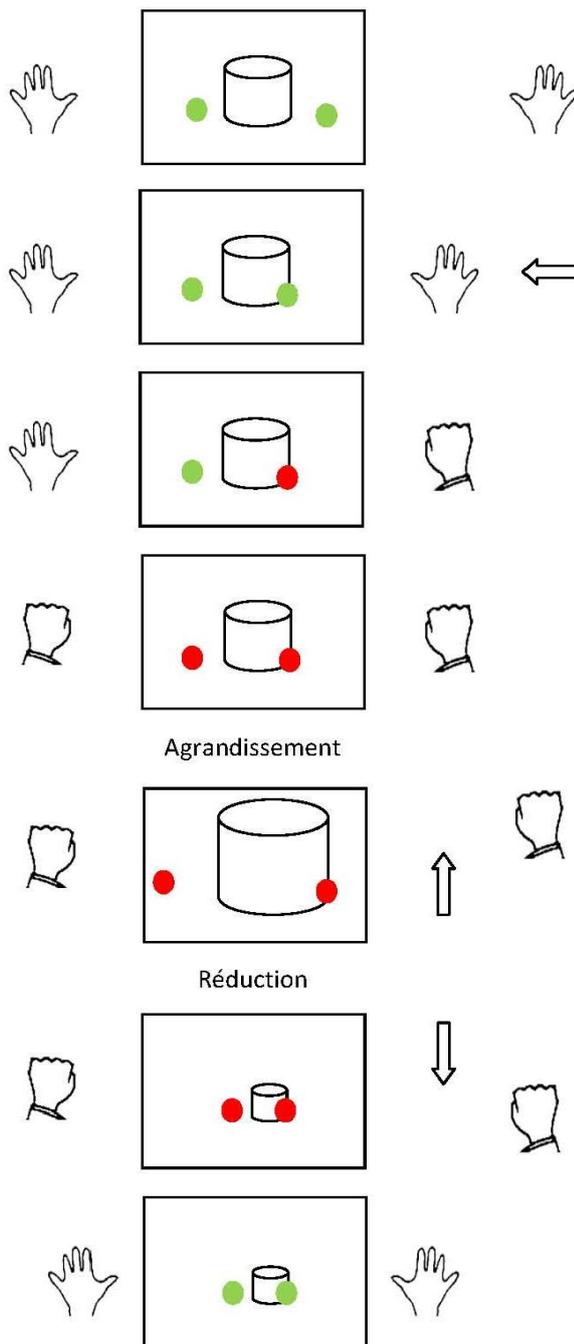
Positionnez votre main dominante fermée sur l'objet et fermez votre deuxième main pour autoriser le redimensionnement. Ecartez ou rapprochez vos mains pour redimensionner l'objet. Ecarter les mains permet d'agrandir l'objet, les rapprocher permet de le réduire. Une fois le redimensionnement effectué, ouvrez vos mains.



## Etude 3 – Fiche explicative sur la modalité 2 du redimensionnement

### Redimensionner un objet 2:

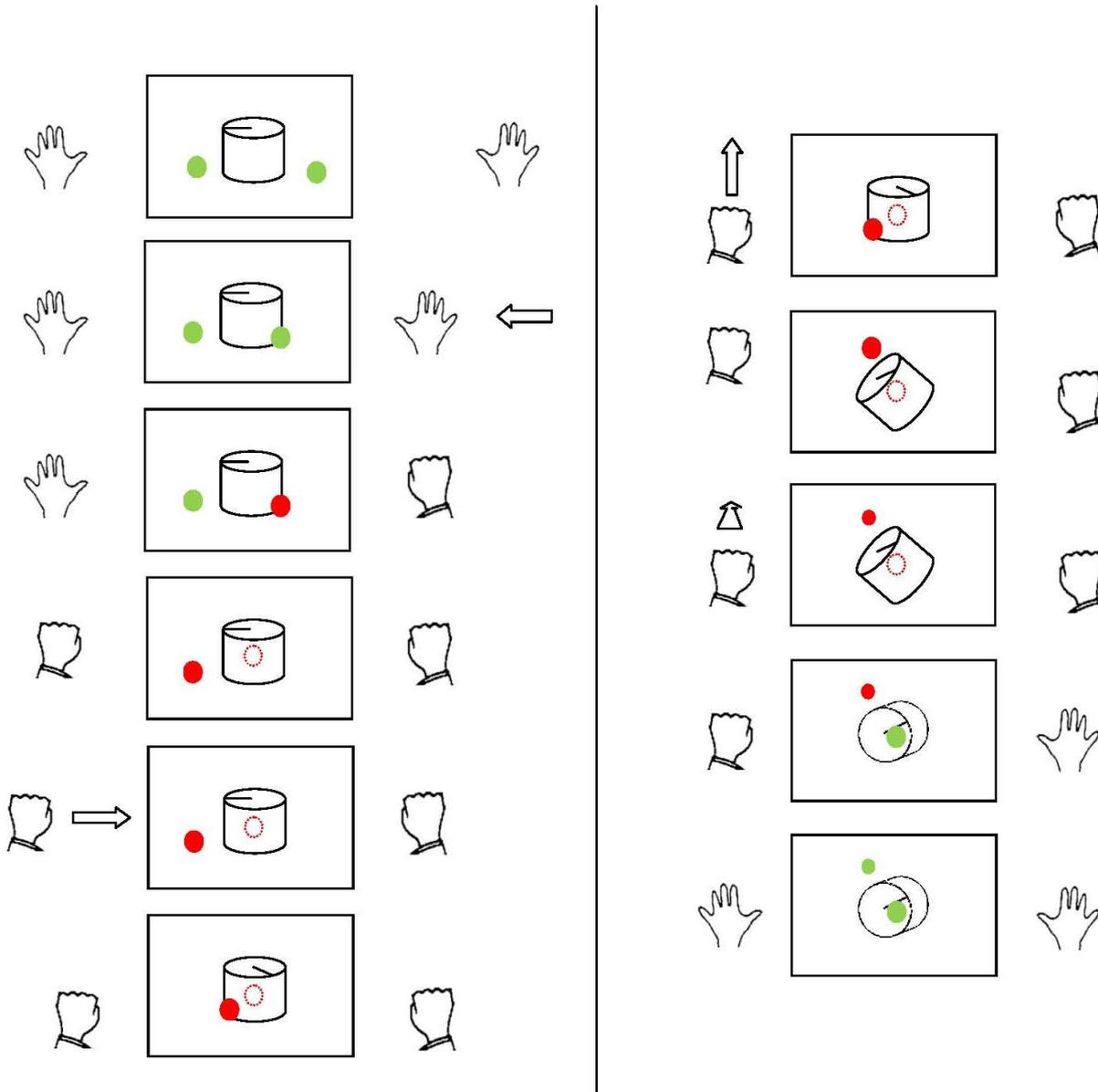
Positionnez votre main dominante fermée sur l'objet et fermez votre deuxième main pour autoriser le redimensionnement. Montez ou descendez vos mains pour redimensionner l'objet. Monter votre main dominante permet d'agrandir l'objet, la descendre permet de le réduire. Une fois le redimensionnement effectué, ouvrez en premier votre main secondaire. Si vous souhaitez réajuster sa taille, refermez votre main secondaire. Lorsque l'objet est à la taille désirée, vous pouvez ouvrir votre main dominante.



### Etude 3 – Fiche explicative sur la modalité 1 de la rotation

#### Rotation d'un objet 1 :

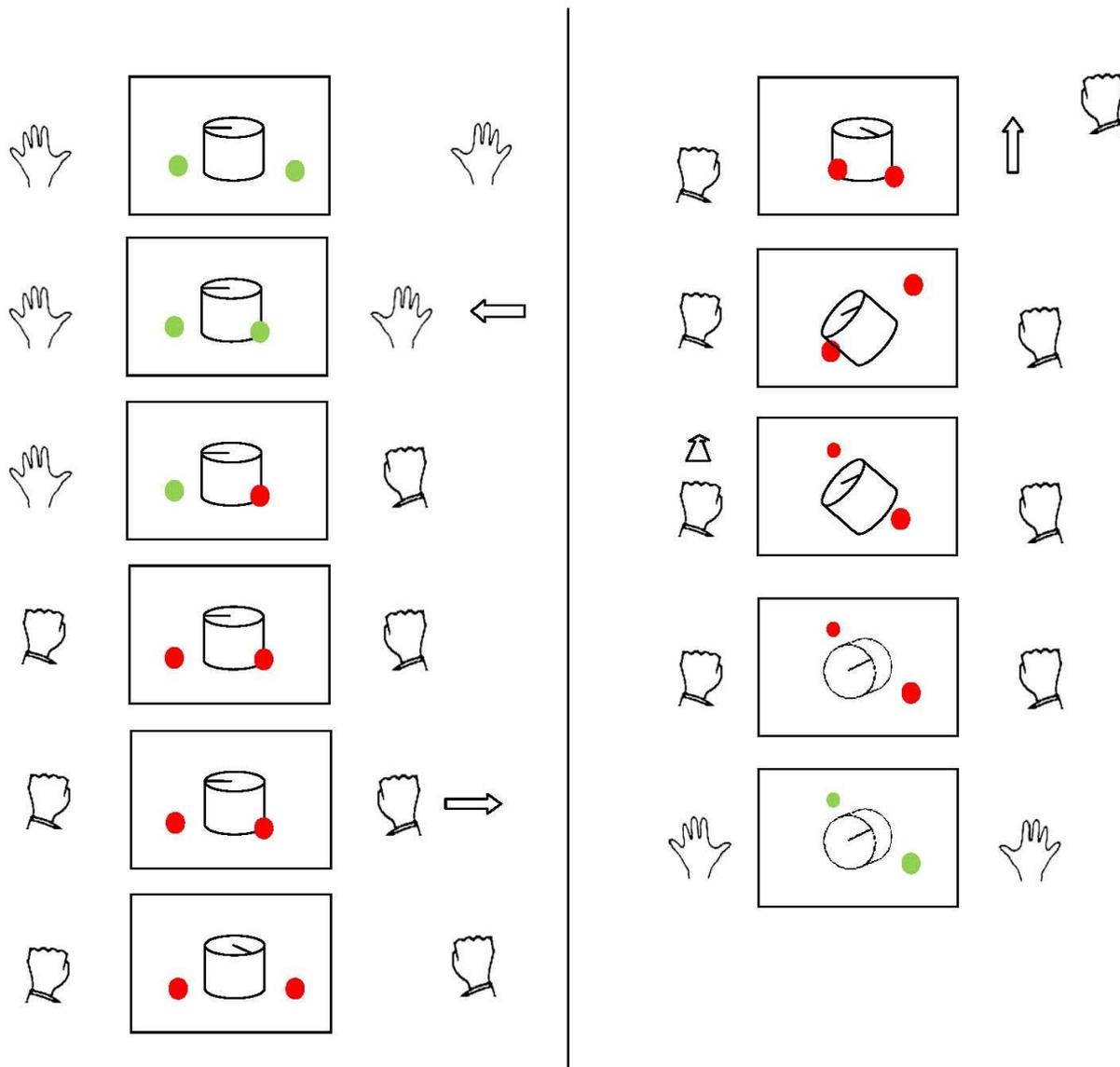
Positionnez votre main dominante fermée sur l'objet, elle disparaîtra et ira se positionner automatiquement au centre de ce dernier. Fermez votre main secondaire pour autoriser la rotation. Déplacer votre main secondaire selon l'axe des X permettra à l'objet de tourner dans cette même direction et ce de façon incrémentale. Il en est de même pour les axes Y et Z. Une fois la rotation effectuée, ouvrez votre main dominante d'abord puis la main secondaire. Plus votre main secondaire est déplacée par rapport à la position où vous l'avez fermée et plus la rotation se fera vite. Un déplacement de faible amplitude, permettra une rotation lente.



### Etude 3 – Fiche explicative sur la modalité 2 de la rotation

#### Rotation d'un objet 2:

Positionnez votre main dominante sur l'objet, fermez la dessus pour autoriser les rotations en X et Y, puis fermez votre main secondaire pour autoriser la rotation en Z. Déplacer votre main dominante selon les axes X et Y pour tourner l'objet selon ces mêmes axes. Déplacer votre main gauche d'avant en arrière pour tourner l'objet selon l'axe Z. Lorsque la rotation est effectuée, ouvrez les deux mains en même temps.



## Etude 3 – Questionnaire d'identification

### PRESENTATION DE L'EXPERIMENTATION

Nous vous invitons à participer à l'évaluation de plusieurs modalités d'interaction permettant la manipulation d'objets 3D en environnement virtuel (E. V.) et ce en utilisant un système de capture de mouvements sans marqueurs.

L'expérimentation vise à effectuer différentes interactions en E. V. qui sont le déplacement, la rotation et le redimensionnement d'objets 3D. Toutes ces actions se regroupent sous le terme de manipulation d'objets et constituent une tâche très importante en réalité virtuelle. Pour cela, vous devrez utiliser vos 2 mains dans l'espace entre vous et l'écran et ouvrir ou fermer les mains pour effectuer les actions demandées. Le but est de comparer deux modalités d'interactions pour chacune des tâches précédemment citées.

Dans un objectif scientifique, la trace de vos interactions ainsi que votre activité gestuelle (déplacement des mains, ouverture, fermeture, etc.) seront enregistrées.

Cette expérimentation préservera votre anonymat conformément à la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Les informations sollicitées dans ce questionnaire nous serviront à préciser votre profil en tant qu'utilisateur de logiciels grand public. Elles resteront confidentielles.

Nous vous remercions d'avoir accepté de participer à cette expérimentation.

#### INFORMATIONS PASSATION

Numéro de passage

Date

#### INFORMATIONS PERSONNELLES

Nom

Tél

Prénom

E-mail

Age

Latéralisation  
 Gaucher  
 Droitier  
 Ambidextre

#### CONNAISSANCES GENERALES

Dernier diplôme obtenu (Niveau - spécialité; ex : master - MNRV) :

Diplôme en cours de préparation (Niveau - spécialité; ex : master - MNRV) :

## COMPETENCES INFORMATIQUES

### Avez-vous des connaissances en réalité virtuelle ?

Pas du tout      Excellentes connaissances

### Quels dispositifs de réalité virtuelle avez-vous déjà utilisés et à quelle fréquence ?

	Jamais	Peu	Parfois	Souvent	Très souvent
Souris 3D	<input type="radio"/>				
Joystick	<input type="radio"/>				
Manette de jeu	<input type="radio"/>				
Bras à retour d'effort	<input type="radio"/>				
Gant de données	<input type="radio"/>				

Autres, précisez :

### Quelles activités informatiques utilisant la 3D pratiquez vous?

	Jamais	Peu	Parfois	Souvent	Très souvent
Jeux vidéo à moteur 3D	<input type="radio"/>				
Logiciels de CAO (Autocad Architectural Desktop, sketchup)	<input type="radio"/>				
Logiciels de dessin, modeling (3DS max, Catia, Maya)	<input type="radio"/>				
Logiciels de création de scènes interactives et temps réel (Virtools, Unity, OpenScene Graph, Ogre)	<input type="radio"/>				

Autres, précisez :

### Avez vous déjà été confronté à des interactions bi-manuelles?

Oui  Non

Si oui, avec quel(s) périphérique(s) :

Et pour quelle(s) application(s) :

## QUESTIONNAIRE POST-PASSATION

### INFORMATIONS PASSATION

Numéro de passage

Date

### IMPRESSIONS GENERALES

⇒ Par quelles modalités avez-vous commencé à chaque fois ?

Modalités 1     Modalités 2

⇒ Etes vous resté concentré tout au long de l'expérimentation ?

Pas du tout                     Tout le temps

⇒ La vue en perspective vous a-t-elle gênée ?

Oui     Non

Si oui, pourquoi ?

⇒ L'expérimentation vous a semblé :

Très longue                        Très courte

Très difficile                        Très facile

Très ennuyeuse                        Très attractive

## QUESTIONS SUR LES DEPLACEMENTS

⇒ Comment évaluez-vous le **degrés de difficulté** d'assimilation des modalités d'interaction du :

	Très difficile	Difficile	Normal	Facile	Très facile
Déplacement 1 :	<input type="radio"/>				
Déplacement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **simplicité** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacement 1 :	<input type="radio"/>				
Déplacement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) l'**intuitivité** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacement 1 :	<input type="radio"/>				
Déplacement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **précision** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Déplacement 1 :	<input type="radio"/>				
Déplacement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très grande à très faible) la **fatigue engendrée** par le :

	Très grande	Grande	Moyenne	Faible	Très faible
Déplacement 1 :	<input type="radio"/>				
Déplacement 2 :	<input type="radio"/>				

⇒ Quelle modalité d'interaction avez vous préféré ?

Déplacement 1    Déplacement 2

⇒ Pourquoi ?

⇒ Quels sont, selon vous, les points positifs du déplacement 1 ?

⇒ Quels sont, selon vous, les points négatifs du déplacement 1 ?

⇒ Quels sont, selon vous, les points positifs du déplacement 2 ?

⇒ Quels sont, selon vous, les points négatifs du déplacement 2 ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction des déplacements ?

## QUESTIONS SUR LES REDIMENSIONNEMENTS

⇒ Comment évaluez-vous le **degrés de difficulté** d'assimilation des modalités d'interaction du :

	Très difficile	Difficile	Normal	Facile	Très facile
Redimensionnement 1 :	<input type="radio"/>				
Redimensionnement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **simplicité** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Redimensionnement 1 :	<input type="radio"/>				
Redimensionnement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) l'**intuitivité** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Redimensionnement 1 :	<input type="radio"/>				
Redimensionnement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **précision** du :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Redimensionnement 1 :	<input type="radio"/>				
Redimensionnement 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très grande à très faible) la **fatigue engendrée** par le :

	Très grande	Grande	Moyenne	Faible	Très faible
Redimensionnement 1 :	<input type="radio"/>				
Redimensionnement 2 :	<input type="radio"/>				

⇒ Quelle modalité d'interaction avez vous préféré ?

Redimensionnement1       Redimensionnement2

⇒ Pourquoi ?

⇒ Quels sont, selon vous, les points positifs du redimensionnement 1 ?

⇒ Quels sont, selon vous, les points négatifs du redimensionnement 1 ?

⇒ Quels sont, selon vous, les points positifs du redimensionnement 2 ?

⇒ Quels sont, selon vous, les points négatifs du redimensionnement 2 ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction du redimensionnement?

## QUESTIONS SUR LES ROTATIONS

⇒ Comment évaluez-vous le **degrés de difficulté** d'assimilation des modalités d'interaction de la :

	Très difficile	Difficile	Normal	Facile	Très facile
Rotation 1 :	<input type="radio"/>				
Rotation 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **simplicité** de la :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Rotation 1 :	<input type="radio"/>				
Rotation 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) l'**intuitivité** de la :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Rotation 1 :	<input type="radio"/>				
Rotation 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très mauvaise à très bonne) la **précision** de la :

	Très mauvaise	Mauvaise	Moyenne	Bonne	Très bonne
Rotation 1 :	<input type="radio"/>				
Rotation 2 :	<input type="radio"/>				

⇒ Comment évaluez-vous (de très grande à très faible) la **fatigue engendrée** par la :

	Très grande	Grande	Moyenne	Faible	Très faible
Rotation 1 :	<input type="radio"/>				
Rotation 2 :	<input type="radio"/>				

⇒ Quelle modalité d'interaction avez vous préféré ?

Rotation1                       Rotation2

⇒ Pourquoi ?

⇒ Quels sont, selon vous, les points positifs de la rotation 1 ?

⇒ Quels sont, selon vous, les points négatifs de la rotation 1 ?

⇒ Quels sont, selon vous, les points positifs de la rotation 2 ?

⇒ Quels sont, selon vous, les points négatifs de la rotation 2 ?

⇒ Quelles sont vos suggestions pour améliorer les modalités d'interaction de la rotation?

Envoyer



## Bibliographie

---



- Argyros, A. A. & Lourakis, M. I. A. (2006). *Binocular hand-tracker and reconstruction based on 2d-shape matching*. Paper presented at the 18th International Conference on Pattern Recognition (ICPR '06).
- Arnaldi, B., Tisseau, J., Berthoz, A., Burkhardt, J.-M., Coquillart, S. & Guitton, P. (2003). *Traité de la Réalité Virtuelle* (3 ed. Vol. Volume 2, L'interfaçage, l'immersion et l'interaction): Les Presses de l'École des Mines de Paris.
- Bach, C. & Scapin, D. (2003). *Ergonomic criteria adapted to human virtual environment interaction*. Paper presented at the 15th French-speaking conference on human-computer interaction, Caen, pp. 24, 31.
- Bailly, G., Demeure, A., Lecolinet, E. & Nigay, L. (2008). *MultiTouch menu (MTM)*. Paper presented at the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, Metz, France, pp. 165-168.
- Beaudouin-Lafon, M. (2004). *Designing interaction, not interfaces*. Paper presented at the Working conference on Advanced Visual Interfaces, Gallipoli, Italy, pp. 15-22.
- Berard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J. R. & Cooperstock, J. R. (2009). *Did Minority Report get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task*. Paper presented at the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II (INTERACT '09), pp. 400-414.
- Bezdicsek, M. & Caldwell, D. G. (2006). *Portable absolute position tracking system for human hand fingertips*. Paper presented at the Virtual Concept 2006, Cancùn, Mexico,.
- Boulic, R., Fua, P., Herda, L., Silaghi, M., Monzani, J.-S., Nedel, L., et al. (1998). *An Anatomic Human Body for Motion Capture*. Paper presented at the EMMSEC'98.
- Bouzit, M., Popescu, G., Burdea, G. & Boian, R. (2002). *The Rutgers Master II-ND Force Feedback Glove*. Paper presented at the Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.
- Bowman, D. A. (1999). *Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application*. Georgia Institute of Technology.
- Bowman, D. A., Kruijff, E., LaViola, J. & Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice*: Addison Wesley Longman Publishing Co., Inc.
- Bowman, D. A. & McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer*, 40(7), 36-43.
- Bray, J. (2006). *Markerless Based Human Motion Capture : A Survey*. Paper presented at the Systems Engineering, pp. 1-44.
- Brethes, L., Menezes, P., Lerasle, F. & Briot, M. (2004). *Segmentation couleur et condensation pour le suivi et la reconnaissance de gestes humains*. Paper presented at the 14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes & Intelligence Artificielle (RFIA'2004), Toulouse, France, pp. 967-975
- Breuer, P., Eckes, C. & Müller, S. (2007). *Hand gesture recognition with a novel IR time-of-flight range camera: a pilot study*. Paper presented at the Proceedings of the 3rd international conference on Computer vision/computer graphics collaboration techniques, Rocquencourt, France.

- Buchholz, B. & Armstrong, T. J. (1992). A kinematic model of the human hand to evaluate its prehensile capabilities. *Journal of biomechanics*, 25(n°2), 149-162.
- Burkhardt, J.-M. (2006). Ergonomie, Facteurs Humains et Réalité Virtuelle. In G. M. P. Fuchs, A. Berthoz & J.-L. Vercher (Ed.), *Le traité de la Réalité Virtuelle* (Vol. 1, pp. 117-150).
- Burkhardt, J.-M. & Fuchs, P. (2006). Introduction à la réalité virtuelle. In *Traité de la réalité virtuelle* (Vol. 2, pp. 5-9): Les Presses de l'École des Mines de Paris.
- Burkhardt, J.-M. & Sperandio, J.-C. (2004). Ergonomie et conception Informatique. In P. Falzoon (Ed.), *Ergonomie* (pp. 437-450). Paris: PUF.
- Büttgen, B., Oggier, T., Lehmann, M. & Kaufmann, R. (2005). *CCD/CMOS lock-in pixel for range imaging: Challenges, limitations and state-of-the-art*. Paper presented at the 1st Range Imaging Research Day, Zurich, Switzerland, pp. 21-32.
- Buttolo, P., Kung, D. & Hannaford, B. (1995). *Manipulation in Real, Virtual and Remote Environments*. Paper presented at the EEE Conference on Systems, Man and Cybernetics pp. 4656 - 4661.
- Callahan, J., Hopkins, D., Weiser, M. & Shneiderman, B. (1988). *An empirical comparison of pie vs. linear menus*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, Washington, D.C., United States.
- Cameron, E. S., Szumski, R. P. & West, J. K. (1991).
- Campos, T. E. d. & Murray, D. W. (2006). *Regression-based Hand Pose Estimation from Multiple Cameras*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 782-789.
- Casiez, G. (2004). *Contribution à l'étude des interfaces haptiques - Le DigiHaptic : un périphérique haptique de bureau à degrés de liberté séparés*. PhD Thesis, University of Lille, Lille.
- Causo, A., Takemura, K., Takamatsu, J., Ogasawara, T., Ueda, E. & Matsumoto, Y. (2010). *Predictive Tracking in Vision-based Hand Pose Estimation Using Unscented Kalman Filter and Multi-viewpoint Cameras*: Daisuke Chugo.
- Chalfoun, J., Massemejean, E., Mouhamed, B., Ouezdou, B. F. & Younès, R. (2003). *Anatomie de la main et relation tendon couple articulaire*. Paper presented at the Congrès International de Modélisation Numérique Appliquée, CIMNA,, Beyrouth, Liban
- Chen, M., Mountford, S. J. & Sellen, A. (1988). A study in interactive 3-D rotation using 2-D control devices. *SIGGRAPH Comput. Graph.*, 22(4), 121-129.
- Cheung, K. M., Kanade, T., Bouget, J.-Y. & Holler, M. (2000). *A real Time System for Robust 3D Voxel Reconstruction of Human Motions*. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00).
- Christmann, O., Carbonell, N. & Richir, S. (2010). Visual search in dynamic 3D visualisations of unstructured picture collections. *Interacting with Computers*, 22(5), 399-416.
- Chua, C.-S., Guan, H. & Ho, Y.-K. (2002). Model-based 3D hand posture estimation from a single 2D image. *Image and Vision Computing*, 20(3), 191-202.
- Collet, C. (1999). *Capture et suivi du regard par un système de vision dans le contexte de la communication homme-machine*. École Normale Supérieure de Cachan.

- Connelly, L., Stoykov, M. E., Yicheng, J., Toro, M. L., Kenyon, R. V. & Kamper, D. G. (2009). *Use of a pneumatic glove for hand rehabilitation following stroke*. Paper presented at the Engineering in Medicine and Biology Society (EMBC 2009), Minneapolis, MN.
- Connelly, L., Yicheng, J., L.Toro, M., Stoykov, M. E., Kenyon, R. V. & Kamper, D. G. (2010). A Pneumatic Glove and Immersive Virtual Reality Environment for Hand Rehabilitative Training After Stroke. *IEEE Engineering in Medicine and Biology Society*, 18(5), 551-559.
- Coquillart, S., Arnaldi, B., Berthoz, A., Burkhardt, J.-M., Fuchs, P. & Guitton, P. (2006). *Le traité de la réalité virtuelle : interfaçage, immersion et interaction en environnement virtuel*: Mines de Paris.
- Coquillart, S., Fuchs, P., Grosjean, J., Hachet, M., Bechmann, D. & Stenberger, L. (2003). Les techniques d'interaction pour les primitives comportementales virtuelles. In *Traité de la réalité virtuelle : Volume 2, L'interfaçage, l'immersion et l'interaction* (3 ed., Vol. 2, pp. 332): Les Presses de l'École des Mines de Paris.
- CSEM. (2009). *Time of Flight Camera Technology*
- Dankers, A., Barnes, N. & Zelinsky, A. (2007). MAP ZDF segmentation and tracking using active stereo vision: Hand tracking case study. *Comput. Vis. Image Underst.*, 108(1-2), 74-86.
- Delamarre, Q. & Faugeras, O. (1998). *Finding Pose of Hand in Video Images: A Stereo-Based Approach*. Paper presented at the third International Conference on Automatic Face and Gesture Recognition, pp. 585-590.
- Delamarre, Q. & Faugeras, O. (1999). *3D Articulated Models and Multi-View Tracking with Silhouettes*. Paper presented at the International Conference on Computer Vision.
- Dewaele, G., Devernay, F. & Horaud, R. (2004). *Hand Motion from 3D Point Trajectories and a Smooth Surface Model*. Paper presented at the European Conference on Computer Vision, pp. 495-507.
- Dorner, B. (1994). *Chasing the Colour Glove: Visual Hand Tracking*. Master's thesis, Simon Fraser University.
- Dvorak, J. C. (1984, february 19). *The San Francisco Examiner*,
- Elmezain, M., Al-Hamadi, A., Appenrodt, J. & Michaelis, B. (2008). *A Hidden Markov Model-based continuous gesture recognition system for hand motion trajectory*. Paper presented at the 19th International Conference on Pattern Recognition, 2008. ICPR 2008, Tampa, FL pp. 1-4.
- Fakespace. (1997). PinchTM Glove System Installation Guide and User Handbook.
- Fischer, H. C., Stubblefield, K., Kline, T., Luo, X., Kenyon, R. V. & Kamper, D. G. (2007). Hand rehabilitation following stroke: a pilot study of assisted finger extension training in a virtual environment. *Topics in Stroke Rehabilitation*, 14(1), 1-12.
- Flückiger, L. (1998). *Interface pour le pilotage et l'analyse des robots basée sur un générateur de cinématiques*. Docteur ès sciences techniques, Ecole polytechnique fédérale de Lausanne, Lausanne.
- Freedman, B., Shpunt, A., Machline, M. & Arieli, Y. (2010). Depth Mapping Using Projected Patterns. United States Patent Application Publication.

- Fuchs, P. & Mathieu, H. (2003). Les interfaces spécifiques de la localisation corporelle - Introduction. In *Traité de la réalité virtuelle : Volume 2, L'interfaçage, l'immersion et l'interaction* (3 ed., Vol. 2, pp. 93-94): Les Presses de l'École des Mines de Paris.
- Fuchs, P. & Moreau, G. (2003). *Traité de la Réalité Virtuelle* (2 ed. Vol. 2): Les Presses de l'Ecole des Mines de Paris.
- Fuchs, P. & Richir, S. (2006). La méthode I2I : Interaction et Immersion pour l'Innovation. *Encyclopédie des Techniques de l'Ingénieur, Télécoms*(TE5910).
- Fujiki, R., Yoshimoto, H., Arita, D. & Taniguchi, R.-i. (2005). *Real-time Model-based Hand Shape Estimation with Stereo Vision*. Paper presented at the Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp. 225-230.
- Gagnon, P. (2006). *Intervalles de confiance pour une différence de deux proportions*. Master thesis, Université de Laval, Laval, Canada.
- Geebelen, G., Maesen, S., Cuypers, T. & Bekaert, P. (2010). *Real-time hand tracking with a colored glove*. Paper presented at the 3D Stereo Media.
- Gentner, R. & Classen, J. (2009). Development and evaluation of a low-cost sensor glove for assessment of human finger movements in neurophysiological settings. *Journal of Neuroscience Methods*, 178(1), 138-147.
- Goebel, W. & Palmer, C. (2009). *Finger motion in piano performance: Touch and tempo*. Paper presented at the International Symposium on Performance Science.
- Gokturk, S. B., Hakan, Y. & Cyrus, B. (2004). *A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions*. Paper presented at the Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04).
- Gorce, M. d. L. & Paragio, N. (2005). *Hand Pose Estimation with Constrained Multi-hypotheses Gradient-Descent*: CERTIS, ENPC
- Gorce, M. d. L., Paragios, N. & Fleet, D. J. (2008). *Model-based hand tracking with texture, shading and self-occlusions*. Paper presented at the Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK
- Gosselin, F., Andriot, C. & Fuchs, P. (2003). Les dispositifs matériels des interfaces à retour d'effort. In *Traité de la réalité virtuelle : Volume 2, L'interfaçage, l'immersion et l'interaction* (3 ed.): Les Presses de l'Ecole des Mines de Paris.
- Green, W. & Jordan, P. (2002). *Pleasure with products: beyond usability* (W. Green P. Jordan ed.). London: Taylor and Francis.
- Guan, C., Hassebrook, L. G., Lau, D. L. & Yalla, V. (2007). Near-Infrared Composite Pattern Projection for Continuous Motion Hand-Computer Interaction. *Real-Time Imaging in Journal of Visual Communication and Image Representation*, 18(2), 141-150.
- Guan, H. & Turk, M. (2005). 3D Hand Pose Reconstruction with ISOSOM. In S. B. Heidelberg (Ed.), *Advances In Visual Computing* (Vol. 3804/2005, pp. 630-635).

- Gumpp, T., Azad, P., Welke, K., Oztop, E., Dillmann, R. & Cheng, G. (2006). *Unconstrained Real-time Markerless Hand Tracking for Humanoid Interaction*. Paper presented at the 6th IEEE-RAS International Conference on Humanoid Robots, Genova pp. 88-93.
- Häger-Ross, C. & Schieber, M. H. (2000). Quantifying the Independence of Human Finger Movements : Comparisons of Digits, Hands, and Movement Frequencies. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(22), 8542-8550.
- Hand, C. (1997). A Survey of 3D Interaction Techniques. *Computer Graphics Forum*, 16(5), 269-281.
- Harrouet, F., Cazeaux, E. & Jourdan, T. (2006). Animation d'humains par cinématique inverse. In *Le traité de la réalité virtuelle - Volume 3 : Les outils et les modèles informatiques des environnements virtuels* (pp. 278): Presses de l'Ecole des Mines de Paris.
- Hartling, P. L., Bierbaum, A. D. & Cruz-Neira, C. (2002). *Tweek: Merging 2D and 3D Interaction in Immersive Environments*. Paper presented at the 6th World Multiconference on Systemics Cybernetics and Informatics.
- Hasegawa, S., Toshiaki, I., Hashimoto, N., Salvati, M., Mitake, H., Koike, Y., et al. (2006). Human-scale haptic interaction with a reactive virtual human in a real-time physics simulator. *Comput. Entertain.*, 4(3), 9.
- Hassanpour, R., Shahbahrami, A., Stephan & Wong. (2008). Adaptive gaussian mixture model for skin color segmentation. *World Academy of Science, Engineering and Technology*, 41.
- Hayward, V. & Astley, O. R. (1996). *Performance Measures for Haptic Interfaces*. Paper presented at the 7th International Symposium on Robotics Research, pp. 195-207.
- Heap, T. & Hogg, D. (1996). *Towards 3D hand tracking using a deformable model*. Paper presented at the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96).
- Herda, L., Fua, P., Plänkers, R., Boulic, R. & Thalmann, D. (2000). *Skeleton-Based Motion Capture for Robust Reconstruction of Human Motion*. Paper presented at the Computer Animation, Philadelphia, PA, USA.
- Herda, L., Fua, P., Plänkers, R., Boulic, R. & Thalmann, D. (2001). Using skeleton-based tracking to increase the reliability of optical motion capture. *Human Movement Science Journal*, 20(3), 313-341.
- Hernandez-Rebollar, J. L., Kyriakopoulos, N. & Lindeman, R. W. (2002). *The AcceleGlove: a whole-hand input device for virtual reality*. Paper presented at the ACM SIGGRAPH 2002 conference abstracts and applications, San Antonio, Texas.
- Hill, J. W. (1977). *Two Measures of Performance in a Peg-inhold Manipulation Task with Force Feedback*. Paper presented at the 13th Annual Conference on Manual Control, pp. 301-309.
- Hirata, Y. & Sato, M. (1992). *D-dimensional Interface Device for Virtual Work Space*. Paper presented at the Virtual Reality Annual International Symposium, Seattle, WA, USA, pp. 331 - 335
- Hollerbach, J. M., Hunter, I. W. & Ballantyne, J. (1992). A comparative analysis of actuator technologies for robotics. In *The robotics review 2* (pp. 299-342): MIT Press.
- Holte, M. B. & Moeslund, T. B. (2007). Gesture Recognition using a Range Camera. *Gesture*.

- Ilayperuma, I., Nanayakkara, G. & Palahepitiya, N. (2009). Prediction of personal stature based on the hand length. *2009*, 14(1), 15-18.
- Immersion. (2011). from [http://www.immersion.fr/index.php?page=shop.product\\_details&flypage=flypage.tpl&product\\_id=316&category\\_id=180&option=com\\_virtuemart&Itemid=5&lang=fr](http://www.immersion.fr/index.php?page=shop.product_details&flypage=flypage.tpl&product_id=316&category_id=180&option=com_virtuemart&Itemid=5&lang=fr)
- Jia, Y., Toro, M. L., Luo, X., Lau, S., Kenyon, R. & Kamper, D. (2007). *Integration of Virtual Reality and an Assistive Device for Hand Rehabilitation Following Stroke*. Paper presented at the IEEE/ICME International Conference on Complex Medical Engineering (CME 2007), Beijing pp. 1331 - 1336
- Jones, M. J. & Rehg, J. M. (2002). Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1), 81-96.
- Kadous, W. (1995). *GRASP: Recognition of Australian Sign Language Using Instrumented Gloves*. Bachelor's thesis, University of New South Wales.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*, 82, 35-45.
- Kamina, P. & Francke, J.-P. (1994). *Arthrologie des Membres*. In *Introduction à la clinique*: Editions Maloine.
- Kang, H. J. & Jung, M. Y. (2008). Human-Computer-Interface (HCI) using Hand Motion, *Final Project Report Computer Vision*.
- Kapandji, A.-I. (1980). *Physiologie articulaire, Tome 1 : Membre supérieur* (5 ed.): Maloine S.A. Editor.
- Kato, M. & Xu, G. (2006). Occlusion-Free Hand Motion Tracking by Multiple Cameras and Particle Filtering with Prediction. *IJCSNS International Journal of Computer Science and Network*, 6(10).
- Kato, M., Xu, G. & Chen, Y.-W. (2007). *Articulated Hand Tracking by ICA-based Hand Model and Multiple Cameras, Scene Reconstruction Pose Estimation and Tracking*: InTech.
- Kessler, G. D., Hodges, L. F. & Walker, N. (1995). Evaluation of the CyberGlove as a whole-hand input device. *ACM Trans. Comput.-Hum. Interact.*, 2(4), 263-283.
- Kolb, A., Barth, E., Koch, R. & Larsen, R. (2009). *Time-of-flight sensors in computer graphics*. Paper presented at the Eurographics.
- Kollorz, E., Penne, J., Hornegger, J. & Barke, A. (2008). Gesture recognition with a Time of Flight camera. *International Journal of Intelligent Systems Technologies and Applications*, 5(3/4), 334-343.
- Kramer, J. (1991). Communication System for Deaf, Deaf-blind, and Non-vocal Individuals Using Instrumented Gloves. US Patent No. 5,047,952, 1991.
- Kramer, J. & Leifer, L. (1987). *The Talking Glove: An Expressive and Receptive 'Verbal' Communication Aid for the Deaf, Deaf-Blind, and Nonvocal*. Paper presented at the Third Annual Conference on computer Technology/Special Education/ Rehabilitation, pp. 335-340.
- Krout, M. H. (1935). Autistic gestures: an experimental study in symbolic movement. *Psychological Monographs*, 46(4), 119-120.

- Kuch, J. J. & Huang, T. S. (1995). *Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration*. Paper presented at the Fifth International Conference on Computer Vision.
- Kurtenbach, G. & Buxton, W. (1991). *Issues in combining marking and direct manipulation techniques*. Paper presented at the 4th annual ACM symposium on User interface software and technology, Hilton Head, South Carolina, United States, pp. 137-144.
- Lahamy, H. & Lichti, D. (2010). Real-time hand gesture recognition using range cameras. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37, 54-59.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafos, J., Green, V., Oliva, D., et al. (2011). Microswitch and keyboard-emulator technology to facilitate the writing performance of persons with extensive motor disabilities. *Research in Developmental Disabilities*, 32(2), 576-582.
- Lange, R. (2000). *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. University of Siegen.
- Lange, R. & Seitz, P. (2000). Seeing distance – A fast time-of-flight 3D camera. *Sensor Review*, 20(3), 212-217.
- Laurentini, A. (1994). The Visual Hull Concept for Silhouette-Based Image Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2), 150-162.
- LaViola, J. J. (1999). *Whole-Hand and Speech Input in Virtual Environments*: Brown University
- Lempereur, M. (2008). *Simulation du mouvement d'entrée dans un véhicule automobile*. Doctorat, Université de Valenciennes et du Hainaut-Cambrésis.
- Li, Z. & Jarvis, R. (2009). *Real time Hand Gesture Recognition using a range camera*. Paper presented at the Australasian Conference on Robotics and Automation 2009 (ACRA09), Sydney, Australia, pp. 21-27.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Lin, J., Wu, Y. & Huang, T. S. (2000). *Modeling the constraints of human hand motion*. Paper presented at the IEEE Workshop on Human Motion (HUMO'00), pp. 121-126.
- Lobo, J., Trindade, P. & Dias, J. (2011). *Observing hand grasp type and contact points using hand distributed accelerometers and instrumented objects*. Paper presented at the IEEE/ICRA 2011: Workshop on Autonomous Grasping, Shanghai, China.
- Loup-Escande, E. (2010). *Vers une conception centrée sur l'utilité : une analyse de la co-construction participative et continue des besoins dans le contexte des technologies émergentes*. PhD thesis, Angers.
- Loup-Escande, E., Burkhardt, J.-M. & Richir, S. (2011). Anticiper et évaluer l'utilité dans la conception ergonomique des technologies émergentes : une revue. *Le Travail Humain*.
- Lu, S., Huang, G., Samaras, D. & Metax, D. (2002). *Model-based Integration of Visual Cues for Hand Tracking*. Paper presented at the Workshop on Motion and Video Computing, p. 118.
- Lu, S., Metaxas, D., Samaras, D. & Oliensis, J. (2003). *Using multiple cues for hand tracking and model refinement* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 443-450.

- Ma, Y., Mao, Z.-H., Jia, W., Li, C., Yang, J. & Sun, M. (2011). Magnetic Hand Tracking for Human-Computer Interface. *Transactions on Magnetics*, 47(5), 970-973.
- MacKenzie, C. L. & Iberall, T. (1994). The Grasping hand. *Advances in Psychology*, 104, 18-38.
- Makower, J., Parnianpour, M. & Nordin., M. (1990). *The Validity Assessment of the Dextrous Hand Master: A Linkage System for the Measurement of the Joints of the Hand*. Paper presented at the the First World Congress of Biomechanics, La Jolla, California, p. 338.
- Malassiotis, S. & Strintzis, M. G. (2008). Real-time hand posture recognition using range data. *Image and Vision Computing*, 26(7), 1027-1037.
- Malik, S. & Laszlo, J. (2004). *Visual touchpad: a two-handed gestural input device*. Paper presented at the Proceedings of the 6th international conference on Multimodal interfaces, State College, PA, USA, pp. 289-296.
- Marieb, E. (1993). *Anatomie et physiologie humaines*: Editions De Boeck Université.
- Massimino, M., J. & Sheridan, T., B. (1994). Teleoperator performance with varying force and visual feedback. *Human Factors*, 36(1), 145-157.
- May, S., Pervoelz, K. & Surmann, H. (2007). 3D Cameras: 3D Computer Vision of Wide Scope. In G. Obinata & A. Dutta (Eds.), *Vision Systems: Application*: InTech.
- Menache, A. (1999). *Understanding Motion Capture for Computer Animation and Video Games*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Meseure, P. & Kheddar, A. (2006). Les outils et les modèles informatiques des environnements virtuels In *Le traité de la réalité virtuelle* (pp. 141): Les Presses de l'École des Mines de Paris.
- Mestre, D., Fuchs, P., Berthoz, A. & Vercher, J.-L. (2006). Immersion et Présence. In *Le traité de la réalité virtuelle : Volume 1, L'homme et l'environnement virtuel* (Vol. 1, pp. 331): Presses de l'Ecole des Mines.
- Michoud, B., Guillou, E. & Bouakaz, S. (2006). *Extension de l'espace d'acquisition pour les méthodes de Shape-from-silhouette*. Paper presented at the Coresa 2006 (COmpression et REpresentation des Signaux Audiovisuels), Caen.
- Miners, B. W. (2001). *Kalman Filtering and Prediction for Hand Tracking*: University of Guelph
- Mistry, P., Maes, P. & Chang, L. (2009). *WUW - wear Ur world: a wearable gestural interface*. Paper presented at the 27th international conference extended abstracts on Human factors in computing systems, Boston, MA, USA, pp. 4111-4116.
- Mitobe, K., Kaiga, T., Yukawa, T., Miura, T., Tamamoto, H., Rodgers, A., et al. (2006). *Development of a motion capture system for a hand using a magnetic three dimensional position sensor*. Paper presented at the ACM SIGGRAPH 2006 Research posters, Boston, Massachusetts, p. 102.
- Mizuuchi, I. (2006). A Musculoskeletal Flexible-Spine Humanoid Kotaro Aiming at the Future in 15 Years Time. In A. Lazinica (Ed.), *Mobile Robots - Towards New Applications* (pp. 44-56): InTech.
- Mo, Z. & Neumann, U. (2006). *Real-time Hand Pose Recognition Using Low-Resolution Depth Images*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, pp. 1499-1505.

- Moeslund, T. B. & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3), 231-268.
- Moeslund, T. B., Hilton, A. & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2), 90-126.
- Mohr, D. & Zachmann, G. (2009). *Continuous Edge Gradient-based Template Matching for Articulated Object*. Paper presented at the International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, pp. 519-524.
- Mohr, D. & Zachmann, G. (2010a). FAST: Fast Adaptive Silhouette Area based Template Matching. In F. Labrosse, R. Zwiggelaar, Y. Liu & B. Tiddeman (Eds.), *Proceedings of the British Machine Vision Conference* (pp. 39.31-39.12): BMVA Press.
- Mohr, D. & Zachmann, G. (2010b). *Silhouette area based similarity measure for template matching in constant time*. Paper presented at the Proceedings of the 6th international conference on Articulated motion and deformable objects, Mallorca, Spain.
- Movea. (2009). MotionPod™ Technology. Retrieved 26/05/2009, 2009, from [http://movea.com/healthcare/motion\\_pod/index.html](http://movea.com/healthcare/motion_pod/index.html)
- Namiki, A., Sugano, R., Mizusawa, S., Yamakawa, Y. & Ishikawa, M. (2009). *High Speed Dexterous Manipulation with High Speed Vision*. Paper presented at the 9th IFAC Symposium on Robot Control (SYROCO2009), pp. 529-534.
- Napier, J. R. (1956). The prehensile movements of the human hand. *The Journal of bone and joint surgery*, 38-B(4), 902-913.
- Naud, M. (2008). *Contribution au développement d'outils d'aide à la conception assistée par la réalité virtuelle et augmentée : Application au domaine du textile/habillement*. Doctorat, Université d'Angers.
- Nedel, L. P., Dal Sasso Freitas, C. M., Jacob, L. J. & Pimenta, M. S. (2003). *Testing the Use of Egocentric Interactive Techniques in Immersive Virtual Environments*. Paper presented at the International Conference on Human-Computer Interaction, Zurich, Switzerland.
- Nielsen, J. (1994). *Usability Engineering*: Morgan Kaufmann.
- Nirei, K., Saito, H., Mochimaru, M. & Ozawa, S. (1996). *Human hand tracking from binocular image sequences*. Paper presented at the 22nd International Conference on Industrial Electronics, Control, and Instrumentation, Taipei, Taiwan, pp. 297-302.
- Noaman, N. M., Ahmmad, R. A. & Aula, A. (2008). Design and Implementation of DHM Glove Using Variable Resistors Sensors. *Journal of Artificial Intelligence*, 1(1), 44-52.
- Norman, D. A. (1988). *The psychology of everyday things*: Basic Books.
- Ouhaddi, H. (1999). *3d hand gesture tracking by model registration*. Paper presented at the International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'99), Santorini, Greece, pp. 70-73.
- Ouhaddi, H. & Horain, P. (1998, 13/03/1998). *Conception et ajustement d'un modèle 3D articulé de la main*. Paper presented at the 6èmes Journées de Travail du GT Réalité Virtuelle, Issy-les-Moulineaux, France, pp. 83-90.

- Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows* (3rd ed.): Open University Press.
- Pamplona, V. F., Fernandes, L. A. F., Prauchner, J., Nedel, L. P. & Oliveira, M. M. (2008). *The Image-Based Data Glove*. Paper presented at the Symposium on Virtual Reality (SVR'2008), Porto Alegre, pp. 204-211.
- Pausch, R. (1991). *Virtual reality on five dollars a day*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, New Orleans, Louisiana, USA, pp. 265-270.
- Pavlovic, V. I., Sharma, R. & Huang, T. S. (1997). Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7), 677-695.
- Perng, J. K., Fisher, B., Hollar, S. & Pister, K. S. J. (1999). *Acceleration Sensing Glove*. Paper presented at the 3rd IEEE International Symposium on Wearable Computers, p. 178.
- Piaget, J. & Chomsky, N. (1979). *Théories du langage, théories de l'apprentissage*: Seuil.
- Potamias, M. & Athitsos, V. (2008). *Nearest neighbor search methods for handshape recognition*. Paper presented at the Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments, Athens, Greece, pp. 1-8.
- Pouliquen, M. (2006). *Proposition d'un modèle de main pour la simulation des interactions homme-machine en environnement virtuel - Application à la prévention des risques d'accidents aux mains*. Doctorat, Ecole centrale de Nantes, Nantes.
- Prisacariu, V. A. & Reid, I. (2011). *Robust 3D hand tracking for human computer interaction*. Paper presented at the IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pp. 368-375.
- Rautaray, S. S. & Agrawal, A. (2011). A Real Time Hand Tracking System for Interactive Applications. *International Journal of Computer Applications*, 18(6), 28-33.
- Rehg, J. M. & Kanade, T. (1994). *Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking*. Paper presented at the Third European Conference - Volume II on Computer Vision, pp. 35-46.
- Rehg, J. M. & Kanade, T. (1995). *Model-based tracking of self-occluding articulated objects*. Paper presented at the Fifth International Conference on Computer Vision, p. 612.
- Rhône-Alpes, I. (2007). La plateforme GrImage. Retrieved 26/05/09, 2009, from <http://grimage.inrialpes.fr/>
- Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D. & Dantzich, M. v. (1998). *Data mountain: using spatial memory for document management*. Paper presented at the 11th annual ACM symposium on User interface software and technology, San Francisco, California, USA, pp. 153-162.
- Rohling, R. N. & Hollerbach, J. M. (1993). *Optimized fingertip mapping for teleoperation of dextrous robot hands*. Paper presented at the IEEE International Conference on Robotics and Automation, pp. 769-775.
- Romero, J. (2007). *Binocular Hand Tracking and Pose Estimation*. Master Thesis, CVAP/CAS, KTH, Stockholm.

- Romero, J., H., K. & D., K. (2010). *Hands in Action: Real-Time 3D Reconstruction of Hands in Interaction with Objects*. Paper presented at the IEEE International Conference on Robotics and Automation, Anchorage, AK, pp. 458-463.
- Romero, J., Kragic, D., Kyrki, V. & Argyros, A. (2008). *Dynamic time warping for binocular hand tracking and reconstruction*. Paper presented at the IEEE International Conference on Robotics and Automation, Pasadena, CA, USA, pp. 2289 – 2294.
- Saponas, T. S., Tan, D. S., Morris, D., Balakrishnan, R., Turner, J. & Landay, J. A. (2009). *Enabling always-available input with muscle-computer interfaces*. Paper presented at the 22nd annual ACM symposium on User interface software and technology, Victoria, BC, Canada, pp. 167-176.
- Saponas, T. S., Tan, D. S., Morris, D., Turner, J. & Landay, J. A. (2010). *Making muscle-computer interfaces more practical*. Paper presented at the 28th international conference on Human factors in computing systems, Atlanta, Georgia, USA, pp. 851-854.
- Sato, M. (2002). *Development of String-based Force Display: SPIDAR*. Paper presented at the 5th Biannual World Automation Congress.
- Savescu, A. (2006). *Analyse et modélisation du mouvement de préhension*. Doctorat, Université Lyon 1, Lyon.
- Saxena, A., Schulte, J. & Ng, A. Y. (2007). *Depth Estimation using Monocular and Stereo Cues*. Paper presented at the 20th international joint conference on Artificial intelligence, Hyderabad, India, pp. 2197-2203.
- Schaller, C. (2010). *Time-of-Flight cameras - new modality for radiotherapy*. PhD thesis, University Erlangen-Nuremberg.
- Scharstein, D. & Szeliski, R. (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision (IJCV)*, 47(1-3), 7-42.
- Schlattmann, M. & Klein, R. (2009). *Efficient Bimanual Symmetric 3D Manipulation for Markerless Hand-Tracking*. Paper presented at the Virtual Reality International Conference, Laval, France.
- Senoo, T., Yamakawa, Y., Mizusawa, S., Namiki, A., Ishikawa, M. & Shimojo, M. (2009). *Skillful Manipulation Based on High-speed Sensory-Motor Fusion*. Paper presented at the 2009 IEEE International Conference on Robotics and Automation, Kobe, pp. 1611-1612.
- Shadow\_Robot\_Company. (2003). *Design of a Dexterous Hand for Advanced CLAWAR Applications*. Paper presented at the 6th International Conference on Climbing and Walking Robots, Catania, Italy.
- Simone, L. K., Sundarajan, N., Luo, X., Jia, Y. & Kamper, D. G. (2007). A low cost instrumented glove for extended monitoring and functional hand assessment. *Journal of Neuroscience Methods*, 160(2), 335-348.
- Slabaugh, G. G., Culbertson, W. B., Malzbender, T. & Schafer, R. W. (2001). *A Survey of Methods for Volumetric Scene Reconstruction from Photographs*. Paper presented at the Volume Graphics, pp. 81-100.
- Smith, K. C., Sandin, D. J., Huang, T. S., Eliason, J. J. & Baum, G. A. (2003). *Real-time 3D hand tracking in a virtual environment*. Paper presented at the Stereoscopic Displays and Virtual Reality Systems, Santa Clara, CA, USA p. 529.

- Stefano, L. D., Marchionni, M. & Mattocchia, S. (2004). A fast area-based stereo matching algorithm. *Image and Vision Computing*, 22(12), 983-1005.
- Stenger, B. (2006). *Template-Based Hand Pose Recognition Using Multiple Cues*. Paper presented at the Computer Vision (ACCV 2006), Hyderabad, India, pp. 551-560.
- Stenger, B., Thayananthan, A., Torr, P. H. S. & Cipolla, R. (2003). *Filtering Using a Tree-Based Estimator*. Paper presented at the Ninth IEEE International Conference on Computer Vision - Volume 2, Nice, France, pp. 1063-1070.
- Stenger, B., Thayananthan, A., Torr, P. H. S. & Cipolla, R. (2006). Model-Based Hand Tracking Using a Hierarchical Bayesian Filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9), 1372-1384.
- Stenger, B., Thayananthan, A., Torr, P. H. S. & Cipolla, R. (2007). Estimating 3D hand pose using hierarchical multi-label classification. *Image Vision Comput.*, 25(12), 1885-1894.
- Sturman, D. J. (1992). *Whole-hand input*. PhD Thesis, Massachusetts Institute of Technology.
- Sturman, D. J. & Zeltzer, D. (1994). A Survey of Glove-based Input. *IEEE Computer Graphics and Applications*, 14(1), 30-39.
- Sun, Z., Bao, G. & Li, X. (2009). Experimental Research of Force Feedback Dataglove Based on Pneumatic Artificial Muscle. *Chinese Journal of Mechanical Engineering*, 22(4).
- Sun, Z., Bao, G., Li, X. & Lu, J. (2008). *Angle Measurement and Calibration of Force Feedback Dataglove*. Paper presented at the the 7th JFPS International Symposium on Fluid Power.
- Suryanarayan, P., Subramanian, A. & Mandalapu, D. (2010). *Dynamic Hand Pose Recognition Using Depth Data*. Paper presented at the 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3105-3108.
- Szeliski, R. (1993). Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1), 23-32.
- TAG. (2008). The economic payback of 3D mice for CAD Design Engineers: research findings, *Technology Assessment Group*.
- Thayananthan, A. (2005). *Template-based Pose Estimation and Tracking of 3D Hand Motion*. PhD Thesis, University of Cambridge.
- Theobalt, C., Albrecht, I., Haber, J., Magnor, M. & Seidel, H.-P. (2004). Pitching a baseball: tracking high-speed motion with multi-exposure images. *ACM Transactions on Graphics*, 23(3), 540-547.
- Tokatli, A. (2005). *3D Hand Tracking in video sequences*. Master thesis.
- Tosas, M. (2006). *Visual Articulated Hand Tracking for Interactive Surfaces*. PhD Thesis, University of Nottingham.
- Tosas, M. & Bai, L. (2007). *Virtual Touch Screen: a vision-based interactive surface*. Paper presented at the Ninth IASTED International Conference on Computer Graphics and Imaging, Innsbruck, Austria, pp. 81-86.

- Tricot, A., Plé gat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G. & Morcillo, A. (2003). *Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH*. Paper presented at the Conférence EIAH 2003, Strasbourg, France.
- Trindade, P., Lobo, J. & Camarinha-Matos, L. (2011). Distributed Accelerometers for Gesture Recognition and Visualization. In *Technological Innovation for Sustainability* (Vol. 349, pp. 215-223): Springer Boston.
- Turki, L. & Coiffet, P. (1995). A survey of dextrous telemanipulation with force feedback using the LRP virtual reality system. *Systems Analysis Modelling Simulation*, 18-19, 555-558.
- Turner, M. L., Findley, R. P., Griffin, W. B., Cutkosky, M. R. & Gomez, D. H. (2000). *Development and Testing of a Telemanipulation System with Arm and Hand Motion*. Paper presented at the ASME Dynamic Systems and Control Division, pp. 1-7.
- Tzafestas, C. & Coiffet, P. (1997). *Computing Optimal Forces for Generalised Kinesthetic Feedback on the Human Hand during Virtual Grasping and Manipulation*. Paper presented at the IEEE International Conference on Robotics and Automation, Albuquerque, New Mexico, USA.
- Ueda, E. (2003). Hand pose estimation for vision-based human interface. *IEEE Transactions on Industrial Electronics*, 50(4), 676-684.
- Wagner, C. (1988). The pianist's hand : anthropometry and biomechanics. *Ergonomics*, 31(1), 97-131.
- Wang, R. Y. (2008). *Real-Time Hand-Tracking as a User-Input Device*. Paper presented at the ACM Symposium on User Interface Software and Technology (UIST 2008).
- Wang, R. Y. (2011). *Practical color-based motion capture*. PhD Thesis, Massachusetts Institute of Technology.
- Wang, R. Y. & Popovi, J. (2009). Real-time hand-tracking with a color glove. *ACM Trans. Graph.*, 28(3), 1-8.
- Welman, C. (1993). *Inverse Kinematics And Geometric Constraints For Articulated Figure Manipulation*. PhD Thesis, Simon Fraser University.
- Wheeler, K. R. & Jorgensen, C. C. (2003). Gestures as Input: Neuroelectric Joysticks and Keyboards. *IEEE Pervasive Computing*, 2(2), 56-61.
- Williams II, R. (1998). Cable-Suspended Haptic Interface. *International Journal of Virtual Reality*, 3(3), 13-21.
- Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209-212.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.): McGraw-Hill.
- Winkler, S., Yu, H. & Zhou, Z. (2007). *Tangible reality desktop for digital media management*. Paper presented at the SPIE Engineering Reality of Virtual Reality, Singapore.
- Wise, S., Gardner, W., Sabelman, E., Valainis, E., Wong, Y., Glass, K., et al. (1990). Evaluation of a Fiber Optic Glove for Semiautomated Goniometric Measurements. *Journal of Rehabilitation Research and Development*, 27(4), 411-424.

- Wu, Y., Lin, J. Y. & Huang, T. S. (2001). *Capturing natural hand articulation* Paper presented at the Eighth IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, BC , Canada pp. 426-432.
- Yamada, K., Walairacht, S., Hasegawa, S., Koike, Y. & Sato, M. (2002). 4 + 4 fingers manipulating virtual objects in mixed-reality environment. *Presence: Teleoper. Virtual Environ.*, 11(2), 134-143.
- Yuan, M., Farbiz, F., Manders, C. M. & Tang, K. Y. (2008). *Robust hand tracking using a simple color classification technique*. Paper presented at the 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, Singapore, pp. 1-5.
- Zhang, S., Royer, D. & Yau, S.-T. (2006). GPU-assisted high-resolution, real-time 3-D shape measurement. *Optics Express*, 14(20), 9120-9129.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 13330-11334.
- Zhao, Y.-j. & Dai, S.-l. (2009). *A Robust and Fast Monocular-Vision-Based Hand Tracking Method for Virtual Touch Screen*. Paper presented at the 2nd International Congress on Image and Signal Processing (CISP '09) Tianjin, pp. 1-5.
- Zhou, H. & Hu, H. (2008). Human motion tracking for rehabilitation - A survey. *Biomedical Signal Processing and Control*, 3(1).
- Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S. & Harvill, Y. (1986). A hand gesture interface device. *SIGCHI Bull.*, 18(4), 189-192.



# CONCEPTION ET EVALUATION D'UN SYSTEME TRANSPARENT DE CAPTURE DE MOUVEMENTS DES MAINS POUR L'INTERACTION 3D TEMPS REEL EN ENVIRONNEMENTS VIRTUELS

**RESUME :** Le but de cette thèse est de proposer et d'évaluer un système transparent de capture de mouvements des mains en temps réel pour permettre la réalisation d'interactions 3D en environnement virtuel (E.V.). Les outils tels que le clavier et la souris ne suffisent plus lorsque l'on travaille dans des E.V. 3D et les systèmes de capture de mouvements actuels ont l'inconvénient du coût et du port de matériel. Les systèmes basés sur des caméras et le traitement d'images comblent en partie ces lacunes, mais ne permettent pas encore une capture de mouvements 3D précise, efficace et temps réel. Notre système propose une solution à ce problème grâce à une caméra 3D. Nous avons implémenté des modalités qui permettent une interaction plus naturelle avec les objets et l'E.V. L'objectif de notre système est d'obtenir des performances au moins égales à celles d'outils couramment utilisés en réalité virtuelle tout en offrant une meilleure acceptabilité globale, mesurée sous l'angle de l'utilité, de l'utilisabilité et du sentiment d'immersion.

Pour répondre à cet objectif, nous avons mené 3 études expérimentales impliquant plus de 100 participants. L'étude 1 a comparé, pour une tâche de sélection, la première version de notre système (basé sur une Camera 3D MESA SwissRanger) à une souris classique. L'expérimentation 2 est axée sur l'étude de la réalisation de tâches de manipulation d'objets (positionnement, orientation, redimensionnement) et de navigation en E.V. Pour cette étude, nous avons comparé la version améliorée de notre système (basée sur la Microsoft Kinect) à des gants de données associés à des capteurs magnétiques. La 3ème étude, complémentaire, porte sur l'évaluation de nouvelles modalités d'interaction, implémentées d'après les commentaires recueillis auprès des participants de la seconde étude.

**Mots clés :** Capture de mouvements des mains, interaction 3D, interaction bi-manuelle, périphérique transparent, temps réel, caméra 3D.

## DESIGN AND EVALUATION OF A MARKERLESS HAND MOTION CAPTURE SYSTEM FOR 3D REAL TIME INTERACTION IN VIRTUAL ENVIRONMENTS

**ABSTRACT:** The purpose of this thesis is to propose and evaluate a markerless system for capturing hand movements in real time to allow 3D interactions in virtual environment (VE). Tools such as keyboard and mouse are not enough for interacting in 3D VE; current motion capture systems are expensive and require wearing equipments. Systems based on cameras and image processing partially fill these gaps, but do not yet allow an accurate, efficient and real-time 3D motion capture. Our system provides a solution to this problem with a 3D camera. We have implemented modalities that allow a more natural interaction with objects and VE. The goal of our system is to obtain performances at least equal to those of common tools in virtual reality while providing a better overall acceptability (i.e., usefulness, usability, immersion).

To achieve this goal, we conducted three experimental studies involving over 100 participants. With the first study, we compared the first version of our system (based on a 3D Camera MESA SwissRanger) to a traditional mouse, for a selection task. The second experiment was focused on the study of the realization of objects-manipulation tasks (position, orientation, scaling) and navigation tasks in VE. For this study, we compared the improved version of our system (based on Microsoft Kinect) with data gloves associated with magnetic sensors. An additional study concerns the evaluation of new modalities of interaction, implemented based on participants' feedback of the second study.

**Keywords:** Hand motion capture, 3D interaction, bimanual interaction, markerless device, real-time, 3D camera.

