



HAL
open science

Interagir en réalité mixte avec un casque : application à la chirurgie augmentée

Charles Bailly

► To cite this version:

Charles Bailly. Interagir en réalité mixte avec un casque : application à la chirurgie augmentée. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALM064 . tel-03185064

HAL Id: tel-03185064

<https://theses.hal.science/tel-03185064>

Submitted on 30 Mar 2021

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

Charles Bailly

Thèse dirigée par **Laurence Nigay**, Professeur, Université Grenoble-Alpes, et
co-encadrée par **François Leitner**

préparée au sein du **Laboratoire d'Informatique de Grenoble**
dans l'**École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Interagir en Réalité Mixte avec un casque : Application à la Chirurgie Augmentée

Thèse soutenue publiquement le **16 décembre 2020**,
devant le jury composé de :

Madame Laurence Nigay

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes,
Directrice de thèse

Monsieur Martin Hachet

DIRECTEUR DE RECHERCHE, INRIA Bordeaux Sud-Ouest, Rapporteur

Monsieur Marcos Serrano

MAITRE DE CONFERENCE, Université de Toulouse III, Rapporteur

Monsieur Frédéric Noël

PROFESSEUR DES UNIVERSITES, Grenoble INP, Président

Monsieur François Faure

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes, Invité

Monsieur François Leitner

INGENIEUR DOCTEUR, Aesculap, Invité et Co-encadrant



Abstract

This thesis is in the field of Human-Computer Interaction (HCI) and specifically focuses on user interaction with Mixed Reality systems that rely on Head-Mounted Displays. Mixed Reality (MR) enables us to create mixed environments that combine virtual content with physical content. The entire workspace, including both physical and virtual objects, can then be used to interact with the system. Unlike traditional graphical interfaces, MR combines the advantages of the physical environment while providing the flexibility of virtual elements. The fields of application are many and varied, ranging from industrial manufacturing to medicine.

Surgery is an application context in which MR is particularly promising. The ability to view medical data directly on the patient's body and around the operating table is a crucial benefit for surgeons. However, the existing techniques for interacting with this virtual information are not suited to the many constraints of the operating room. New approaches are needed to enable surgeons to view, manipulate and edit virtual content during an operation.

Our research is driven by this need for MR interaction techniques for augmented surgery. Thanks to our partnership with the company Aesculap for this CIFRE thesis, our work is dedicated to knee prosthesis operations. We focus on the use of Head-Mounted Displays (HMDs), through which the augmented field of view is directly related to the position of the head.

We start by providing a design space for MR head-based interaction. We explore this design space by devising new interaction techniques with menus. These techniques involve the entire reality-virtuality continuum by considering different physicalities for targets and implementing transitions between mixed reality and virtual reality. Our goal is to fulfill the surgery constraints by taking advantage of the objects already present in the operating room.

After exploring head-based techniques, we broaden our area of research by considering surgical stages where surgeons have at least one hand available to interact with the system. Our contributions on this aspect are twofold. We first continue our study of menu techniques by considering a key aspect of the surgeon's work : visual attention to the operating field. We thus propose three new techniques allowing the surgeon to bring virtual widgets back into his field of vision without having to look away from the operating field. We then study the exploration of MR 3D scenes with multiple views. For instance, multiple views of the knee are displayed during the surgical planning step. The originality of the work lies in the different nature of the multiple views : virtual views and mixed views. We propose two new complementary interaction techniques that allow users to control the level of coupling between these different views and implement transitions between virtual reality and mixed reality.

Résumé

Cette thèse s'inscrit dans le domaine de l'Interaction Homme-Machine (IHM) et est dédiée à l'interaction en Réalité Mixte (RM) avec un casque. La RM permet de créer des environnements mixtes mêlant du contenu virtuel à du contenu physique. L'ensemble de l'espace de travail, objets physiques comme virtuels, peuvent alors servir à interagir avec le système. Contrairement aux interfaces graphiques traditionnelles, la RM combine les avantages de l'environnement physique tout en offrant la flexibilité des éléments virtuels. Les domaines d'application sont nombreux et variés, allant de la fabrication industrielle à la médecine.

La chirurgie est un contexte applicatif dans lequel la RM est particulièrement prometteuse. La possibilité de visualiser des données médicales directement sur le corps du patient et autour de la table d'opération est en effet un atout crucial pour les chirurgiens. Cependant, les techniques existantes permettant d'interagir avec ces informations virtuelles ne sont pas adaptées aux nombreuses contraintes du bloc opératoire. De nouvelles approches sont nécessaires pour permettre aux chirurgiens de visualiser, manipuler et éditer du contenu virtuel pendant une opération.

Notre recherche est motivée par ce besoin de techniques d'interaction en RM pour la chirurgie augmentée. Grâce à notre partenariat avec la société Aesculap pour cette thèse CIFRE, notre travail est dédié aux opérations de prothèse de genou. Nous nous concentrons sur l'utilisation de casques de RM (*Head-Mounted Display*), avec lesquels le champ de vision augmenté est directement lié à la position de la tête.

Nous commençons par proposer un espace de conception pour l'interaction basée sur les mouvements de la tête en RM. Nous explorons cet espace de conception en concevant de nouvelles techniques d'interaction pour interagir avec des menus. Ces techniques impliquent l'ensemble du continuum réalité-virtualité en considérant différentes physicalités pour les cibles et des transitions entre la réalité mixte et la réalité virtuelle. Notre but est de respecter les contraintes du chirurgien en tirant partie des objets déjà présents en bloc opératoire.

Après avoir exploré les techniques basées sur la tête, nous élargissons notre domaine de recherche en considérant les étapes chirurgicales où les chirurgiens ont au moins une main disponible pour interagir avec le système. Nos contributions sur cet aspect sont doubles. Nous poursuivons tout d'abord notre étude de l'interaction avec les menus en considérant un aspect clé du travail du chirurgien : l'attention visuelle sur le champ opératoire. Nous proposons ainsi trois nouvelles techniques permettant au chirurgien de ramener des widgets virtuels dans son champ de vision sans avoir besoin de détourner le regard du champ opératoire. Ensuite, nous étudions l'exploration de scène 3D en RM avec des vues multiples en nous inspirant de l'étape du *planning* chirurgical. L'originalité des travaux réside dans la nature différente des vues multiples considérées : vue virtuelle et vue mixte. Nous proposons deux nouvelles techniques d'interaction complémentaires qui permettent de contrôler le niveau de couplage entre ces vues de nature différente et implémentent des transitions entre réalité augmentée et réalité virtuelle.

Remerciements

Je voudrais commencer par remercier Laurence Nigay et François Leitner. En tant que directrice de thèse, Laurence a été d'une aide précieuse tout au long de ces trois années (3,5 années même en comptant le stage de MoSIG!). Ses conseils au cours de nos longues discussions et sa disponibilité malgré ses nombreuses responsabilités font d'elle une mentor remarquable envers qui je suis particulièrement reconnaissant. François n'est pas en reste : son encadrement côté entreprise a rendu cette thèse Cifre possible. C'est grâce à son soutien continu que mes recherches ont pu être appliquées à un domaine d'application concret qui a du sens pour moi.

J'en profite pour remercier les membres de l'équipe d'Aesculap et de l'équipe IIHM. Que ce soit les discussions autour d'un café ou simplement à la cafétéria, les coups de pouce pour le développement ou les nombreuses exps, ces années de thèse ont été pour moi très agréables et enrichissantes, et c'est grâce à vous. Un grand merci aux doctorants/post-docs de l'équipe pour les échanges, les rires et les soirées à l'indétrônable Docteur D. Les choses n'auraient pas été pareilles sans vous.

Je souhaite également remercier les rapporteurs de cette thèse, Martin Hachet et Marcos Serrano. Merci également à Frédéric Noël et François Faure d'avoir accepté de faire partie du jury en tant qu'examineur et invité. Cette fin d'année 2020 aura été particulièrement agitée et je vous suis tous reconnaissant de prendre le temps de découvrir et d'échanger autour de mes travaux de recherche.

Je remercie également ma famille pour son soutien sans faille. Oui, Grenoble est relativement éloigné de votre chère campagne Bazadaise et de la célèbre ville rose, mais cela ne vous a pas empêché de toujours m'encourager et me pousser vers l'avant. Même après presque trois mois de cohabitation à Sansoutic pendant le 1er confinement.

Enfin, je tiens à remercier ma fidèle paire de sandales qui m'a été d'une aide précieuse au cours de ces cinq années à Grenoble. De nombreux moustiques ont cru pouvoir ruiner mes nuits en toute impunité. Ils ont eu tort.

Table des matières

Introduction	13
I Cadre de nos travaux : Réalité Mixte et Chirurgie Augmentée	16
1 Réalité Mixte	18
1.1 Définitions de la Réalité Mixte	19
1.1.1 Définition centrée sur le type d'affichage	19
1.1.2 Définition basée sur l'interaction	28
1.1.3 Réalité Mixte dans le cadre de nos travaux	29
1.2 Interaction en Réalité Mixte	31
1.2.1 Objets virtuels proches : Manipulation directe	31
1.2.2 Objets virtuels distants : Pointage	34
1.2.3 Sélection de commandes	41
2 Chirurgie Augmentée	45
2.1 Réalité Mixte en bloc opératoire	46
2.1.1 Chirurgie orthopédique classique	46
2.1.2 Apports de la Réalité Mixte pour la chirurgie	50
2.1.3 Limites de la réalité mixte en bloc opératoire	53
2.2 Interaction en bloc opératoire	59
2.2.1 Contraintes existantes en bloc opératoire	60
2.2.2 Techniques existantes utilisables en bloc opératoire	62
2.2.3 Bilan de l'interaction en bloc opératoire	65
II Interaction : techniques de menu et de navigation	69
3 Interaction avec la tête : techniques de menu	71
3.1 Espace de conception pour l'interaction avec les mouvements de la tête en RM dans le cas de menus	72
3.1.1 Élément d'interaction contrôlé par les mouvements de la tête .	73
3.1.2 Physicalité de la cible	75
3.2 Conception de nouvelles techniques de menu	78
3.2.1 Principes de conception	78
3.2.2 Implémentation des techniques	79
3.3 Études expérimentales	81
3.3.1 Étude expérimentale 1 : HCC et HCM	81

3.3.2	Étude expérimentale 2 : transitions entre les modes RA et RV	89
3.3.3	Étude expérimentale 3 : physicalité des menus	91
3.4	Leçons apprises et discussion	102
3.4.1	HCC et HCM	103
3.4.2	Menus virtuels	103
3.4.3	Menus mixtes	103
3.4.4	Menus physiques	104
4	Interaction tête + main : techniques de menu <i>Bring2Me</i>	107
4.1	Éléments de conception	108
4.1.1	Sélection d'un menu	109
4.1.2	Déplacement du menu	110
4.1.3	Sélection d'un élément du menu	110
4.2	Conception de nouvelles techniques de menu : <i>Bring2Me</i>	110
4.3	Étude expérimentale	112
4.3.1	Protocole expérimental	112
4.3.2	Résultats	115
4.4	Leçons apprises et discussion	119
4.4.1	Tourner la tête	119
4.4.2	Techniques <i>Bring2Me</i>	119
4.4.3	Perspectives	121
5	Interaction tête + main : techniques de navigation dans une scène 3D en RM	125
5.1	Éléments de conception	128
5.1.1	Couplage des vues	128
5.1.2	Navigation : changement de point de vue	130
5.2	Conception de nouvelles techniques pour changer le point de vue	132
5.3	Études expérimentales	133
5.3.1	Installation expérimentale	134
5.3.2	Étude pilote : exploration informelle de l'espace de conception	135
5.3.3	Deuxième étude : approfondissement de la navigation virtuelle	137
5.4	Leçons apprises et discussion	143
5.4.1	Stratégies	143
5.4.2	Vue de référence	144
5.4.3	Transitions entre RA et RV	145
6	Prototype d'OrthoPilot® en RM	147
6.1	Fonctionnement général du prototype	148
6.1.1	OrthoPilot® sans Réalité Mixte	148
6.1.2	Ajout des techniques de Réalité Mixte	149
6.1.3	Visualisation d'informations médicales	149
6.2	Techniques de menus basées sur les mouvements de la tête	151
6.2.1	Curseur virtuel contrôlé avec la tête	151
6.2.2	Menu virtuel contrôlé avec la tête	152
6.3	Techniques d'interaction tête + main : <i>Bring2Me</i>	152
6.4	Évaluation du prototype	154

Conclusion	156
A Questionnaire Raw-TLX	165
B Questionnaire final études 1 et 2	166
C Questionnaire final étude 3	167
D Questionnaire final étude 4	168
E Questionnaire étude 5	169

Table des figures

A	Chirurgie classique vs chirurgie augmentée	13
B	Plan du mémoire	15
1.1	Plan du mémoire	18
1.2	Continuum Réalité-Virtualité	19
1.3	Réalité virtuelle	19
1.4	Exemple de virtualité augmentée	21
1.5	Augmenter une scène virtuelle	21
1.6	Exemple de Réalité Remixée	22
1.7	Exemple de scène en Réalité Augmentée	23
1.8	Intégration spatiale d'objets virtuels	23
1.9	Modalités d'affichage en RA	25
1.10	Exemple de Réalité Diminuée	27
1.11	Sous-ensembles de la Réalité Mixte	27
1.12	Augmenter une scène virtuelle	28
1.13	Interaction directe avec un gant	32
1.14	Exemple de RA tangible	33
1.15	Technique d'interaction bimanuelle en RA tangible	34
1.16	Modélisation d'une tâche de pointage	35
1.17	Illustration des paramètres de la steering law	36
1.18	Techniques de pointage en Réalité Mixte	37
1.19	Taille des champs de vision en Réalité Mixte	39
1.20	Peephole pointing	39
1.21	Pointer sans utiliser le regard	40
1.22	Interaction gestuelle via un gant électronique	41
1.23	Exemple de menu invoqué	43
1.24	Combiner commandes vocales et gestes pour interagir	44
2.1	Plan du mémoire	45
2.2	Robot chirurgical da Vinci	47
2.3	Matériel lié à l'OrthoPilot®	48
2.4	Captures d'écran de l'OrthoPilot®	49
2.5	Pointage avec l'OrthoPilot® en bloc opératoire	50
2.6	Organe augmenté en RA	51
2.7	Maxillaire augmentée	52
2.8	Entraînement guidé en RM	52
2.9	Annotations virtuelles en RM	53
2.10	Précision de la réalité mixte pour une tâche manuelle	55
2.11	Conflit accommodation-convergence	56

2.12	Techniques de visualisation pour la perception de la profondeur	57
2.13	Cécité d'inattention en RA	58
2.14	Gestion de l'espace en bloc opératoire	61
2.15	Pédales chirurgicales	64
2.16	Intersection de l'IHM et de la chirurgie augmentée	65
2.17	Étapes d'une opération et interaction possibles	66
2.18	Techniques d'interaction existantes et bloc opératoire	67
3.1	Plan du mémoire	71
3.2	Étapes d'une opération et interaction possibles	72
3.3	Exemple de menu de l'OrthoPilot®	73
3.4	Contrôler un curseur virtuel avec la tête	74
3.5	Contrôler un menu virtuel avec la tête	74
3.6	Premier axe de l'espace de conception	75
3.7	Exemples de menus virtuels, mixtes et physiques	76
3.8	Second axe de l'espace de conception	76
3.9	Espace de conception complet	77
3.10	Rigid-bodies utilisés comme ancres physiques	77
3.11	Nouvelles techniques par rapport à l'espace de conception	78
3.12	Le casque Epson Moverio BT-300	79
3.13	Suivi des mouvements la tête de l'utilisateur	80
3.14	Installation de la première expérimentation	82
3.15	Techniques HCC et HCM lors de la première expérimentation	83
3.16	Déroulement de l'étude 1	84
3.17	Étude 1 : temps de sélection moyens	85
3.18	Étude 1 : erreurs moyennes	86
3.19	Étude 1 : retours qualitatifs	87
3.20	Transition de RA à RV au cours du pointage	90
3.21	Étude 2 : temps de sélection	91
3.22	Étude 3 : Exemples de menus virtuels, mixtes et physiques	92
3.23	Étude 3 : installation expérimentale	94
3.24	Étude 3 : Comparaison des trois types de menu	95
3.25	Étude 3 : Déroulement d'une étape	96
3.26	Étude 3 : Guidage textuel	97
3.27	Déroulement de l'étude 3	98
3.28	Étude 3 : temps de sélection	99
3.29	Étude 3 : transfert asymétrique de compétence	100
3.30	Étude 3 : taux d'erreur moyen	101
3.31	Étude 3 : retours qualitatifs	102
3.32	Étude 3 : avantages et limitations de chaque type de menu	105
4.1	Plan du mémoire	107
4.2	Étapes d'une opération et interaction possibles	108
4.3	Principe des techniques <i>Bring2Me</i>	109
4.4	Fonctionnement des techniques <i>Bring2Me</i>	111
4.5	Étude 4 : installation expérimentale	113
4.6	Étude 4 : déroulement de l'expérimentation	115
4.7	Étude 4 : temps de sélection	116
4.8	Étude 4 : taux d'erreur moyens	117

4.9	Étude 4 : retours qualitatifs	118
4.10	Étude 4 : Bénéfices et limitation des techniques étudiées	123
5.1	Plan du mémoire	125
5.2	Étapes d'une opération et interaction possibles	126
5.3	Étape de planning chirurgical	127
5.4	Combiner vue mixte et vue virtuelle en RM	128
5.5	Mode VR temporaire	131
5.6	3ème espace de conception	132
5.7	La Boîte et le Patron	132
5.8	Les 4 techniques de changement de point de vue	133
5.9	Pilote étude 5	134
5.10	Comparaison des faces spéciales	138
5.11	Étude 5 : Temps de complétion et erreurs	140
5.12	Étude 5 : retours qualitatifs	141
5.13	Stratégies utilisées par les participants	142
5.14	Étude 5 : Stratégies et performances	143
6.1	Plan du mémoire	147
6.2	OrthoPilot®sans Réalité Mixte	148
6.3	Prototype	150
6.4	Technique HCC→Virtuel au sein du prototype	152
6.5	Technique HCC→Mixte au sein du prototype	152
6.6	Technique HCM→Mixte au sein du prototype	153
C	Résumé des contributions	158

Liste des tableaux

4.1	Caractéristiques des techniques <i>Bring2Me</i>	109
5.1	Étude 5 : recommandations proposées	145

Introduction

Il y a quelques années, la notion de réalité augmentée était encore un concept futuriste et nébuleux pour le grand public. La réalité virtuelle, légèrement en avance d'un point de vue technologique, n'était encore qu'un joli oxymore aux yeux de bien des gens. Et pourtant... Des casques et lunettes de réalité augmentée commerciaux sont à présent disponibles. Des smartphones permettent de visualiser un meuble virtuel dans son salon. Qui n'a jamais entendu parler d'applications comme le fameux Pokémon Go et sa chasse aux créatures virtuelles dans les rues et les parcs municipaux ?

À l'heure actuelle, les différentes technologies mêlant environnement virtuel et environnement réel se sont multipliées. La réalité mixte commence à impacter de nombreux domaines d'applications tels que la maintenance industrielle, l'urbanisme ou encore la médecine. Citons par exemple l'utilisation de casques de réalité mixte dans certains hôpitaux britanniques pour réduire le temps d'exposition du personnel soignant aux patients atteints de la Covid-19 [54].

Cette thèse CIFRE a été réalisée en collaboration avec la société Aesculap, qui produit des prothèses médicales et des logiciels chirurgicaux. Le but de tels logiciels est d'aider le chirurgien à mener à bien l'opération, par exemple en simulant les conséquences de plusieurs tailles de prothèse et en affichant des informations médicales sur un écran. Cet écran ne peut néanmoins pas être à proximité immédiate du patient, ce qui force le chirurgien à interagir à distance avec le système (Figure Aa).



FIGURE A – a) Chirurgie classique avec écran distant. b) Chirurgie augmentée avec casque de RM. Images originales fournies par Aesculap.

La réalité mixte présente de nombreux atouts pour aider le travail du chirurgien et ne plus dépendre d'un unique écran distant. En effet, elle permet de voir des

informations médicales sous la forme d'objets virtuels 2D et 3D directement superposées sur le corps du patient. Des objets virtuels tels que des radios et des menus virtuels peuvent être intégrés autour du patient afin de faciliter la prise de décision du chirurgien et son interaction avec le système (Figure Ab).

Nos travaux en Interaction Homme-Machine s'inscrivent dans ce contexte de chirurgie (orthopédique) augmentée. Nous nous intéressons aux techniques d'interaction avec un casque de réalité mixte qui transforme l'ensemble du bloc opératoire en un environnement interactif. Nous nous éloignons ainsi de l'interaction confinée à un écran (interface graphique WIMP) pour nous concentrer sur l'interaction dans l'environnement mixte 3D autour de l'utilisateur (post-WIMP). De nombreuses possibilités d'interaction sont envisageables. Néanmoins, le bloc opératoire impose de nombreuses contraintes au personnel médical lors d'une opération, notamment les contraintes de stérilité des équipements proches du patient. Interagir avec des objets virtuels (hologrammes) en bloc opératoire soulève donc de nombreuses questions.

Nos travaux s'intéressent à la conception de nouvelles techniques permettant d'interagir avec des objets virtuels dans le cas de la réalité mixte avec un casque. En particulier, nous nous concentrons sur des domaines d'application où les utilisateurs ont de fortes contraintes qui limitent leurs possibilités d'interaction, notamment les chirurgiens lors d'une opération. Ainsi, nous explorons la question de recherche suivante :

Quelles techniques d'interaction permettent à un utilisateur contraint d'interagir avec du contenu virtuel en réalité mixte ?

Pour tenter de répondre à cette question, nous commençons dans une première partie (Chapitres 1 et 2) par cerner la notion de réalité mixte (Chapitre 1) afin de définir le contexte de nos travaux. Nous soulignons l'absence de consensus concernant la terminologie à employer et discutons des définitions parfois contradictoires présentes dans la littérature. Nous mettons en évidence les différences entre les notions de réalité augmentée, réalité mixte et réalité virtuelle. Nous présentons ensuite une synthèse de l'état de l'art sur les techniques d'interaction existantes. Cette synthèse récapitule les différentes classes de techniques d'interaction existantes en RM. Nous exposons les enjeux et besoins de notre contexte applicatif principal : la chirurgie augmentée (Chapitre 2). Nous y précisons pourquoi les techniques d'interaction existantes en réalité mixte ne sont pas adaptées pour des utilisateurs contraints comme les chirurgiens.

Pour répondre à ce besoin de nouvelles techniques utilisables en bloc opératoire, nous présentons nos travaux sur l'interaction en bloc opératoire au sein d'une seconde partie (Chapitres 3 à 6). Notre approche scientifique allie 1) des travaux conceptuels avec l'établissement d'espaces de conception et 2) des travaux expérimentaux en laboratoire afin d'évaluer les techniques d'interaction conçues et développées. Notre approche scientifique est donc itérative reposant sur une démarche inductive et déductive. Cette approche se retrouve dans la structure commune des trois chapitres (3, 4 et 5) : une section sur les éléments et espaces de conception, une section décrivant les techniques d'interaction conçues à partir de l'espace de conception défini et une section sur les études expérimentales des techniques conçues.

Nous commençons par détailler un espace de conception pour l'interaction avec des menus basée sur les mouvements de la tête (Chapitre 3). Nous présentons ensuite

les techniques d'interaction qui sont issues de cet espace de conception ainsi que les trois évaluations expérimentales de ces techniques. Dans les deux chapitres suivants, nous élargissons nos recherches à l'interaction tête + main. Les chirurgiens peuvent en effet avoir une main libre lors de certaines étapes de la chirurgie. Dans un premier temps, nous poursuivons notre étude de l'interaction avec des menus en nous concentrant sur l'attention visuelle de l'utilisateur, qui est un facteur clé en chirurgie. Nous proposons et étudions expérimentalement de nouvelles techniques d'interaction permettant de ramener des widgets virtuels dans le champ de vision du chirurgien (Chapitre 4). Dans un second temps (Chapitre 5), nous étudions l'exploration d'une scène 3D avec un système de RM combinant des vues de différentes natures : une vue mixte (la jambe augmentée du patient) et une vue virtuelle (une copie virtuelle de la jambe). Pour cela, nous partons d'une étape chirurgicale existante : le planning chirurgical. Nous proposons et évaluons expérimentalement de nouvelles techniques d'interaction pour la tâche d'exploration qui impliquent l'utilisation d'une main.

La structure globale du mémoire est schématisée à la Figure B. Nous concluons ce mémoire en résumant nos contributions et en présentant les perspectives à court terme de nos travaux de recherche, puis à long terme. Ce mémoire contient enfin une bibliographie et 5 documents annexes.

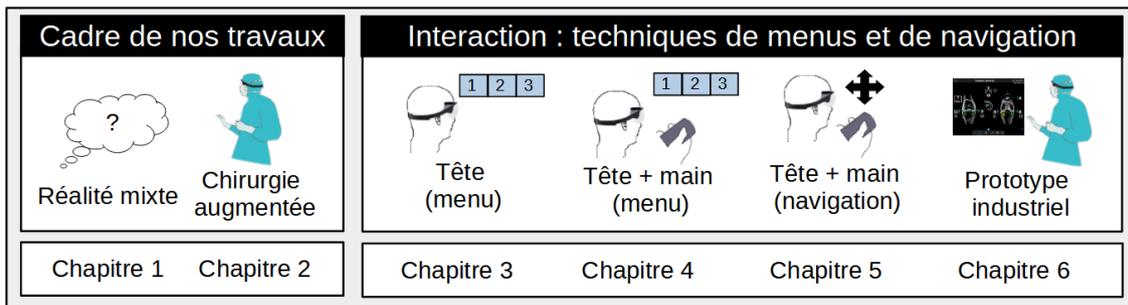


FIGURE B – Plan du mémoire

Première partie

Cadre de nos travaux : Réalité Mixte et Chirurgie Augmentée

Chapitre 1

Réalité Mixte

Sommaire

1.1 Définitions de la Réalité Mixte	19
1.1.1 Définition centrée sur le type d’affichage	19
1.1.2 Définition basée sur l’interaction	28
1.1.3 Réalité Mixte dans le cadre de nos travaux	29
1.2 Interaction en Réalité Mixte	31
1.2.1 Objets virtuels proches : Manipulation directe	31
1.2.2 Objets virtuels distants : Pointage	34
1.2.3 Sélection de commandes	41

Introduction

L’objectif de ce chapitre est d’expliquer ce que nous désignons par Réalité Mixte (RM) dans nos travaux. De nombreux concepts sont liés à l’augmentation ou au remplacement de l’environnement réel, de la Réalité Augmentée (RA) à la Réalité Virtuelle (RV). Il est donc nécessaire de commencer par définir chacun de ces termes plus ou moins équivoques. Au delà du célèbre Continuum Réalité-Virtualité de Milgram [146], nous verrons que les contours précis de la Réalité Mixte sont encore difficiles à cerner. Plusieurs définitions et facettes de cette technologie existent au sein de la communauté scientifique. Nous positionnons donc nos travaux par rapport aux discussions présentes dans la littérature. Dans un second temps, nous présentons également les grandes classes de techniques d’interaction liées à la RM. Pour cela, nous passons en revue les techniques de manipulation directe d’objets virtuels, de pointage et de déclenchement de commandes en RM.

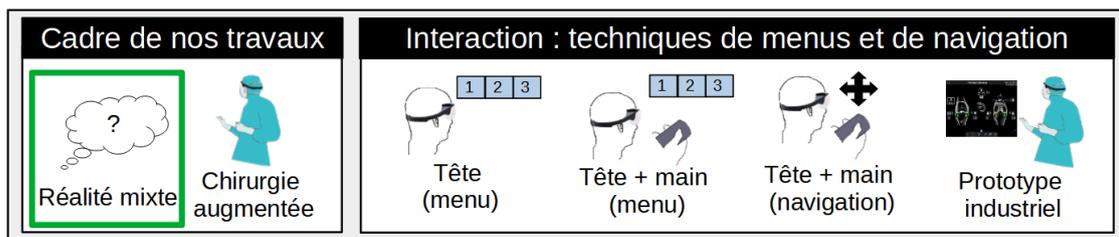


FIGURE 1.1 – Plan du mémoire (Figure B) - Focus sur le premier chapitre.

1.1 Définitions de la Réalité Mixte

1.1.1 Définition centrée sur le type d'affichage

La définition la plus connue et la plus utilisée du concept de Réalité Mixte est probablement celle issue des travaux de Milgram et al. en 1994 [146]. Les auteurs proposent de classer différentes technologies basées sur des environnements réels et/ou virtuels selon un axe. Cet axe, ou plutôt ce continuum, est basé sur le degré de *virtualité* de ces technologies.

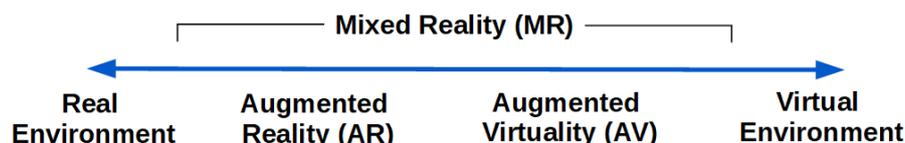


FIGURE 1.2 – Continuum Réalité-Virtualité défini par Milgram[146].

À gauche, la première extrémité du continuum correspond à l'environnement réel, c'est à dire sans aucun élément virtuel. Les interactions possibles dans cet environnement sont basées sur des objets purement physiques. Autrement dit, l'utilisateur accomplit sa tâche dans le monde réel sans contenu virtuel perceptible. À l'autre extrémité du Continuum Réalité-Virtualité de Milgram et al. se trouve le cas opposé : l'environnement purement virtuel [146]. Dans ce cas, tout ce que perçoit l'utilisateur est généré et contrôlé par le système. Cette approche est qualifiée de Réalité Virtuelle (RV). L'exemple le plus courant de RV repose sur l'utilisation d'un casque électronique qui masque la vue de son porteur. Au lieu de percevoir le monde physique qui l'entoure, l'utilisateur perçoit un environnement entièrement virtuel, comme illustré à la Figure 1.3. Une autre méthode d'affichage consiste à utiliser des systèmes CAVE (*Cave Augmented Virtual Environments*), qui sont des salles spécialisées dans lesquelles plusieurs murs servent de support à la projection de contenu virtuel.



FIGURE 1.3 – Exemple de système de Réalité Virtuelle (RV). a) Un utilisateur portant un casque de RV. b) Environnement entièrement numérique perçu par l'utilisateur. Image modifiée issue de https://en.wikipedia.org/wiki/File:Reality_check_ESA384313.jpg

Différentes façons de définir la RV sont présentes dans la littérature [222]. La définition générale se base sur la notion d'environnement numérique. L'accent est mis sur le fait qu'un système informatique génère ce que perçoit l'utilisateur et que divers dispositifs (gants électroniques, contrôleurs, etc) lui permettent d'interagir avec cet environnement numérique [222]. L'un des concepts fondamentaux de la RV est la notion d'immersion. Steuer insiste sur cette sensation d'être plongé dans un environnement, qualifiée de téléprésence [188]. Le sentiment d'immersion ne se limite d'ailleurs pas à la vue. D'autres sens comme l'ouïe [150], le toucher [1] voir l'odorat [59] peuvent être stimulés afin de renforcer l'impression d'appartenir à l'environnement. Dans les deux cas, l'environnement généré peut être la reproduction d'un environnement réel ou représenter un monde complètement imaginaire.

Plusieurs casques de RV sont aujourd'hui disponibles dans le commerce. Pour le grand public, de tels équipements sont essentiellement centrés sur le divertissement, comme par exemple les jeux-vidéo. Au niveau professionnel, la maturité technologique de la RV permet également de l'utiliser comme un outil de formation. La capacité à simuler un environnement complet (à 360 degrés, à différentes échelles, etc) ou de répéter indéfiniment des situations complexes est en effet un atout précieux pour l'enseignement et l'entraînement. On peut ainsi citer l'exemple d'apprentis chirurgiens se formant à certaines procédures chirurgicales en RV [4].

Les concepts liés aux deux extrémités du Continuum Réalité-Virtualité sont clairement définis et cernés. En revanche, la zone au milieu du Continuum regroupe différents concepts aux limites plus confuses. Ci-dessous, nous présentons ces concepts généraux dans l'ordre décroissant de leur niveau de virtualité (RV \rightarrow Environnement physique).

Virtualité Augmentée

Moins connue que la RV, la Virtualité Augmentée (VA) repose elle aussi sur l'immersion de l'utilisateur dans un environnement virtuel. Cependant, contrairement à la RV, la Virtualité Augmentée superpose des éléments de l'environnement physique sur le monde simulé. Le plus souvent, des caméras enregistrent l'environnement physique de l'utilisateur. Le flux vidéo est analysé et filtré pour ne garder que certains éléments choisis, qui seront ensuite superposés à l'environnement virtuel. Dans ce cas, l'utilisateur porte un casque électronique afin de percevoir le résultat. Ce principe est illustré à la Figure 1.4.

Cette approche est encore relativement peu explorée dans la littérature par rapport à la Réalité Virtuelle et la Réalité Augmentée. Certains travaux comme ceux de Choo et al. utilisent les capacités de la Virtualité Augmentée pour étudier l'accessibilité d'applications mobiles [41]. Dans leur étude, l'utilisateur est plongé dans une scène virtuelle. On y voit un téléphone virtuel sur lequel tourne une application mobile. En parallèle, le participant manipule un téléphone physique. Les images des doigts du participant sont superposés en temps réel au téléphone virtuel. D'un côté, l'environnement virtuel permet de simuler facilement un handicap visuel en déformant ou floutant la scène. De l'autre côté, se servir simultanément d'un téléphone physique permet de bénéficier d'un retour haptique [41]. Ce dispositif est illustré à la Figure 1.5.

Nous pouvons également citer l'exemple des travaux de Regenbrecht et al., qui

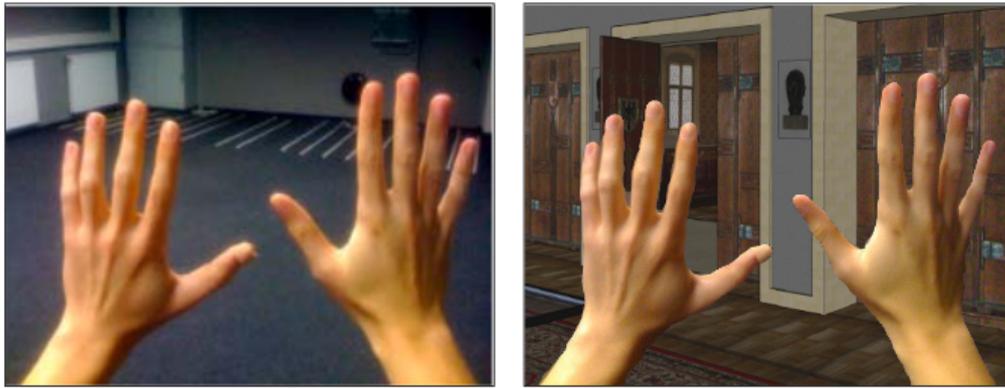


FIGURE 1.4 – Exemple de scène en Virtualité Augmentée. L'utilisateur perçoit un environnement virtuel comme en RV, mais des éléments de l'environnement réel (ici, les mains et bras de l'utilisateur) sont superposés au virtuel. Images issues de [34].

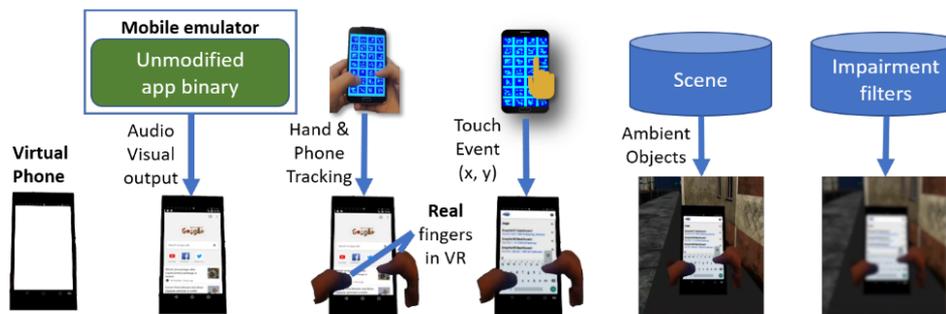


FIGURE 1.5 – Le système *Empath-D* utilisé par Choo et al. [41]. La scène virtuelle est augmentée avec le flux vidéo des doigts de l'utilisateur. Images issues de [41].

ont proposé un système de vidéoconférence en VA [171]. Les participants sont plongés dans une salle virtuelle mais voient les autres membres de la réunion en temps réel grâce au flux vidéo de leur webcam.

Réalité Remixée

Lindlbauer et al. ont récemment proposé le terme de *Réalité Remixée* pour désigner un autre aspect de la notion d'environnement mixte [127]. Avec la Réalité Remixée, l'utilisateur perçoit une reconstruction 3D en temps réel de son environnement physique grâce à des caméras. Par exemple, il peut porter un casque occultant sa vision (comme en RV) et ne percevoir que le retour vidéo des caméras de ce casque. L'objectif est de tirer parti de ce flux vidéo qui peut être retouché numériquement pour modifier l'environnement perçu par l'utilisateur. Un tel équipement est qualifié de dispositif vidéo (*Video see-through*). Plutôt que d'ajouter des éléments réels à une scène virtuelle comme en Virtualité Augmentée, le but est ici de modifier la perception de l'environnement réel. Ce concept de Réalité Remixée est postérieur aux travaux de Milgram et al. et par conséquent n'apparaît pas au sein du Continuum. Néanmoins, nous pouvons la placer à gauche de la VA puisqu'elle a un degré de virtualité plus faible que celle-ci (Figure 1.2).

Lindlbauer et al. ont conçu un système où l'utilisateur voit la salle dans la-

quelle il se trouve grâce à plusieurs caméras externes [127]. Plusieurs types de modifications de l'environnement réel sont proposées. Tout d'abord, des modifications spatiales permettent de déplacer, déformer, dupliquer ou de changer l'échelle d'un objet de l'environnement réel. De manière similaire, des altérations graphiques telles que changer la couleur, la luminosité ou la texture d'un objet sont possibles [127]. Par ailleurs, il est également possible de faire évoluer le point de vue dynamiquement via des téléportations ou des mouvements libres (i.e. impossible à effectuer en temps normal, comme voler dans la pièce). Enfin, les auteurs proposent également des modifications temporelles comme mettre en pause le flux vidéo, ou enregistrer une séquence pour la revoir plus tard à une vitesse différente [127]. Une illustration de ce système de Réalité Remixée est disponible à la figure 1.6. Pour un aperçu plus complet des possibilités évoquées ci-dessus, nous vous recommandons de visionner les vidéos liées à ces travaux ([https://dl-acm-org.ins2i.bib.cnrs.fr/doi/10.1145/3173574.3173703](https://dl.acm.org/ins2i.bib.cnrs.fr/doi/10.1145/3173574.3173703)).

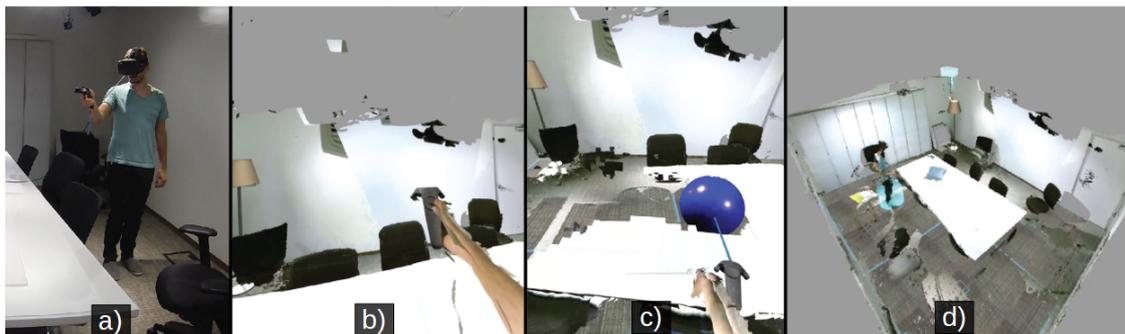


FIGURE 1.6 – Le système de Réalité Remixée proposé par Lindlebauer et al. [127]. a) Utilisateur portant un casque. b) Environnement 3D perçu via le flux vidéo du casque. c) Environnement remixé : suppression de la table. d) Environnement remixé : modification du point de vue. Images issues de [127].

Réalité Augmentée

Enfin, la dernière partie du Continuum à détailler est la Réalité Augmentée (RA). L'objectif est ici d'ajouter des éléments virtuels à l'environnement réel. A première vue, le concept de RA paraît relativement simple. Des objets virtuels, parfois appelés hologrammes, sont superposés au monde physique afin de créer un environnement mixte. Les objets virtuels peuvent être "attachés", ancrés à des objets physiques dans le but de créer des objets mixtes. En 1997, Azuma a proposé une synthèse de l'état de l'art en RA [13]. Selon l'auteur, un système de RA doit remplir trois critères : combiner des éléments virtuels et réels (C1) dans un environnement en trois dimensions (C2) et permettre d'interagir en temps réel avec cet environnement (C3) [13]. Un exemple de scène en réalité augmentée est présenté à la Figure 1.7.

Caractéristiques des systèmes de RA Les trois critères utilisés par Azuma excluent plusieurs systèmes. Ainsi, une salle de cinéma projetant des films en 3D n'est pas considérée comme un système de RA car elle ne respecte pas C3. En revanche, vérifier que C1 est respecté s'avère parfois plus délicat. La notion de combinaison d'éléments virtuels et réels est assez vague et sujette à interprétation. Un film en



FIGURE 1.7 – Exemple de scène en réalité augmentée. a) Scène originale : environnement réel. b) Scène augmentée avec des objets virtuels superposés à l'environnement réel.

3D est une projection d'images sur un écran physique 2D, mais crée une illusion de profondeur dans la scène. Est-ce suffisant pour valider le critère C1 ? De même, si les augmentations virtuelles sont toutes en 2D et sans réel lien spatial avec les objets du monde physique, est-il pertinent de considérer qu'il s'agit d'un environnement mixte en 3D (C2) ? De telles questions soulignent les contours flous de cette technologie. Deux points de divergence ressortent particulièrement : l'intégration spatiale (*spatial registration*) et le degré d'interaction avec l'environnement.

A partir de quand peut-on dire qu'un objet virtuel augmente l'environnement ? Contrairement à ce que l'on pourrait croire, la réponse à cette question n'est pas des plus évidentes. Dans le but d'illustrer cette notion, nous prendrons ici l'exemple d'une exposition en réalité augmentée illustré à la Figure 1.8.

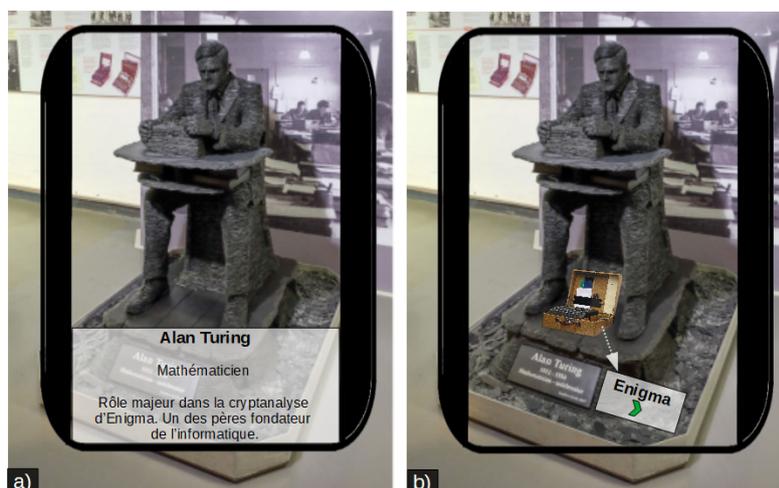


FIGURE 1.8 – Illustration du concept d'intégration spatiale des objets virtuels. a) Un panneau virtuel non intégré spatialement. b) Exemple d'objet virtuel intégré spatialement : la représentation virtuelle d'Enigma, "posée" entre les pieds de la statue, tient compte de l'environnement physique. Image modifiée issue de https://commons.wikimedia.org/wiki/File:Alan_Turing_by_Stephen_Kettle_2007.jpg.

Un premier point de vue consiste à dire qu'à partir du moment où l'objet virtuel est contextuel et visible, alors il augmente l'environnement réel. Imaginons qu'un visiteur arrive devant une statue d'Alan Turing. Lorsqu'il oriente la tablette vers l'œuvre d'art, un panneau explicatif apparaît en bas de l'écran et renseigne le visiteur sur la vie du célèbre scientifique, comme illustré à la Figure 1.8a. Ces informations sont liées à l'œuvre d'art qui intéresse actuellement le visiteur : il s'agit d'une augmentation contextuelle. Néanmoins, ce panneau virtuel n'est pas intégré spatialement à l'environnement. Ce n'est qu'un objet en 2D qui n'a pas de réelle position spatiale : il est simplement placé quelque part sur l'écran de la tablette, et aurait très bien pu apparaître plus haut ou sur un côté. On peut considérer que cela reste une augmentation du réel puisque la scène mixte obtenue n'est visible qu'au travers de la tablette. Néanmoins, certains chercheurs et industriels estiment que le contenu virtuel doit être intégré spatialement à l'environnement afin de pouvoir parler de Réalité Augmentée [187].

Intéressons-nous à présent au second cas (Figure 1.8b). Cette fois, quand le visiteur oriente la tablette vers la statue, il peut voir une représentation virtuelle de la machine Enigma, posée entre les pieds du mathématicien britannique. Le contenu virtuel est ici spatialement intégré à l'environnement : la position de l'objet virtuel 3D prend en compte les objets physiques aux alentours.

Un élément central de cette question de l'intégration spatiale est celle du repère utilisé. Dans notre premier cas, la position du panneau textuel est définie par rapport à la tablette ("en bas de l'écran"). Le repère n'est donc pas celui du monde physique, mais uniquement celui du système d'affichage. Si le visiteur déplace la tablette tout en restant à proximité de la statue, le panneau restera au même endroit sur l'écran. Dans le second cas, au contraire, l'objet virtuel est intégré spatialement : sa position a un sens par rapport à l'environnement physique. Si le visiteur déplace la tablette, l'objet virtuel ne "suivra" pas le mouvement, mais restera affiché à sa place entre les pieds de la statue. Le repère de cet objet virtuel est donc bien l'environnement réel.

Un second point de divergence au sujet de la définition de la RA vient du degré d'interaction avec l'environnement mixte. Pour une partie de la communauté, augmenter les perceptions de l'utilisateur suffit pour parler de RA [187]. Dans ce cas, la visualisation d'un environnement mixte serait alors suffisante, même si aucune interaction avec le système n'est possible. Au contraire, d'autres auteurs considèrent qu'un système de RA doit permettre à l'utilisateur d'interagir avec son environnement, ce qui rejoint la définition d'Azuma (C3) [13]. Différents cas sont possibles : 1) interagir avec les objets purement physiques (non augmentés), 2) avec des objets mixtes (objets physiques augmentés) voire 3) également avec des objets purement virtuels.

Méthode d'affichage des objets virtuels Il existe plusieurs méthodes pour afficher des objets virtuels superposés à un environnement réel en RA. La première est d'utiliser un système de projection externe pour augmenter la scène [170]. L'utilisateur peut alors visualiser l'environnement mixte sans avoir à porter sur lui un dispositif électronique. Cette méthode d'affichage est aussi appelée RA spatiale (*Spatial Augmented Reality*). Un aperçu de cette méthode est présenté à la Figure 1.9b.

Une seconde méthode pour visualiser l'environnement mixte passe par un dispositif mobile comme un téléphone portable ou une tablette [138, 162]. Ces équipements disposent en effet d'une caméra et de multiples capteurs (gyroscope, accéléromètre, magnétomètre, etc). Cela permet de les utiliser comme des dispositifs mobiles de RA : l'utilisateur peut les tenir avec une ou deux mains et visualiser l'environnement mixte au travers de l'écran du dispositif. En anglais, ces méthodes sont qualifiées de *handed AR* et de *mobile AR*. L'avantage de ce type d'affichage est lié à leur mobilité, indispensable dans certains domaines d'application comme la maintenance industrielle [162]. En revanche, les limites de la RA mobile incluent la taille limitée de l'écran (et donc du champ de vision augmenté) ainsi que la fatigue physique engendrée par le maintien du dispositif mobile dans les airs. Un aperçu de cette méthode est présenté à la Figure 1.9c.

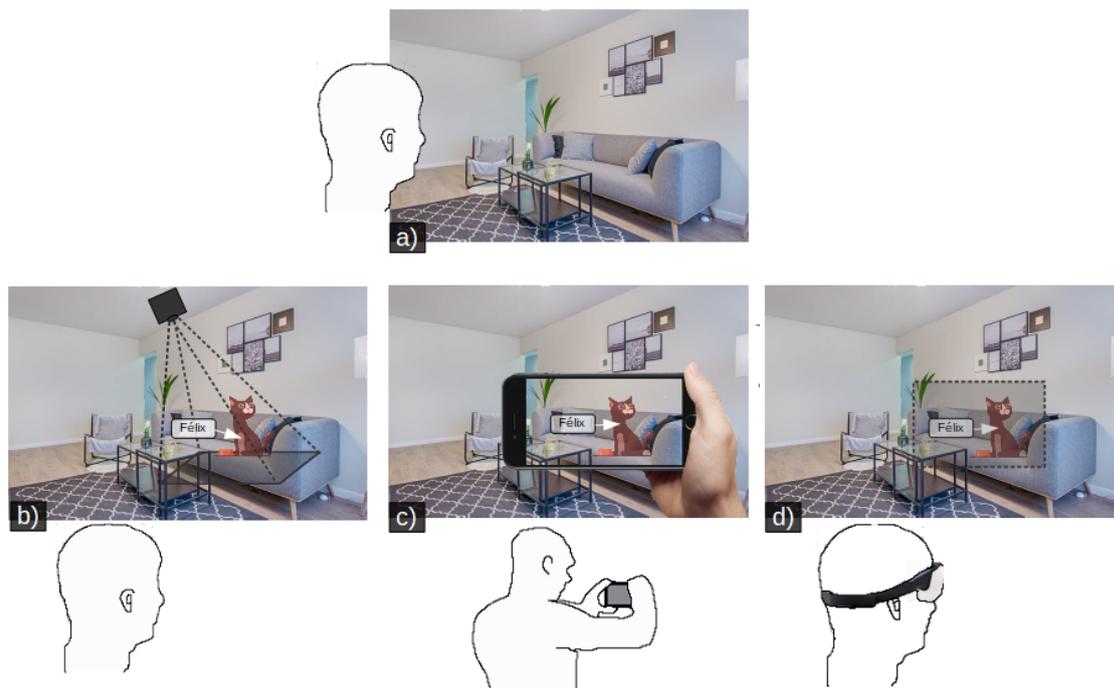


FIGURE 1.9 – Les différentes modalités possibles pour afficher du contenu virtuel. a) Scène non augmentée. b) RA par projection. b) RA mobile avec un téléphone portable tenu à la main. c) RA avec un casque.

Une troisième manière de percevoir un environnement mixte est basée sur le port de lunettes ou d'un casque électronique. En anglais, on parle alors de *Head-Mounted Display*, ou *HMD*. Cette méthode est illustrée à la Figure 1.9d. Il existe deux types d'équipement dans cette catégorie : 1) les dispositifs vidéo, que nous avons déjà abordé avec la Réalité Remixée et 2) les dispositifs optiques (*Optical see-through*). Les dispositifs optiques sont basés sur la transparence : l'utilisateur voit l'environnement qui l'entoure comme s'il portait une paire de lunettes classiques. Par un jeu de miroirs, le contenu virtuel est projeté au niveau des verres ou de la visière transparente du casque pour donner l'illusion que les objets virtuels sont superposés à l'environnement.

Enfin, il existe également des systèmes hybrides qui combinent ces différentes méthodes. Par exemple, Berard et Louis ont proposé un système de RA projective

centré autour d'une sphère de polystyrène [26]. Assis sur une chaise, l'utilisateur tient la sphère entre ses mains et voit un objet 3D "à l'intérieur" de celle-ci. La sphère est augmentée à l'aide d'un projecteur au plafond : il s'agit donc de RA spatial. La sphère peut être tournée afin d'examiner l'objet sous différents angles. Un tel système se rapproche donc aussi de la RA mobile, où l'utilisateur contrôle le champ de vision augmenté avec ses mains. Ce dispositif se rapproche également de la RA avec casque puisque l'utilisateur doit porter des lunettes stéréoscopiques. Enfin, ces lunettes comportent différents marqueurs optiques traqués par un ensemble de caméras externes afin de déterminer l'emplacement de la tête de l'utilisateur [26], comme le ferait un casque de RA.

Un autre exemple d'approche hybride combine projection et dispositif tenu à la main. Certains travaux se basent en effet sur des projecteurs tenus à la main [110]. Dans de tels cas, l'utilisateur contrôle le champ de vision augmenté à la main, indépendamment de la position de sa tête.

Domaines d'application L'essor des smartphones et tablettes a également permis au grand public de bénéficier d'applications de RA. Outre le divertissement, la RA ouvre des portes pour l'essayage virtuel : vêtements, produits de beauté [57] ou encore pour tester de nouveaux meubles dans son salon [56]. Pour l'instant, l'offre de casques de RA à destination des particuliers est encore limitée. Néanmoins, l'engouement pour ce type de technologie pourrait aboutir à une démocratisation future de ce type d'équipement, à l'image des casques de RV aujourd'hui.

La RA peut aussi bénéficier à de nombreux domaines d'applications professionnels. Ainsi, elle est en passe de devenir un outil précieux pour l'assemblage, la maintenance et la réparation d'équipements industriels [13, 154]. Fournir en temps réel des instructions ou de l'aide contextuelle directement sur une machine complexe, accéder facilement à une documentation particulière, travailler en coopération avec un autre technicien distant. Les applications possibles sont multiples pour aider les travaux et l'entraînement des opérateurs industriels [154]. Le domaine militaire étudie également cette technologie [89], notamment pour l'entraînement à la prise de décision stratégiques [135]. De même, certains secteurs comme l'aviation (militaire ou civile) [74] et l'automobile [58] s'intéressent de près aux habitacles augmentés. Par ailleurs, la médecine est également un domaine qui bénéficie grandement des possibilités offertes par la RA, comme détaillé au Chapitre 2.

Réalité Diminuée Voisine de la RA, la Réalité Diminuée a pour objectif de filtrer ou remplacer certains éléments de l'environnement réel [148]. Cela permet par exemple de masquer un élément gênant de la scène ou de voir au travers d'un obstacle [111, 169]. Ce principe est illustré à la Figure 1.10. Pour des raisons techniques, la Réalité Diminuée se base essentiellement sur des dispositifs vidéo plutôt que sur des dispositifs optiques.

Première définition de la Réalité Mixte

Réalité Virtuelle, Virtualité Augmentée, Réalité Remixée, Augmentée ou Diminuée... En précisant le sens de chacun de ces termes, nous venons de balayer l'ensemble du Continuum de Milgram et al. [146]. Selon les auteurs, la RM se situe entre les deux extrémités du Continuum. Tout élément compris entre les bornes du

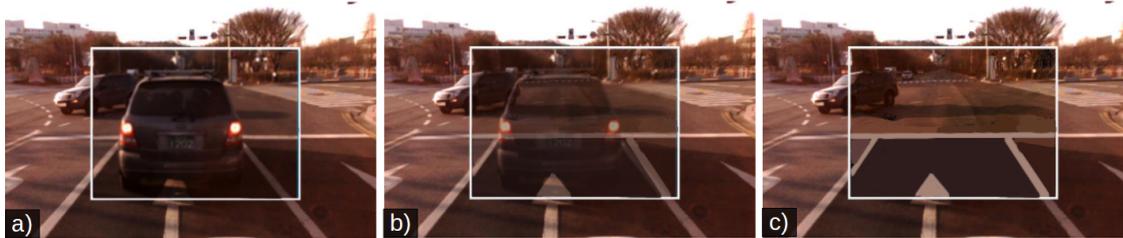


FIGURE 1.10 – Exemple de Réalité Diminuée. a) Un véhicule masque une partie de la route. b) Vue diminuée pour voir au travers du véhicule. c) Vue diminuée où le véhicule est complètement masqué. Images modifiées de [127].

Continuum Réalité-Virtualité ferait donc partie de la RM. Autrement dit, la RM serait l'union de chacune des approches que nous venons d'aborder (Réalité Augmentée, Remixée, etc). Ainsi, la RM se caractériserait par la présence d'objets du monde réel et du monde virtuel au sein d'un même environnement. La Figure 1.11 illustre cette définition. La définition de la Réalité Mixte proposée par Milgram et

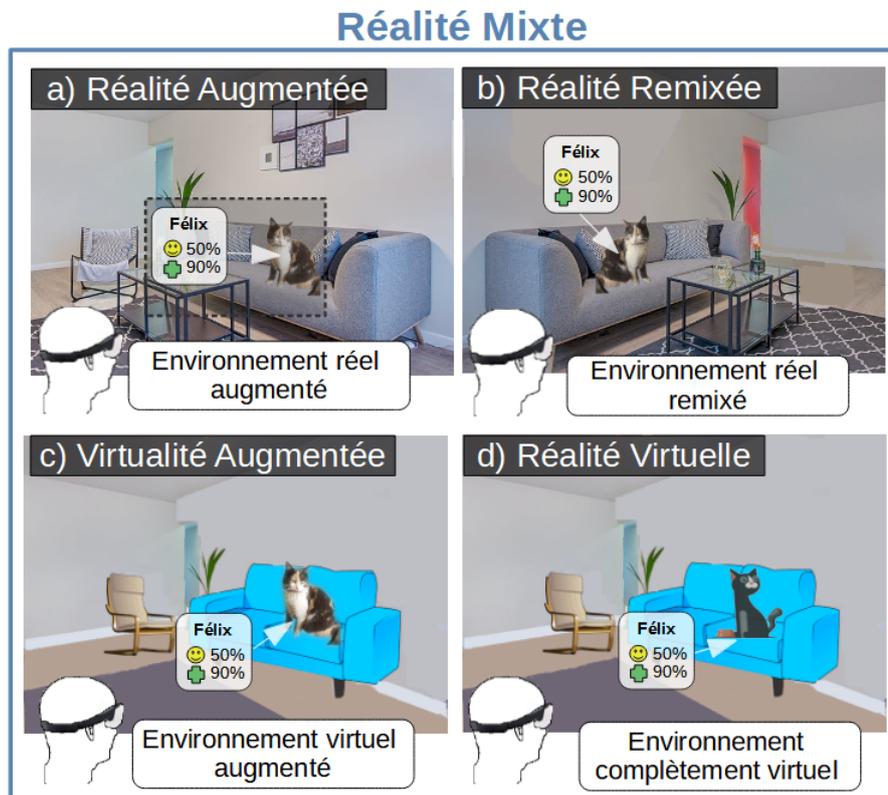


FIGURE 1.11 – Les différentes approches possibles mêlant environnement réel et virtuel. Selon Milgram et al., l'union de ces approches forme la Réalité Mixte [146].

al. est probablement la plus couramment utilisée à l'heure actuelle au sein de la communauté en IHM [187]. L'un des principaux avantages de cette définition est sa simplicité : tout ce qui est compris entre les extrémités du Continuum fait partie de la Réalité Mixte. Milgram et al. se basent avant tout sur l'aspect visuel et l'affichage de l'environnement mixte. Ainsi, la taxonomie proposée inclue par exemple des axes comme la fidélité de la reproduction et du sentiment d'immersion (*Extent*

of *Presence Metaphor*).

1.1.2 Définition basée sur l'interaction

Plutôt que de se concentrer sur des caractéristiques visuelles, Dubois et al. ont présenté une classification centrée sur l'interaction [62]. Les auteurs utilisent notamment deux éléments pour établir cette classification : la cible de la tâche et le type d'augmentation.

Le premier élément renvoie au type d'environnement ciblé et à l'intention de l'utilisateur. Si l'utilisateur cherche avant tout à affecter le monde physique lors de son interaction avec le système, alors les auteurs considèrent qu'il s'agit d'une interaction en RA [62]. Dans ce cas, le but de l'environnement mixte est d'aider l'utilisateur (par exemple, un technicien) à accomplir sa tâche dans le monde physique (par exemple, réparer une machine en panne). Au contraire, si l'utilisateur cherche avant tout à impacter l'environnement virtuel, Dubois et al. considèrent qu'il s'agit d'une interaction en Virtualité Augmentée. Au lieu d'un unique Continuum Réalité-Virtualité, les auteurs proposent deux continua parallèles 1) Virtualité Augmentée : du monde réel vers l'environnement virtuel ($R \rightarrow V$) et 2) Réalité Augmentée : de l'environnement virtuel vers le monde réel ($V \rightarrow R$) [62].

Par ailleurs, les auteurs distinguent également deux types d'augmentation : l'augmentation de l'exécution (interface en entrée) et l'augmentation de l'évaluation (interface en sortie). Selon les auteurs, un système de RA peut augmenter l'exécution d'une action effectuée dans le monde physique, par exemple en ouvrant automatiquement le capot d'une machine. Superposer de l'information virtuelle sur cette machine revient à augmenter l'évaluation du technicien chargé de la réparer. À l'inverse, en Virtualité Augmentée, les deux types d'augmentation se basent sur des éléments du monde physique dans le but d'affecter l'environnement virtuel. Dans le cadre des travaux de Choo et al. [41], le système augmente l'évaluation de la scène en superposant les doigts de l'utilisateur à la scène virtuelle (Figure 1.12). Un exemple possible d'augmentation de l'exécution serait de modifier le floutage de la scène virtuelle représentant le handicap visuel suite à un micro-geste des doigts sur le téléphone physique tenu par l'utilisateur. Dans ce cas, le système augmente l'interaction avec l'environnement virtuel par un geste sur un dispositif physique du monde réel.

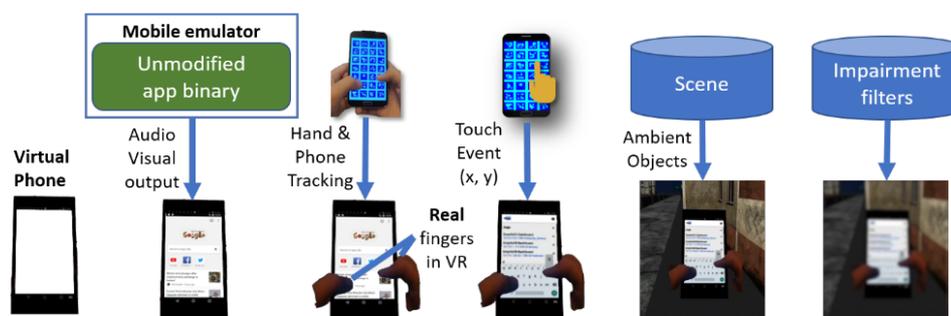


FIGURE 1.12 – Copie de la Figure 1.5. Le système *Empath-D* utilisé par Choo et al. [41]. La scène virtuelle est augmentée avec le flux vidéo des doigts de l'utilisateur. Image issues de [41].

Une telle classification propose une autre façon de différencier les concepts de RA et de Virtualité Augmentée. La notion plus récente de Réalité Remixée [127] peut également s'inscrire dans cette classification. Ainsi, les systèmes de Réalité Remixée augmentent l'évaluation de l'environnement mixte, par exemple via un changement de point de vue ou d'échelle. En fonction de l'intention de l'utilisateur (qui cible soit le monde physique soit l'environnement virtuel), ces systèmes se positionnent au sein du Continuum de Réalité Augmentée (tâche dans le monde physique) ou de celui de Virtualité Augmentée (tâche dans le monde virtuel). La Réalité Remixée apparaît alors comme une approche hybride qui peut s'inscrire en RA ou en Virtualité Augmentée.

1.1.3 Réalité Mixte dans le cadre de nos travaux

Nous venons de présenter deux définitions très différentes de la RM : l'une basée sur le Continuum Réalité Virtualité et l'autre sur l'interaction avec l'environnement mixte. Néanmoins, aucune définition ne fait pour l'instant consensus au sein de la communauté de l'IHM.

Limites d'une définition unique

La simplicité de la définition de RM proposée de Milgram et al. est à double tranchant. En effet, le centre du Continuum Réalité-Virtualité regroupe plusieurs visions différentes de l'environnement mixte, dont les frontières exactes sont parfois floues. Les exemples de confusion entre les différentes technologies mêlant environnement réel et virtuel sont légions. Au sein du grand public, la différence entre RA et RM est encore imprécise, voire inconnue. Cette confusion se reflète d'ailleurs dans certains articles de presse, où les concepts de RA, RV et RM sont allègrement mélangés [158]. Même la presse plus spécialisée n'échappe pas à ce phénomène. Récemment, l'Usine Digitale a publié un article sur un système de formation en RV pour les chirurgiens britanniques tout en l'illustrant avec un casque de RA [55]. Casque qui selon son constructeur (Microsoft) est en fait un casque de RM.

Plus surprenant, on retrouve une absence de consensus sur ces questions au sein des communautés scientifiques et industrielles. Les caractéristiques précises d'un système de Réalité Virtuelle sont globalement partagées, ce qui n'est pas le cas des systèmes au centre du Continuum. Ainsi, nous avons évoqué en Section 1.1.1 les divergences au sujet des caractéristiques de la RA. D'autres questions similaires se posent également pour les autres approches au centre du Continuum. Par exemple, la Réalité Diminuée peut être considérée de deux manières. D'un côté, elle se rapproche beaucoup de la RA, dont elle est le pendant inverse. D'un autre côté, il est possible de faire l'équivalent de la Réalité Diminuée (filtrer ou remplacer un objet) en Réalité Remixée. Les deux approches se basent en effet sur des dispositifs vidéos. Dans ce cas, la Réalité Diminuée pourrait être considérée comme un simple sous-ensemble de la Réalité Remixée. Définir la Réalité Mixte comme l'union des différentes approches existants sur le Continuum Réalité-Virtualité revient donc à englober plusieurs concepts aux frontières floues.

Les différentes définitions de ces concepts peuvent être incompatibles entre elles. Ainsi, la VA telle que définie par Milgram et al. [146] peut être considérée comme de la RA dans la classification de Dubois et al. [62] si l'utilisateur interagit avec le

système dans le but d'impacter le monde physique. Par ailleurs, ces deux définitions ne se basent pas sur la même granularité. Milgram et al. considèrent le système dans son ensemble en fonction de ses caractéristiques visuelles alors que Dubois et al. considèrent la tâche de l'utilisateur. Dans ce second cas, un système peut très bien appartenir à plusieurs catégories s'il comporte différents types de tâches (tâches $R \rightarrow V$ et $V \rightarrow R$) [62].

Par ailleurs, les travaux de Milgram et al. se concentrent uniquement sur les caractéristiques visuelles d'un environnement mixte [146]. Les auteurs proposent une taxonomie des systèmes de RM, mais chaque axe proposé ne tient compte que du canal visuel. Ainsi, le degré d'immersion dans la scène (*Extent of presence metaphor*) ne repose que sur les différents types d'affichage de l'information, d'un simple écran physique à un système parfait, c'est à dire indiscernable de la réalité (en terme de latence, champ de vision, etc). Les autres sens tels que l'ouïe, le toucher et l'odorat sont complètement absents de ces travaux.

Enfin, une définition basée sur le Continuum se concentre sur un seul et unique environnement mixte. La possibilité de travailler avec plusieurs utilisateurs ayant chacun leur environnement, c'est à dire d'avoir un système multi-environnements n'est donc pas considérée.

Co-existence de multiples définitions

Au delà des limites des définitions discutées ci-dessus, nous retrouvons dans la littérature différents emplois des termes "Réalité Mixte". Les récents travaux de Speicher et al. mettent en valeur les multiples définitions existantes de cette notion [187]. Certains auteurs caractérisent la Réalité Mixte comme une RA "plus forte", où davantage d'interactions avec les éléments de l'environnement mixte sont possibles [187]. La RA est alors réduite à la perception d'un environnement augmenté, avec lequel il est peu ou pas possible d'interagir.

Nous retrouvons également des articles où les deux appellations sont synonymes. Dans ce cas, certains auteurs choisissent de n'utiliser qu'un seul terme [12] ou d'évoquer les deux à la fois via des termes tels que "AR/RM" [123, 128]. Toujours dans la même idée que RA et RM désignent un concept similaire, certains chercheurs et industriels considèrent que la Réalité Mixte est avant tout un terme marketing pour parler de RA [187]. Force est de constater que la Réalité Mixte est encore à l'heure actuelle un concept difficile à cerner. Plusieurs définitions parfois contradictoires co-existent et de nombreuses questions restent ouvertes et sujettes à interprétation.

Positionnement de nos travaux

L'objectif de cette thèse n'est pas de proposer une définition de ce qu'est la Réalité Mixte. Notre travaux portent sur la conception et l'évaluation de techniques d'interaction applicables à la chirurgie augmentée. Pour cela, nous nous basons essentiellement sur des techniques de Réalité Augmentée (RA) avec un casque. Néanmoins, nous étudions également certaines approches inspirées de la Réalité Virtuelle (RV) : plusieurs techniques d'interaction que nous proposons incluent des transitions entre un mode RA et un mode RV. Enfin, certaines de nos contributions comme les techniques *Bring2Me* (Chapitre 4) peuvent potentiellement être utilisées aussi bien en RA qu'en RV.

Tous ces arguments nous ont poussé à préférer utiliser la notion de Réalité Mixte plutôt que de restreindre nos travaux à la RA. Il était donc important de commencer par préciser les notions sous-jacentes à ce terme ainsi que les discussions actuelles de la communauté sur le sujet. Dans le cadre de nos travaux, le concept de RM n'est donc pas équivalent à celui de RA. Nous préférons nous rapprocher d'une des définitions relevée par Speicher et al. dans leur revue de la littérature : la RM comme combinaison de RA et de RV [187]. Cette définition est celle qui correspond le plus à nos travaux puisque nous n'incluons pas les paradigmes comme la Réalité Remixée ou la Virtualité Augmentée dans nos recherches. De même, une définition de la RM comme un type de collaboration multi-utilisateur en RA et RV [187] pourrait être intéressante dans le cadre de la chirurgie augmentée, mais ne fait pas partie du champ de nos recherches. Nos contributions sont concentrées sur la combinaison de RA et de RV avec un seul chirurgien.

Dans cette première section, nous avons passé en revue les différentes approches mêlant environnement réel et environnement virtuel. Notre objectif était de clarifier chacun de ces concepts aux contours parfois encore nébuleux. Au travers des discussions actuelles, nous avons positionné nos travaux par rapport à l'état de l'art. Nos recherches portent donc sur des techniques d'interaction en Réalité Mixte (RM) pour la chirurgie augmentée. L'objectif de la prochaine section est d'établir un état de l'art des classes de techniques d'interaction existantes en RM.

1.2 Interaction en Réalité Mixte

La première section de ce chapitre nous a permis de préciser la notion de Réalité Mixte (RM) et les concepts qui y sont associés. L'objectif de cette seconde section est de développer cette notion grâce à une revue des classes de techniques d'interaction en RM que nous organisons en trois points. Dans un premier temps, nous nous intéressons aux techniques permettant d'interagir avec des objets virtuels proches. Nous étudions ensuite les techniques de pointage pour interagir avec du contenu virtuel distant. Enfin, nous présentons les travaux existants liés aux techniques d'interaction liées à des objets interactifs comme des menus et des palettes. Comme expliqué à la section précédente, nous mettons l'accent sur les techniques basées sur la Réalité Augmentée (RA), mais considérons aussi certaines approches basées sur la Réalité Virtuelle (RV).

1.2.1 Objets virtuels proches : Manipulation directe

Comment peut-on interagir avec un objet purement virtuel ? A l'heure actuelle, les interfaces graphiques "classiques" auxquelles le grand public est habitué sont limitées à un écran physique. En Interaction Homme-Machine, ces interfaces sont désignées sous le terme WIMP, pour *Windows, Icons, Menus and Pointing*. Que ce soit avec un clavier, une souris ou un écran tactile, nous travaillons dans des fenêtres, sélectionnons des icônes et des dossiers, nous naviguons dans des menus pour déclencher des commandes. Autrement dit, nous passons par des outils intermédiaires pour faire le lien entre le monde physique dans lequel nous évoluons et l'environnement numérique. Cela peut être un bouton permettant de dessiner un rectangle dans une application de dessin, ou la barre de défilement qui nous permet

de lire la page suivante de ce manuscrit. Dans sa théorie de l'interaction instrumentale Beaudouin-Lafon parle de médiateurs, d'instruments d'interaction qui servent de passerelle à double sens entre les deux mondes [22]. Au travers d'un dispositif physique (par exemple, une souris), l'utilisateur peut activer un instrument (la barre de défilement vertical) pour en prendre le contrôle et activer son effet.

La RM part d'un paradigme très différent de celui des interfaces WIMP. L'ensemble de l'environnement mixte et 3D peut être interactif. La RM ouvre ainsi la voie aux interfaces post-WIMP [22]. Dans cette première section, nous nous intéressons aux techniques d'interaction liées aux objets proches de l'utilisateur, c'est à dire à portée de bras.

Manipulation directe et *Virtual Hand*

La méthode la plus immédiate pour interagir avec un objet virtuel consiste à manipuler directement ledit objet. Cette idée a conduit à la métaphore de la main virtuelle (*Virtual Hand*) [166]. Cette technique est basée sur une manipulation directe et égocentrique de l'objet virtuel ciblé. Pour la mettre en œuvre, certains travaux se basent sur des équipements spécifiques au niveau des mains [35, 123]. Nous pouvons ainsi citer les travaux de Buchmann et al. où l'utilisateur peut se saisir d'immeubles virtuels grâce à un gant comportant plusieurs marqueurs [35]. L'avantage d'une telle approche est que la détection et le suivi des gestes de la main sont facilités par l'ajout de cet équipement. Utiliser des gants électroniques permet également de bénéficier d'un retour haptique, par exemple de vibrations lorsque l'utilisateur "entre en contact" avec un objet virtuel [123]. Néanmoins, une difficulté à prendre en compte est la gestion correcte de l'occultation visuelle entre la main et l'objet virtuel manipulé [35]. Le système proposé par les auteurs est illustré à la Figure 1.13.

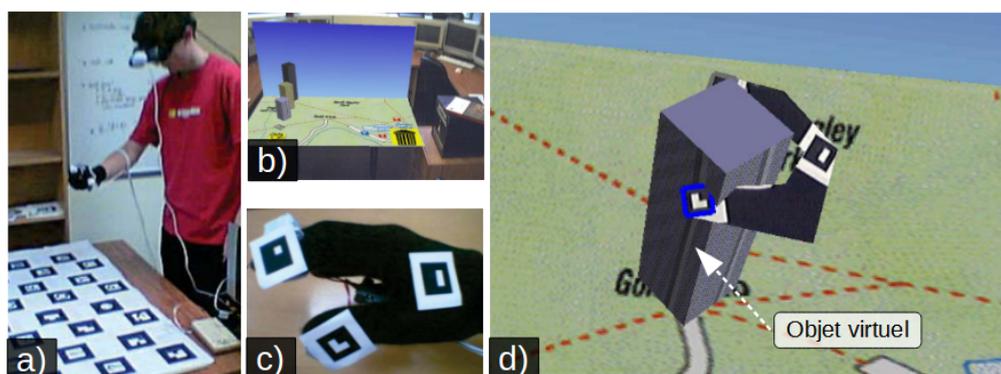


FIGURE 1.13 – Exemple de technique d'interaction basée sur un gant pour manipuler des objets virtuels. a) Utilisateur portant le gant. b) Espace de travail augmenté. c) Marqueurs sur le gant. d) Participant manipulant un objet virtuel à l'aide du gant. Images issues de [35].

Pour éviter à l'utilisateur de porter un équipement supplémentaire, il est également possible de détecter directement la position des mains et des doigts de l'utilisateur [25, 177]. Ainsi, Benko et al. ont proposé un système de RA projective basé sur une caméra de profondeur Kinect [25]. Le flux vidéo de cette caméra est analysé

pour y détecter la forme des mains, notamment grâce à la différence de couleur entre la peau et le reste de l'environnement. Il devient alors possible d'interagir avec des objets virtuels projetés dans l'espace de travail sans porter de dispositif additionnel au niveau des mains.

Une difficulté persiste néanmoins avec ce type d'approche : l'absence de retour haptique lors de l'interaction. L'objet virtuel étant par définition intangible, l'utilisateur perd la sensation de contact, l'impression de manipuler un objet réel. Des informations comme la forme de l'objet, sa texture ou encore son poids sont alors perdues, ce qui peut nuire à l'interaction. Comblé ce manque de retour haptique lors de l'interaction avec des objets virtuels est un sujet actif de recherche, en particulier en RV [1, 149, 220].

Une des solutions pour compenser cette absence de retour haptique consiste à utiliser des objets physiques pour interagir avec le système. Billinghamurst et al. ont étudié le concept de RA tangible en se basant sur les principes de conception des interfaces tangibles [31]. Selon les auteurs, *"Les interfaces de RA tangibles sont celles dans lesquelles 1) chaque objet virtuel est lié à un objet physique et 2) l'utilisateur interagit avec des objets virtuels en manipulant les objets physiques correspondants"* [31]. Les auteurs citent notamment la possibilité d'utiliser des techniques d'interaction 3D bimanuelles ou encore la collaboration entre plusieurs participants. Un exemple de système de RA tangible est présenté à la Figure 1.14.

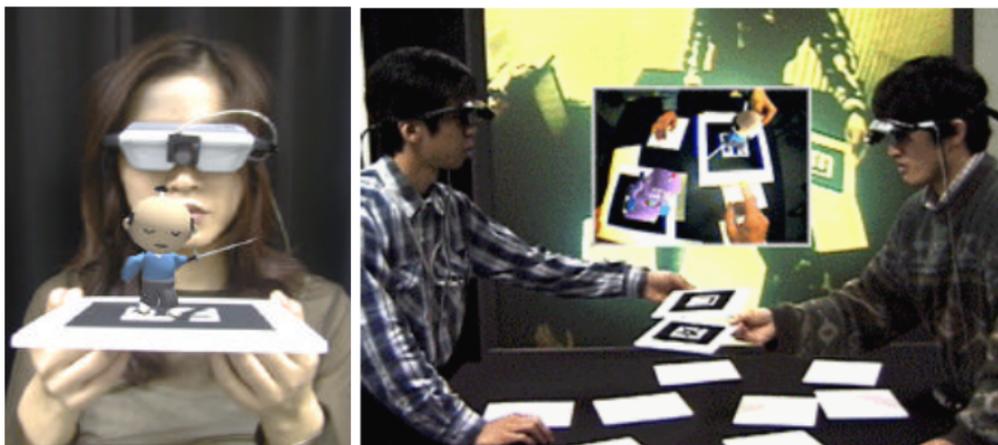


FIGURE 1.14 – Exemple de système de RA tangible basé sur des tuiles physiques. Images issues de [31].

De nombreux travaux sur la RA tangible se basent sur des tuiles physiques comme support de l'interaction. Chaque tuile comporte une marque ou un symbole qui lui est propre. Chacune de ces marques est prédéfinie à l'avance : ses caractéristiques précises sont donc connues. Le système traque ces marques grâce à des caméras afin de détecter la position et l'orientation des tuiles. Cette approche permet ainsi d'afficher correctement l'objet virtuel associé à chaque tuile. White et al. ont proposé les *Shake menus*, des menus radiaux virtuels centrés sur une tuile physique que l'utilisateur doit secouer pour activer le menu [207]. Certains auteurs combinent plusieurs tuiles pour créer un cube physique qui peut être manipulé par l'utilisateur pour déclencher différentes actions. Par exemple, Sin et al. ont proposé un outil pour apprendre l'astronomie en RA tangible [182]. L'utilisateur peut explorer un système

solitaire virtuel miniature grâce à différentes translations et rotations du cube tangible. Dans le même esprit, Lee et al. ont étudié différentes techniques d'interaction bimanuelles basées sur la combinaison de plusieurs de ces cubes tangibles [121]. Une illustration de leurs travaux est disponible à la Figure 1.15.

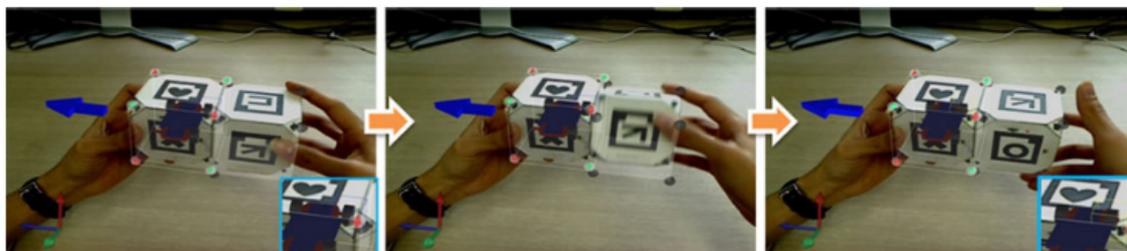


FIGURE 1.15 – Exemple d'interaction bimanuelle en RA tangible basé sur un assemblage de cubes physiques. Images issues de [121].

Les techniques d'interaction que nous venons de présenter permettent d'interagir avec du contenu virtuel. Néanmoins, elles sont limitées aux objets relativement proches de l'utilisateur. Dans la section suivante, nous étudions le cas de l'interaction avec des objets virtuels distants.

1.2.2 Objets virtuels distants : Pointage

La sélection d'objets à distance, ou pointage, est une tâche élémentaire particulièrement étudiée en IHM [119, 151, 185] car présent dans de nombreuses interactions. Dans un premier temps, nous nous intéressons à la modélisation de ces tâches de pointage. Nous explorons ensuite la littérature concernant le pointage en RM vers des cibles visibles, puis vers des cibles hors du champ de vision augmenté. Enfin, nous détaillons les techniques de pointage existantes en RM qui préservent le contexte visuel de l'utilisateur. Par ailleurs, nous nous concentrons sur les techniques de pointage liées à la RM avec un casque. Ce choix est motivé par notre domaine d'application principal, la chirurgie augmentée, où le chirurgien porte un casque de RM. Les techniques de pointage de RA sur dispositifs mobiles [200] sont donc exclues de cet état de l'art.

Modèle de performance d'une tâche de pointage

Le pointage est un geste *déictique* [112], c'est à dire qui montre un élément, qui pointe vers une cible. Ainsi, montrer du doigt une constellation dans un planétarium ou déplacer le curseur de la souris vers une icône à l'écran sont des actions déictiques. Intuitivement, ce type de geste peut faire émerger des remarques comme : "Il est plus dur de désigner une cible éloignée", ou encore "Plus on va vite, moins on est précis". En IHM, les performances d'un geste de pointage ont été modélisées par la loi de Fitts [72]. Cette loi décrit le compromis entre la vitesse et la précision d'un geste de pointage vers une cible. Depuis les travaux de MacKenzie [133], la formulation suivante est souvent adoptée :

$$MT = a + b * \log_2\left(\frac{D}{d} + 1\right) \quad (1.1)$$

Autrement dit, le temps MT pour atteindre le plus rapidement possible un cible dépend de la distance D à parcourir (entre la position initiale et le centre de la cible) et de la taille d de cette cible. Les coefficients empiriques a et b sont déterminés par régression linéaire en fonction de l'expérimentation. La loi de Fitts permet également de formuler la difficulté d'une tâche de pointage, ou Indice de Difficulté (ID) de la tâche :

$$ID = \log_2\left(\frac{D}{d} + 1\right) \quad (1.2)$$

Une tâche de pointage avec un ID faible comporte une cible relativement proche ou de grande taille. Au contraire, une tâche difficile avec une cible petite et éloignée aura un indice de difficulté élevé. Pour donner un ordre de grandeur, Soukoreff et al. suggèrent d'utiliser des tâches de pointage avec un ID compris en 2 (facile) et 8 (difficile) [185]. La loi de Fitts a été initialement formulée pour des tâches en une dimension, mais peut être adaptée pour des tâches en 2D [81]. Wingrave et Bowman ont ainsi étudié la loi de Fitts dans le cadre d'une tâche de pointage avec une technique de raycasting [209]. Les auteurs ont observé qu'utiliser la distance angulaire D_{ang} que doit parcourir la main et la taille visuelle de la cible T_{vis} permet de modéliser de manière précise le temps de sélection des cibles :

$$MT = a + b * \log_2\left(\frac{D_{ang}}{T_{vis}} + 1\right) \quad (1.3)$$

Meyer et al. se sont également servi de la loi de Fitts afin d'établir un modèle précis de prévision de performance pour le pointage en 2D [142]. Les auteurs s'intéressent particulièrement aux imprécisions du système moteur humain. Pour cela, ils émettent l'hypothèse que les mouvements rapides vers une cible ont toujours une part d'aléatoire, compensée au fur et à mesure par des sous-mouvements. Selon les auteurs, un geste de pointage se décompose donc en plusieurs temps : une phase balistique initiale pour approcher la cible (phase *préliminaire*), puis un ou plusieurs sous-mouvements qui corrigent la déviation aléatoire des mouvements précédents [142]. Un aperçu de ce modèle est présenté en Figure 1.16.

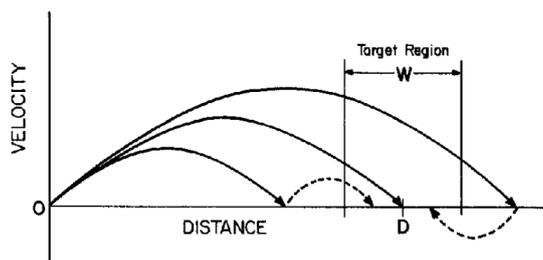


FIGURE 1.16 – Modèle des sous-mouvements stochastiques proposé par Meyer et al. illustré avec 3 exemples de tâches de pointage [142]. Les flèches continues représentent les phases préliminaires du pointage et les flèches en pointillé les sous-mouvements qui leur succèdent pour atteindre la cible.

Un autre modèle existant pour le pointage en 2D est la *Steering law* [2]. Cette loi modélise une trajectoire au sein d'un tunnel. L'idée est que ce mouvement possède une certaine "marge de manœuvre", une largeur variable mais bornée par les murs du tunnel. Pour les trajectoires simples, c'est à dire rectilignes, la formulation a des

similarités avec celle de la loi de Fitts :

$$MT = a + b * \frac{A}{W} \quad (1.4)$$

Comme auparavant, a et b sont des coefficients empiriques. Ici, A représente la longueur du chemin à parcourir et W la largeur du chemin, comme illustré à la Figure 1.17a. Le parallèle avec la distance D jusqu'au centre de la cible et la taille d de la cible est immédiat. Dans le cas général avec des trajectoires non-rectilignes (Figure 1.17b), la formulation de la steering law est néanmoins plus complexe. Elle dépend alors du chemin C (les bornes du tunnel) et de la largeur du chemin W à la position s . Laviola et al. suggèrent que la *Steering law* peut aussi être un modèle de performance adapté à la 3D [119].

$$MT = a + b \int_C \frac{ds}{W(s)} \quad (1.5)$$

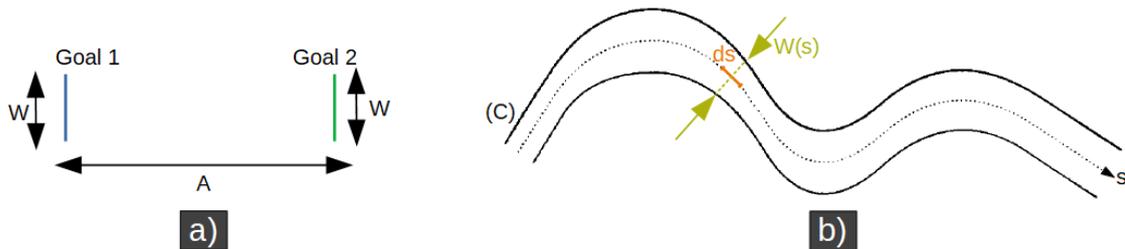


FIGURE 1.17 – Illustration des paramètres de la steering law. a) Pour une trajectoire rectiligne en 2D. b) Pour une trajectoire non-rectiligne en 2D. Images modifiées issues de [2].

Pour les tâches de pointage en 3D, de nombreux travaux ont également cherché à étendre la loi de Fitts [38, 80]. Ainsi, l'étude menée par Grossman et al. suggère que toutes les directions ne sont pas équivalentes pour le pointage en 3D : les déplacements latéraux semblent plus rapides que les changements de profondeur. Cha et al. ont proposé une version de Fitts 3D qui tient compte de la distance jusqu'à la cible, la taille de cette cible ainsi que la position 3D de celle-ci, qui inclut deux angles d'inclinaison. Les auteurs ont observé qu'un tel modèle basé sur un système de coordonnées sphériques était un meilleur prédicteur de performance que la formulation classique de la loi de Fitts.

Ces différentes lois peuvent guider la conception de techniques de pointage. Il convient néanmoins de noter que ces modèles de performance prennent comme hypothèse le fait que l'utilisateur pointe vers une cible visible.

Pointage vers une cible visible

Il existe deux approches principales pour sélectionner des cibles distantes et visibles dans un espace en trois dimensions : une extension de la main virtuelle et le raycasting [7]. Un aperçu de ces deux types de technique est disponible à la Figure 1.18.

La première approche est basée sur la métaphore de la main qui se saisit d'un objet physique, à l'instar de la technique *Virtual Hand* présentée en Section 1.2.1. L'utilisateur contrôle une main virtuelle qui reproduit les mouvements de sa propre main. La différence vient ici du fait qu'il est possible d'adapter la distance entre les deux types de mains, comme si le bras virtuel s'allongeait pour atteindre les cibles éloignées [166] (Figure 1.18a). Cette approche nécessite néanmoins davantage de mouvements que les techniques de raycasting, ce qui peut augmenter la fatigue physique de l'utilisateur.

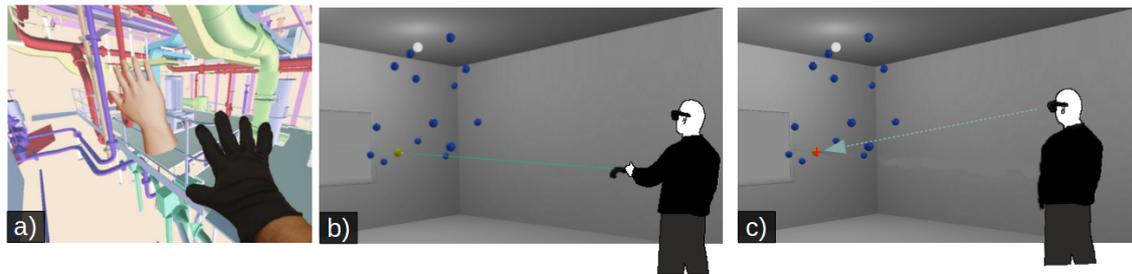


FIGURE 1.18 – Exemple de technique de pointage en RM. a) Main virtuelle, image issue de [7]. b) Raycasting à la main. c) Raycasting avec la tête. Images modifiées issues de [19]

La seconde approche, le raycasting, est basée sur la projection d'un rayon dans une direction donnée. L'utilisateur peut contrôler la direction de ce rayon afin de pointer vers une cible distante. Les techniques de raycasting ont été largement étudiées dans la littérature [66, 155, 209], et des travaux ont montré qu'elles étaient efficaces pour la sélection spatiale d'objets [155]. L'approche la plus classique consiste à utiliser un dispositif tenu à la main (i.e. un contrôleur, une télécommande) comme origine du rayon [19, 66], comme illustré à la Figure 1.18b. Par exemple, Baloup et al. ont proposé la technique *RayCursor* basée sur le contrôleur du HTC Vive [19], un casque de VR disponible dans le commerce [201]. Un curseur virtuel est affiché sur le rayon émit par le contrôleur. L'utilisateur peut utiliser le pavé tactile présent sur cette manette afin de déplacer le curseur le long de ce rayon. L'objectif de cette technique est de permettre à l'utilisateur d'affiner sa sélection afin de lever toute ambiguïté potentielle entre plusieurs objets proches de la cible [19]. Ainsi, il devient possible de sélectionner des cibles partiellement masquées par d'autres cibles.

Raycasting basé sur la tête Lorsque les mains de l'utilisateur ne sont pas disponibles, une alternative consiste à utiliser les mouvements de la tête pour interagir [10, 118, 155, 168]. L'une des particularités de la RM avec un casque est que le champ de vision augmentée est directement lié à la position de la tête. Pour voir une cible distante au centre de son champ de vision, l'utilisateur va donc devoir tourner la tête dans la direction correspondante. Il est donc possible d'exploiter ce mouvement déjà existant afin de sélectionner une cible. Ainsi, des techniques de raycasting existantes se basent sur les mouvements de la tête pour contrôler un curseur virtuel [118, 155, 168]. Jagacinski et Monk ont d'ailleurs montré en 1985 que la loi de Fitts (avec une formulation proche de l'Équation 1.1) est un bon indicateur de la performance d'une tâche de pointage basée sur la tête [103]. Ozacar et al. ont comparé cette approche

avec la technique *GyroWand*, où l'origine des rayons est fixée sur le menton de l'utilisateur et sa direction est contrôlée à l'aide d'un dispositif tenu à la main [155]. Les auteurs ont constaté que le curseur de tête surpassait *GyroWand* en terme de vitesse. Lee et al. suggèrent que le raycasting basé sur la main surpasse le raycasting basé sur la tête en termes de temps d'exécution et de précision, mais le second a l'avantage de ne pas nécessiter les mains des utilisateurs au moment de l'interaction [124]. Un aperçu de ce type de technique de raycasting est disponible à la Figure 1.18c.

Pointage basé sur le regard Toujours dans le cas où les mains ne sont pas disponibles, des travaux ont exploré l'utilisation du regard (*eye-gaze selection techniques* en anglais) pour pointer vers des cibles distantes en RM avec un casque [104, 118, 168]. Déplacer son point de focus visuel est très rapide et peut être réalisé sans bouger la tête. En revanche, les techniques basées sur le regard doivent faire face au problème bien connu du toucher de Midas [102]. Pointer avec le regard crée en effet des problèmes de sélections de cibles involontaires et donc non désirées lors de l'interaction. Pour compenser cet effet indésirable, Kyto et al. [118] recensent dans la littérature des approches basées sur le temps de survol de la cible (*dwell time*) [102], sur la poursuite de la cible [199] ou encore sur des gestes du regard [96].

Le pointage avec le regard a souvent été comparé avec le pointage basé sur la tête [104, 118]. Si les techniques basées sur le regard sont plus rapides, elle ont généralement été jugées moins confortables et moins précises que les techniques basées sur les mouvements de tête [104, 118]. Néanmoins, il est intéressant de noter que certains casques de RA récents comme l'Hololens 2 [145] ont choisi d'intégrer nativement un système de détection du regard.

Pointage vers une cible hors du champ de vision

La plupart des casques de RM actuels n'offrent qu'un champ de vision augmenté de petite taille. Ce problème est particulièrement marqué en RA, où le champ de vision des casques (autonomes, c'est à dire non relié en filaire à un ordinateur) ne dépasse pas les 52 ° en diagonale [145]. Les casques de VR actuels ont généralement un champ de vision proche de 110° (<https://benchmarks.ul.com/compare/best-vr-headsets>). L'environnement mixte est par conséquent bien plus large que le champ de vision augmenté. La Figure 1.19 illustre la taille de ces deux types de champs de vision par rapport aux capacités de l'œil humain.

Cette limitation implique que certaines des cibles peuvent ne pas être visibles. Dans leurs travaux sur le pointage direct hors écran avec un appareil mobile, Ens et al. ont signalé trois stratégies pour gérer les cibles hors écran [67] :

- **S1** : Mettre à l'échelle de l'espace de travail afin d'englober les cibles visées (zoom et dé-zoom)
- **S2** : Déplacer l'espace de travail pour faire apparaître les cibles dans la fenêtre d'affichage (défilement)
- **S3** : Déplacer la fenêtre vers les cibles (*peephole pointing*).

Cependant, l'adaptation de ces stratégies à la RM avec un casque soulève plusieurs problèmes car la fenêtre d'affichage (le champ de vision augmenté) est égocentrique et les cibles sont situées dans l'ensemble de l'environnement 3D. Les stratégies S1 et S2 briseraient le lien spatial entre les objets physiques et leurs augmentations

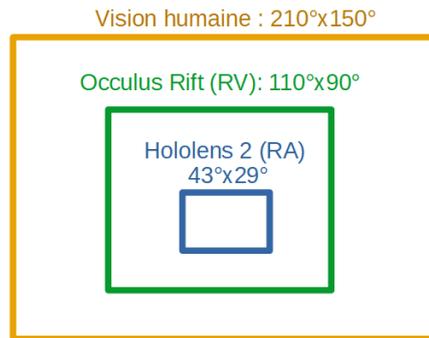


FIGURE 1.19 – Comparaison du champ de vision d’un casque de RA et d’un casque de VR classiques par rapport au champ de vision humain à l’œil nu.

virtuelles, cassant ainsi la métaphore de l’objet mixte. De plus, la mise à l’échelle de l’espace de travail (S1) rendrait également les cibles trop petites (dé-zoom trop important).

Le déplacement de la fenêtre vers les cibles (S3) peut être effectué en RM avec un casque en tournant la tête vers les emplacements cibles. Cette situation se rapproche fortement du *peephole pointing*, où une fenêtre virtuelle est déplacée pour révéler les différents pans d’un grand espace de travail [110]. Une illustration de cette stratégie pour l’interaction en 2D est disponible à la Figure 1.20.

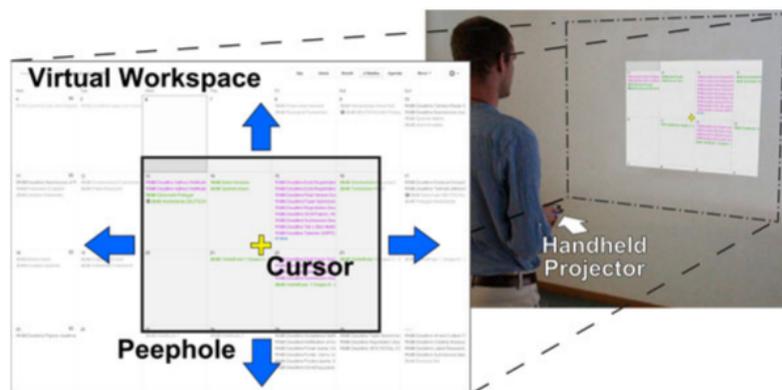


FIGURE 1.20 – Exemple de peephole pointing avec un projecteur tenu à la main. Image issue de [110].

En RM comme avec une technique de *peephole pointing*, l’utilisateur peut ne pas connaître la position exacte de la cible. Dans ce cas, une phase d’exploration de l’environnement est nécessaire avant tout pointage. Cao et al. ont étudié ce problème et proposé un modèle de performance pour le peephole pointing [36]. Ce modèle, adapté de la loi de Fitts, tient compte de cette absence potentielle de savoir *a priori* sur la position des cibles. Ces travaux sont néanmoins dédiés à une tâche de pointage 1D sur un écran [36]. Ens et al. ont confirmé les résultats de ces travaux au sein d’une CAVE (*Cave Augmented Virtual Environments*) dans laquelle le champ de vision de l’utilisateur était réduit. L’objectif était de simuler le champ de vision limité d’un casque de RM [66].

Tourner la tête pour voir une cible avant de la sélectionner est néanmoins à double tranchant. En déplaçant le champ de vision augmenté, l’utilisateur perd

également le contexte visuel initial. Cela pose problème dans des contextes applicatifs où l'attention visuelle est principalement dédiée à une zone particulière liée à la tâche principale, par exemple pour les pilotes de ligne [176].

Pointage n'utilisant pas le regard

Des travaux ont cherché à pallier ce problème de perte de contexte visuel en RM. Pour cela, une solution consiste à pointer vers une cible sans la regarder, comme illustré à la Figure 1.21. Li et al. ont proposé un système permettant de déclencher des raccourcis sans regarder en orientant un appareil mobile vers des icônes virtuelles placées autour de l'utilisateur [125]. Les sélections font appel à la mémoire kinesthésique et spatiale de l'utilisateur. Les auteurs ont observé que le système surpasse l'interface classique d'un appareil Nokia pour les tâches courantes sur un téléphone mobile [125].



FIGURE 1.21 – Un participant avec un casque de VR pointant vers un widget de sélection de couleur sans tourner la tête pour le regarder. Image issue de [213].

Plus récemment, Yan et al. ont étudié l'acquisition de cibles autour du corps en RV sans regarder [213]. Les auteurs ont mené plusieurs études pour évaluer l'amplitude des positions utilisables pour des tâches de pointage, la précision du contrôle de l'acquisition de la cible et ont finalement comparé la sélection avec ou sans l'utilisation du regard. Les sélections n'impliquant pas le regard se sont révélées plus rapides et moins fatigantes que les sélections nécessitant le regard tout en atteignant un niveau de précision satisfaisant [213] (de l'ordre de 92 % contre 98 % avec le regard). De plus, la plupart des participants ont préféré ne pas avoir à tourner la tête pour regarder la cible. De tels résultats sont encourageants pour les techniques d'interaction sans le regard, en particulier dans le cas de la RM où le champ de vision est réduit. De plus, pointer sans regarder peut bénéficier de la mémoire spatiale et de la proprioception. Des travaux comme ceux de Perrault et al. suggèrent que la mémoire spatiale et la création de liens sémantiques peuvent être de puissants outils pour permettre à l'utilisateur de retenir un grand nombre d'éléments [164]. Mémoriser un environnement mixte afin d'être capable de pointer vers l'un de ses éléments sans le regarder semble donc envisageable.

1.2.3 Sélection de commandes

Les techniques d'interaction que nous venons de décrire permettent de sélectionner un objet virtuel proche ou à distance. L'étape qui suit la sélection d'un objet virtuel est l'étape de sélection d'une commande. De plus, lorsque l'interaction en RM n'est pas dédiée à la manipulation d'un objet virtuel, elle consiste alors souvent à déclencher une commande générale à l'application.

Commandes gestuelles

L'une des approches les plus présentes dans la littérature pour déclencher une commande se base sur des gestes manuels [46, 52, 87, 123]. Comme pour la main virtuelle pour sélectionner un objet virtuel (Section 1.2.1), des travaux proposent d'utiliser des équipements spécifiques afin de détecter la position des mains et des doigts. Ainsi, Lee et al. ont proposé des techniques d'interaction basées sur des gants électroniques [123]. Les auteurs ont conçu un langage gestuel bimanuel permettant de déclencher différentes commandes pour manipuler des objets virtuels. Un signal sonore indique lorsqu'un geste valide est reconnu par le système. Un aperçu de ces travaux est disponible à la Figure 1.22.

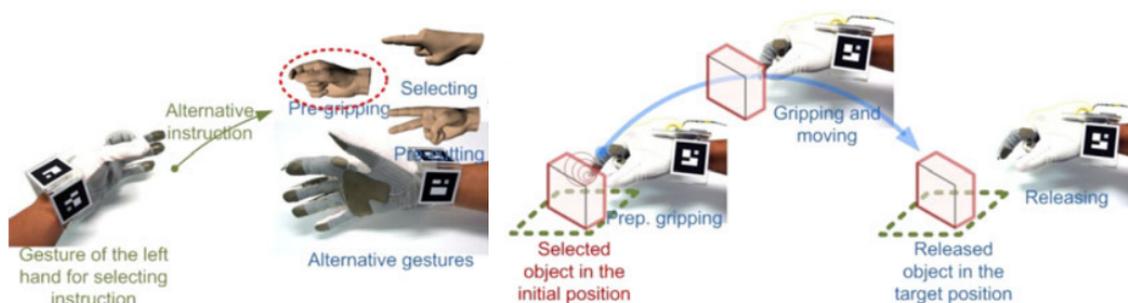


FIGURE 1.22 – Exemple d'une technique d'interaction gestuelle basée sur des gants électroniques. Image issue de [123].

Une autre catégorie de techniques d'interaction se base sur les gestes "mains libres" (*free-hand gestures*) [46, 52, 87]. Dans le même esprit que les travaux de Benko et al. [25] décrits en Section 1.2.1, un système de caméras permet de détecter en temps réel la position des mains nues. Ainsi, Colaço et al. ont conçu un capteur de gestes 3D basé sur une caméra RGB [46]. Ce capteur peut se fixer à un casque de RM. Les auteurs illustrent les capacités de leur capteurs en montrant plusieurs gestes et techniques d'interaction possibles, notamment avec des menus [46]. Il est intéressant de noter que certains casques de RA très récents comme le HoloLens 2 intègre désormais nativement le suivi des doigts de l'utilisateur [145].

Lorsque les mains de l'utilisateur ne sont pas disponibles pour interagir, la tête peut également être utilisée pour sélectionner des commandes [136, 214, 215]. Dans les travaux de Mardanbegi et al., les utilisateurs doivent ainsi effectuer différents gestes de la tête en fonction de la commande [136]. Cette approche pose le problème d'une fatigue physique accrue, notamment lors d'utilisation intensive et/ou prolongée.

Outre la fatigue physique, utiliser des gestes pour interagir présente deux autres inconvénients majeurs. Tout d'abord, la position de la tête et des mains doit être

correctement reconnue et interprétée par le système. Il est essentiel que le système soit robuste et puisse distinguer les faux positifs tels que les entrées involontaires [214]. Cela peut être délicat avec un casque de RM où le champ de vision augmenté est lié à la position de la tête. Cela signifie 1) que l'utilisateur doit s'entraîner à effectuer les gestes corrects et 2) que la tête et/ou les mains doivent être bien visibles au moment d'effectuer le geste. Si le système se base sur la ou les caméras intégrées au casque de RM, alors la zone de détection des gestes peut être relativement limitée. Ensuite, le second inconvénient vient du fait que les participants doivent apprendre la correspondance entre les gestes et les actions qui leur sont associées. Le nombre de gestes possibles peut certes être limité ou les métaphores être explicites (par exemple, mimer une paire de ciseaux pour l'action "Couper" dans [123]). Néanmoins, mémoriser les gestes et les actions associées augmente la charge cognitive des utilisateurs. Pour pallier ce problème, un rappel visuel de chaque correspondance peut être ajouté à l'environnement mixte. Une telle approche force cependant à tourner la tête pour consulter le mémo avant chaque interaction, ce qui interrompt la tâche et fait momentanément perdre de vue le contexte visuel initial.

Augmentation du corps de l'utilisateur Une des approches qui limite la fatigue physique consiste à directement augmenter le corps de l'utilisateur. Azai et al. ont par exemple proposé un menu affiché sur l'avant-bras de l'utilisateur [12]. La sélection d'un élément se fait par contact direct sur le bras augmenté avec l'autre main. Les auteurs proposent également d'autres techniques d'interaction, comme une glissade le long du bras (*slide interaction*) servant à révéler la sous-hiérarchie du menu [12]. De telles techniques bénéficient non seulement d'un retour haptique lors de l'interaction, mais également de la proprioception [194]. Cette faculté de situer son propre corps dans l'espace est précieuse car elle peut permettre d'interagir avec le système sans avoir besoin de regarder ce que l'on fait. Par exemple, un utilisateur expert avec le menu d'Azai et al. [12] saura que la commande pour copier un objet se trouve sur son poignet et pourra l'activer sans quitter des yeux sa zone de travail. En RM, cela évite à l'utilisateur de tourner sa tête et préserve donc le contexte visuel.

Ce problème de perte du contexte visuel initial peut aussi avoir lieu lors de l'interaction avec un menu. Si l'utilisateur n'a pas mémorisé l'agencement des items du menu ou la commande associée à l'item qui l'intéresse, il est alors forcé de tourner la tête pour le regarder. Une alternative consiste alors à considérer des *menus invoqués*. Ces menus n'ont pas de positions physiques permanentes dans l'environnement : ils apparaissent directement dans le champ de vision des utilisateurs lorsqu'ils sont appelés, et disparaissent après la sélection d'un élément ou lorsque les utilisateurs n'ont plus besoin d'interagir avec eux. Contrairement aux menus *persistants* qui ont une position stable et permanente dans l'environnement, aucune mémoire spatiale n'est impliquée lors de l'utilisation de menus temporaires. Par exemple, Datcu et al. ont proposé un système où un menu apparaît à l'emplacement de la main non dominante, comme illustré en Figure 1.23. Il devient alors possible de sélectionner des éléments du menu en les pointant avec la main dominante [52].

Les auteurs ont considéré un seul menu, mais différents gestes de la main pourraient être utilisés pour faire apparaître les différents menus disponibles. Le principal inconvénient de cette approche est que les utilisateurs doivent apprendre les gestes



FIGURE 1.23 – Exemple de menu invoqué apparaissant grâce à des gestes manuels. Image issue de [52].

et la correspondance des gestes avec les menus. Une telle interaction ne peut pas tirer parti de la mémoire spatiale car les menus n'ont pas de position spécifique dans l'environnement.

Interaction au pied L'interaction avec les pieds est une alternative possible pour éviter à l'utilisateur d'utiliser ses mains pour interagir. Dans le contexte chirurgical, certains systèmes comme l'OrthoPilot® choisissent d'intégrer une pédale chirurgicale [3], comme illustré en Figure 2.15. Le chirurgien peut utiliser ses pieds afin de valider une commande. Dans la classification des techniques d'interaction au pied proposée par Rovers et al., un tel dispositif est comparable à un simple interrupteur [175]. La pédale ne permet pas de contrôler un paramètre continu, mais uniquement d'actionner un bouton, par exemple pour passer à l'étape chirurgicale suivante [3].

Commandes vocales

Les commandes vocales ont l'avantage de ne pas solliciter les mains ni l'attention visuelle de l'utilisateur [45, 114]. En pratique, la voix se combine particulièrement bien avec d'autres modalités en entrée. L'un des travaux les plus célèbres sur le sujet est le "*Put that there*" de Bolt [32]. Cette étude explore la combinaison de gestes déictiques et d'entrées vocales pour interagir en 2D avec un écran distant. Le système analyse à la fois le discours pour comprendre l'action (par exemple, déplacer un objet) et les gestes de l'utilisateur qui complètent la commande (par exemple, un objet est d'abord désigné comme cible, puis l'utilisateur énonce la commande "déplacer" avant de pointer vers le nouvel emplacement désiré).

De nombreuses techniques d'interaction multimodale en Réalité Mixte impliquent d'utiliser la voix [98, 107, 205, 208]. Ainsi, Irawati et al. ont proposé un système combinant l'interaction gestuelle via des tuiles tactiles avec des commandes vocales [98]. L'utilisateur peut par exemple énoncer la commande pour sélectionner un meuble, déplacer la tuile correspondante jusqu'à une nouvelle position puis spécifier vocalement qu'il souhaite placer le meuble à cet endroit [98]. Le système utilisé par les auteurs est illustré à la Figure 1.24. Kaiser et al. explorent également des techniques d'interaction multimodale impliquant la voix et des gestes, mais se concentrent sur la gestion des ambiguïtés [107]. Les auteurs étudient comment éliminer les ambiguïtés individuelles de la parole et des gestes en combinant les deux approches grâce à

une fusion d'éléments statistiques et symboliques.

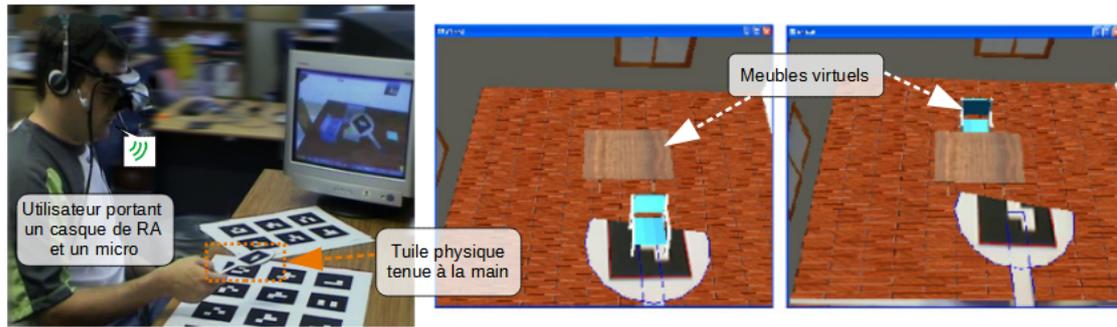


FIGURE 1.24 – Exemple de techniques d'interaction multimodale basées sur des gestes et des commandes vocales. Image issue de [98].

Conclusion

Dans ce chapitre, nous avons commencé par introduire la notion de RM et positionner nos travaux par rapport aux divergences liées à ce concept. Nous avons ensuite introduit les grandes familles de techniques d'interaction existantes en RM. Pour cela, nous avons passé en revue les techniques de manipulation directe d'objets virtuels, de pointage et de déclenchement de commandes. Le chapitre suivant se concentre sur le contexte applicatif de nos travaux : la chirurgie augmentée. Nous y annonçons les atouts de la RM pour le chirurgien ainsi que les techniques d'interaction utilisables en bloc opératoire. Ces dernières constituent un sous-ensemble de l'espace des possibles décrit dans ce chapitre.

Chapitre 2

Chirurgie Augmentée

Sommaire

2.1	Réalité Mixte en bloc opératoire	46
2.1.1	Chirurgie orthopédique classique	46
2.1.2	Apports de la Réalité Mixte pour la chirurgie	50
2.1.3	Limites de la réalité mixte en bloc opératoire	53
2.2	Interaction en bloc opératoire	59
2.2.1	Contraintes existantes en bloc opératoire	60
2.2.2	Techniques existantes utilisables en bloc opératoire	62
2.2.3	Bilan de l'interaction en bloc opératoire	65

Introduction

Nous venons de voir dans le chapitre précédent de nombreuses techniques permettant d'interagir avec du contenu virtuel en Réalité Mixte (RM). Nombre de ces techniques ont été étudiées lors d'expériences en laboratoire et ont pour but d'être généralisables à plusieurs contextes applicatifs. Nos travaux se concentrent sur un contexte applicatif précis : la chirurgie orthopédique augmentée. L'objectif de chapitre est de présenter les spécificités et les enjeux de ce domaine d'application. Nous commençons par présenter le contexte d'une chirurgie orthopédique classique, c'est à dire sans réalité mixte. Dans un second temps, nous détaillons les avantages et les limites de la réalité mixte pour le travail des chirurgiens.

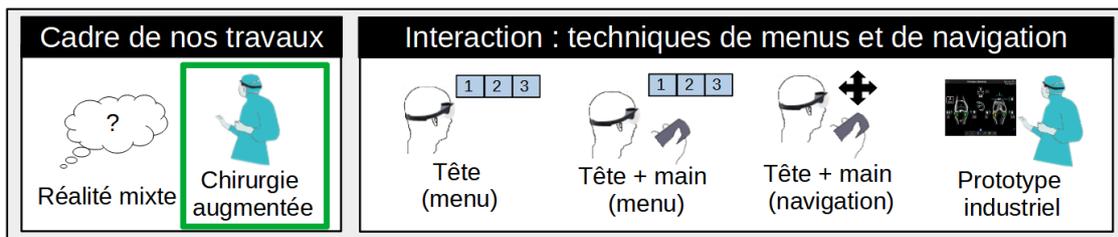


FIGURE 2.1 – Plan du mémoire (Figure B) - Focus sur le second chapitre.

2.1 Réalité Mixte en bloc opératoire

2.1.1 Chirurgie orthopédique classique

Principes généraux

Selon le dictionnaire Larousse Médical, la chirurgie orthopédique est "la spécialité chirurgicale traitant les maladies, les accidents et les déformations de l'appareil locomoteur (os, articulations, ligaments, tendons, muscles)" [139]. Si les interventions les plus connues sont sans doute liées au traitement des fractures, la chirurgie orthopédique concerne également les différentes interventions visant à soulager les articulations abîmées ainsi que les greffes osseuses.

La chirurgie orthopédique s'est aujourd'hui largement développée pour répondre aux besoins croissants de ce genre d'interventions. Vieillesse de la population, obésité, mouvements répétitifs et activité physique mal contrôlée sont autant de facteurs de risques pour l'organisme, et plus particulièrement pour les articulations. Les problèmes d'arthrose, c'est à dire d'usure anormale et de destruction du cartilage au niveau d'une articulation, touchaient en 2005 entre 9 et 10 millions de français [120]. L'arthrose est la seconde cause de consultation dans les pays développés après les maladies cardio-vasculaires. Elle est une cause majeure d'invalidité, surtout lorsque des articulations telles que le genou, la hanche ou la colonne vertébrale sont touchées [101, 120].

Pour soulager des douleurs ou suite à un traumatisme, il est possible de remplacer en partie ou totalement une articulation par une prothèse articulaire. Le terme médical consacré est celui d'*arthoplastie* partielle ou totale. Bien que les prothèses d'épaule et de coude existent, la plupart des interventions concernent la pose de prothèse de hanche et de genou ou des chirurgies du rachis.

Chirurgie assistée par ordinateur

Aux commencements de la chirurgie, le succès d'une opération reposait uniquement sur le savoir-faire et l'expertise du chirurgien. L'ensemble des gestes techniques devaient par conséquent être effectués "à la main", sur la seule base de l'observation de l'anatomie du patient. Les progrès technologiques à la fin des années 80 ont néanmoins permis le développement de la Chirurgie Assistée par Ordinateur (CAO), notamment pour la neurochirurgie [23].

Certains systèmes de CAO peuvent accompagner l'ensemble du processus chirurgical. Avant l'opération, leur objectif est de faciliter la planification des gestes chirurgicaux. Les résultats des examens d'imagerie médicale peuvent être analysés afin d'aider le chirurgien à comprendre les spécificités de l'anatomie du patient. Le *TSolution One*® est un exemple de solution commerciale comportant de tels aspects pour la chirurgie orthopédique (<https://thinksurgical.com/professionals/>). Par ailleurs, la CAO permet aussi aux chirurgiens de se former et de s'entraîner en amont avec des scénarios virtuels.

Deux types de systèmes de CAO peuvent assister l'équipe médicale au cours d'une opération : les systèmes *actifs* et les systèmes *passifs*. Les systèmes actifs tels que les robots chirurgicaux effectuent une partie des gestes techniques sur le corps du patient [195]. Il existe notamment des systèmes synergiques, où le chirurgien

et le robot coopèrent pour effectuer certaines procédures. De nombreux systèmes de télé-intervention entrent dans cette catégorie. Dans ce cas, le chirurgien peut travailler sur une station de contrôle éloignée du patient et pilote à distance des bras articulés à retour d'effort comportant différents outils. Pour la laparoscopie (chirurgie de l'abdomen), nous trouvons par exemple le robot da Vinci® illustré à la Figure 2.2. Pour la chirurgie orthopédique, un exemple de système synergique est le système Navio® de Smith and Nephew (<https://www.smith-nephew.com/professional/microsites/navio/navio-technology/product-overview/>).



FIGURE 2.2 – Exemple de télé-chirurgie avec le système da Vinci ®. ©2020 Intuitive Surgical, Inc.

En opposition à ces systèmes actifs de CAO, les systèmes passifs ne font que fournir de l'information au chirurgien, qui effectue lui-même l'ensemble des gestes chirurgicaux [62, 195]. Ces dispositifs suivent la position dans l'espace des outils ou de repères sur le corps du patient. Ils permettent au chirurgien d'accéder à des mesures, à des images médicales, ou encore d'effectuer des simulations.

L'entreprise Aesculap, partenaire de cette thèse CIFRE, a mis au point un système de CAO passif nommé OrthoPilot®[3]. Spécialement conçu pour la pose de prothèses de genou et de hanche, l'OrthoPilot® est un logiciel de navigation chirurgicale. Le terme *navigation* est différent du sens traditionnellement donné en IHM : il indique ici que le logiciel va guider le chirurgien lors de chaque étape de l'opération afin que la prothèse soit la mieux placée possible. La prothèse est alors qualifiée de prothèse naviguée.

Pose de prothèse naviguée avec l'OrthoPilot®

La fonctionnalité principale de l'OrthoPilot® est de simuler la pose de prothèse en se basant sur l'anatomie du patient. Le chirurgien peut ainsi tester différentes tailles et positions de prothèses, planifier la configuration qui semble la plus adaptée au patient, puis poser la prothèse en vérifiant au fur et à mesure qu'il respecte le plan établi. En plus des simulations, le logiciel aide également à positionner de façon optimale plusieurs outils chirurgicaux comme les guides de coupe. Enfin, il est possible d'enregistrer la trajectoire d'une articulation en mouvement. Plusieurs propriétés cinématiques peuvent en effet intéresser le chirurgien, qui peut alors s'assurer que la prothèse joue correctement son rôle lorsqu'elle est sollicitée.

Pour détecter la position du corps du patient en temps réel, l'OrthoPilot® comprend un chariot placé au sein du bloc opératoire. Comme illustré à la Figure 2.3a, cette structure comporte un bras articulé avec un système optique composé d'un émetteur et de deux caméras infra-rouge. Ces composants permettent de suivre en temps réel la position de marqueurs sphériques réfléchissants la lumière infra-rouge.

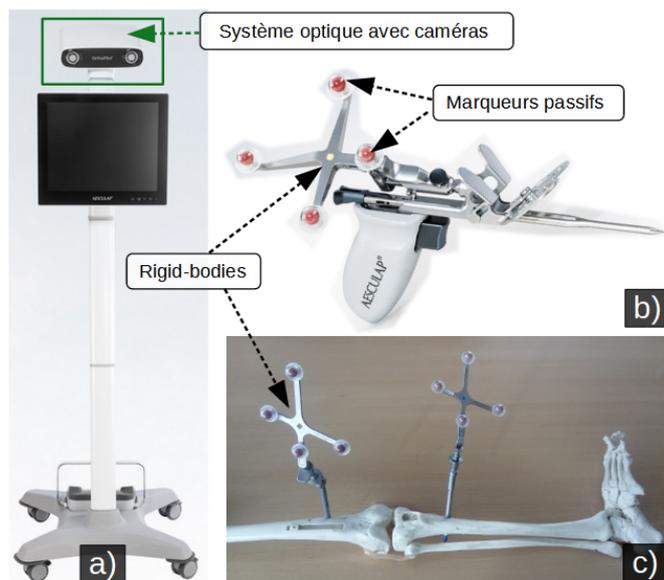


FIGURE 2.3 – Matériel lié à l'OrthoPilot®. a) Structure roulante avec système optique. b) Outil chirurgical comportant un rigid-body afin d'être détecté par les caméras. c) Rigid-bodies fixés sur le tibia et le fémur d'une maquette en plastique. Images originales fournies par Aesculap.

Ces marqueurs sont regroupés sur des branches métalliques appelées *rigid-bodies*. Chaque rigid-body a une configuration spatiale de marqueurs qui lui est propre, ce qui permet au système de l'identifier. Ces branches métalliques sont installées sur les outils chirurgicaux (Figure 2.3b) ou temporairement fixées sur le corps du patient (Figure 2.3c) afin d'être détectées par la caméra. Au début de l'opération, le chirurgien enregistre la position de certains repères anatomiques précis sur le corps du patient à l'aide du palpeur, l'un des outils suivis ressemblant à un tournevis. Cela permet au système d'initialiser les valeurs d'un modèle de l'anatomie du patient, qui est indispensable pour pouvoir ensuite simuler des positions de prothèses correctes.

En plus du système optique, la structure roulante de l'OrthoPilot® comporte également un ordinateur et un écran tactile où est affiché l'interface graphique du logiciel. Cette interface permet au chirurgien d'accéder à des informations médicales concernant le patient comme des données anatomiques et des mesures en temps réel. Le chirurgien peut également se servir de plusieurs menus pour déclencher des commandes telles que passer à l'étape suivante ou précédente ou enregistrer une capture d'écran (Figure 2.4a). Lors de certaines étapes de l'opération, le chirurgien est également amené à modifier des valeurs à l'écran (Figure 2.4b). Cela est notamment le cas lorsqu'il souhaite enregistrer son plan d'action afin de pouvoir comparer ce plan initial au résultat concret observé sur le patient une fois le geste chirurgical effectué.

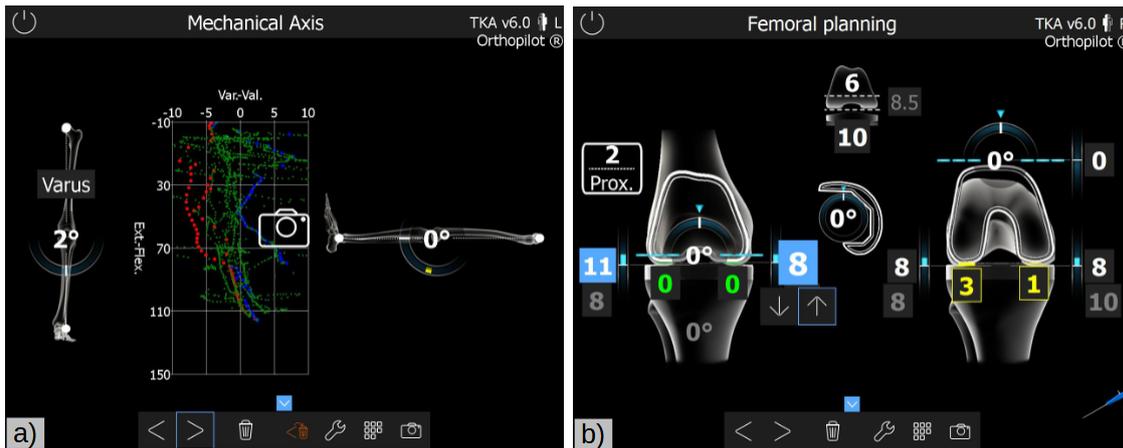


FIGURE 2.4 – Captures d'écran de l'OrthoPilot®. a) Informations médicales et un menu. b) Édition de valeurs pour définir la future position de la prothèse. Images originales fournies par Aesculap.

L'intérêt d'un tel système de navigation chirurgicale est multiple. Tout d'abord, l'OrthoPilot® permet au chirurgien de placer la prothèse avec une grande précision (de l'ordre du millimètre). Les mesures indiquées par le système viennent d'une détection en temps réel du corps du patient et sont donc d'autant plus précieuses pour le chirurgien que de simples radios prises en amont de l'opération. Cela permet également de se passer de coûteux systèmes d'imagerie tels que l'Imagerie à Résonance Magnétique (IRM). Le temps opératoire n'est pas forcément réduit, il est même souvent légèrement allongé. En revanche, l'OrthoPilot® permet de connaître immédiatement les conséquences des choix réalisés (gestes chirurgicaux effectués sur les ligaments par exemple). Ces informations sont importantes pour le chirurgien, qui peut davantage adapter la procédure au patient actuel. Une prothèse mieux posée permettra une meilleure récupération post-opératoire : le patient aura besoin de moins de soins (moins de douleur résiduelle ressentie), moins de rééducation et se rétablira plus rapidement. De plus, la prothèse s'usera moins vite et durera ainsi plus longtemps : le risque de ré-intervention est donc diminué.

Le chirurgien dispose de 3 méthodes pour interagir avec le système. La première méthode possible consiste à demander à un tiers d'utiliser l'écran tactile. Toucher l'écran tactile va néanmoins à l'encontre des consignes de protection contre les agents infectieux en bloc opératoire. La personne désignée devra recommencer un long processus de désinfection. Cela implique également que cette personne doit se rendre disponible pour assister le chirurgien, mettant temporairement de côté sa tâche initiale. Il s'agit donc là d'une méthode de dernier recours.

Une seconde approche consiste à utiliser une pédale orthopédique médicale. Ce dispositif placé sous la table d'opération est relié par câble à l'ordinateur. Contrairement aux bras et aux mains, les membres inférieurs sont en effet sujet à des règles de stérilité moins strictes. Le chirurgien peut donc se servir de ses pieds pour appuyer sur l'un des trois boutons de la pédale et ainsi interagir avec le système.

Enfin, la troisième approche permettant au chirurgien d'interagir avec l'OrthoPilot® est basée sur une technique de pointage à distance. La structure roulante avec l'écran étant éloignée de la table d'opération pour des raisons sanitaires, la

distance qui le sépare du chirurgien peut varier de un à trois mètres [82]. Grâce à ses outils traqués dans l'espace par la caméra, le chirurgien peut pointer vers les widgets affichés sur l'écran en utilisant une technique de raycasting (voir Section 1.2.2). Une illustration de cette approche est disponible à la Figure 2.5. Plusieurs mécanismes de confirmation de la sélection sont alors possibles : 1) une technique de temporisation (*dwell-time*), 2) un clic mécanique via une gâchette présente sur certains outils chirurgicaux tenus à la main ou 3) un appui sur la pédale pour confirmer la sélection.

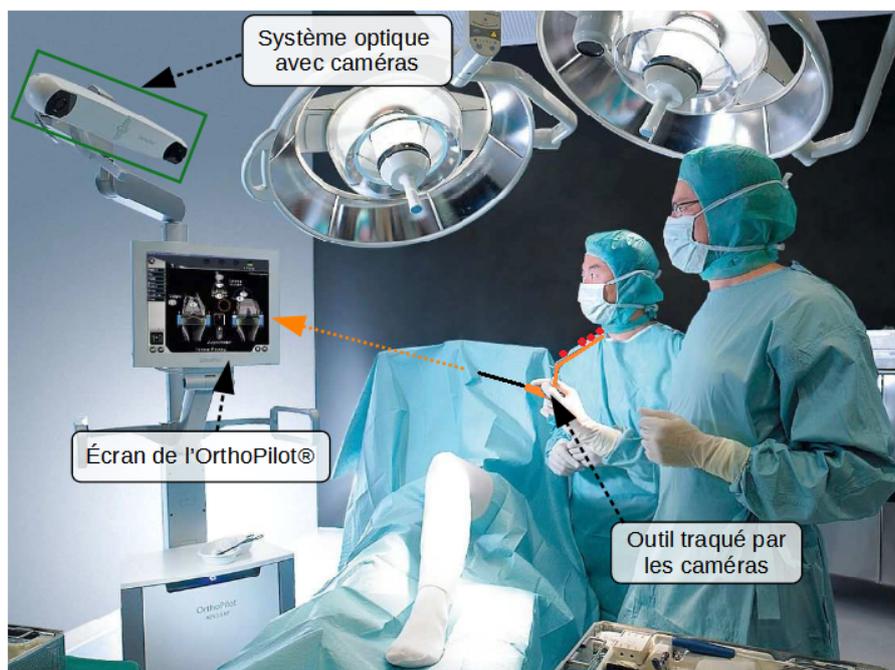


FIGURE 2.5 – Un chirurgien en train d’interagir avec l’OrthoPilot® par raycasting. Les couleurs de l’outil chirurgical tenu par le chirurgien ont été retouchées pour en améliorer la visibilité. Image originale fournie par Aesculap.

A l’heure actuelle, l’OrthoPilot® n’intègre aucune fonctionnalité ni technique d’interaction en RM.

2.1.2 Apports de la Réalité Mixte pour la chirurgie

Visualisation d’informations médicales en RA

Un apport très important de la Réalité Augmentée (RA) pour la chirurgie est la capacité de visualiser des informations médicales directement superposées sur le corps du patient. Plutôt que de dépendre de radiographies effectuées en amont ou d’un écran distant, les données médicales deviennent visible en temps réel là où elles sont le plus pertinentes. Il s’agit d’une aide précieuse pour mieux visualiser les structures anatomiques internes [108, 189]. Ainsi, Katic et al. ont conçu une solution en RA permettant de visualiser les parties vitales à l’intérieur d’un organe (par exemple, des vaisseaux sanguins) [108]. Le chirurgien a alors une meilleure appréciation des zones à éviter et limite ainsi les risques de complications. Un aperçu de ce système de RA est disponible à la Figure 2.6.

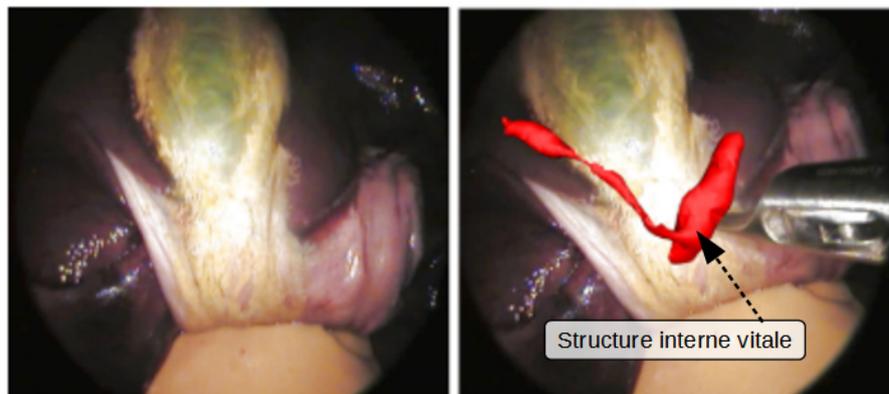


FIGURE 2.6 – Comparaison entre un organe non augmenté (gauche) et ce même organe en RA (droite) où une structure interne est mise en évidence. Images modifiées issues de [108].

Les données médicales affichées sur le corps du patient ne sont pas limitées à mettre en valeur des structures anatomiques existantes. Des informations plus abstraites peuvent également aider le chirurgien à mener à bien l'opération. Par exemple, en chirurgie orthopédique augmentée, Fallavollita et al. proposent d'afficher l'axe mécanique de la jambe [71]. Cette droite abstraite représente l'axe par lequel la majeure partie du poids du corps va passer au travers de la jambe. Il s'agit d'une donnée essentielle pour le succès d'une opération de chirurgie orthopédique : si cet axe est mal aligné avec la jambe, les articulations de tout le membre inférieur en souffriront et s'abîmeront rapidement. Être capable de voir cette information directement sur la jambe du patient est donc important pour le chirurgien.

Visualiser des informations directement superposées au patient permet de vérifier si les gestes chirurgicaux effectués jusque là ont l'effet escompté et de détecter plus rapidement d'éventuelles erreurs. Cela comprend également les erreurs liées au logiciel de navigation en lui-même. Par exemple, Badiali et al. ont proposé un système de RA pour la chirurgie maxillo-faciale [14]. Comme le montre la Figure 2.7, des croix virtuelles sont superposées sur plusieurs points-clés de la maxillaire du patient. Une simple inspection visuelle permet alors de vérifier la bonne calibration du système, ce qui n'est pas toujours immédiat avec un logiciel de navigation chirurgicale classique.

Enseignement et entraînement

Plusieurs études soulignent les bénéfices de la RM pour l'enseignement du geste chirurgical et l'entraînement des chirurgiens [4, 76, 198]. La formation d'internes est un processus long et complexe : en plus d'apprendre les procédures et gestes techniques, les apprentis doivent développer leur intuition afin de pouvoir prendre des décisions parfois radicales [198]. Or, le manque de temps réellement consacré à la formation et le manque de pratique de certains procédures ont un sérieux impact sur la qualité de l'apprentissage des internes [20, 147]. L'intérêt de nouvelles approches pour la formation des chirurgiens est donc identifié. Par exemple, en 2014, la première opération chirurgicale retransmise en direct à l'aide des Google Glass a été suivie par près de 13000 apprentis chirurgiens issus de 115 pays différents [206].

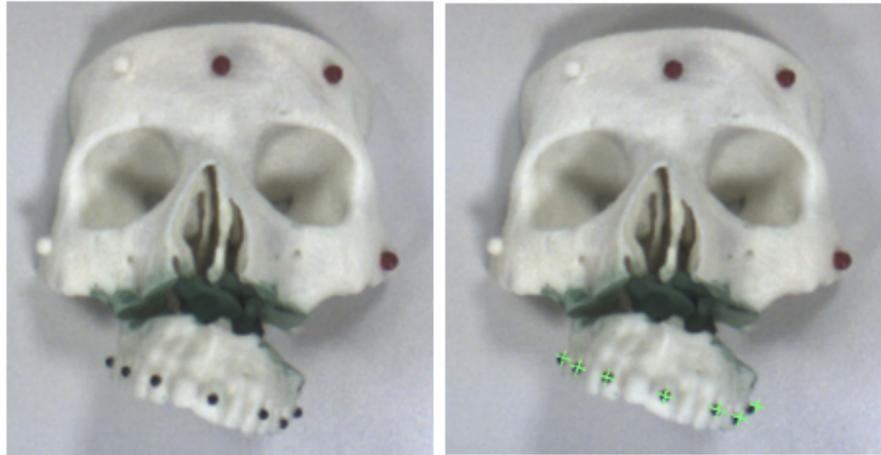


FIGURE 2.7 – Maxillaire augmentée en réalité mixte. Le positionnement correct des objets virtuels traduit l'exactitude de la calibration du système. Images issues de [14].

Travailler sur une réplique virtuelle et à l'échelle d'un corps humain permet de se familiariser avec une procédure chirurgicale sans inclure de risque pour les patients. Des scénarios rares ou particulièrement délicats peuvent être facilement simulés et répétés. De plus, la RM aide les apprentis à s'entraîner en autonomie, ce qui réduit la dépendance à un instructeur.

D'autres travaux se focalisent sur l'utilisation d'un mannequin d'entraînement. C'est par exemple le cas de l'étude de Ma et al. visant à enseigner aux apprentis chirurgiens la pose d'un cathéter au niveau de la veine jugulaire [130]. Comme illustré à la Figure 2.8, des mains virtuelles montrent le positionnement correct de l'instrument alors que des instructions sont décrites sous forme textuelle. Gasques Rodrigues et al. soulignent également que la RM permet d'augmenter le réalisme de la scène, par exemple en ajoutant des écoulements de sang virtuels [76].

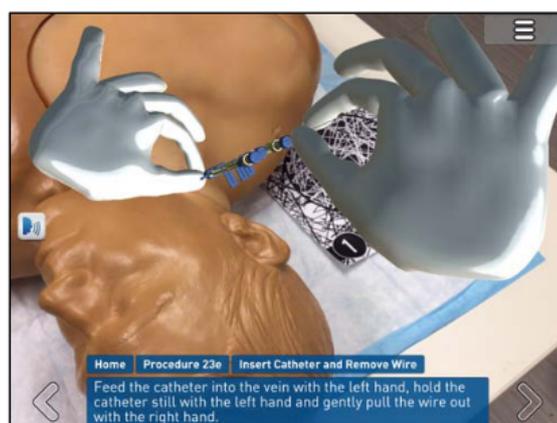


FIGURE 2.8 – Exemple de système de réalité mixte pour apprendre un geste chirurgical. Image issues de [130].

Utiliser un casque de RM permet également d'enregistrer des photos et vidéos de l'opération du point de vue du chirurgien. Une étude de Borgmann et al. sur l'utilisation de RA en urologie suggère que presque tous les chirurgiens ont estimé de telles

fonctionnalités utiles voir très utiles au cours de multiples opérations [33]. L'enregistrement de données peut servir à la fois pour l'enseignement et l'entraînement des internes, mais peut aussi être utilisé afin de créer de la documentation.

Coopération à distance

Outre l'enseignement et l'entraînement, la RM peut également faciliter la coopération à distance entre chirurgiens. Lors d'opérations délicates ou nécessitant une procédure particulière, il arrive qu'un chirurgien fasse appel à un confrère expert de cette procédure pour l'assister. Le mentor en question n'est néanmoins pas forcément sur place et peut donc avoir à guider à distance l'opération. C'est dans l'optique d'améliorer cette coopération à distance que des travaux comme ceux d'Andersen et al. utilisent la RA [6]. Le prototype proposé par les auteurs est basé sur une tablette maintenue au dessus du patient par un bras mécanique. Le flux vidéo de la caméra de cette tablette est envoyé au mentor distant, qui peut donc voir le champ opératoire en temps réel. Comme illustré à la Figure 2.9, le mentor peut ajouter des annotations sur le corps du patient grâce à sa propre tablette tactile. Ces annotations sont ensuite reflétées sur l'écran du chirurgien en train d'opérer afin de guider ses gestes [6].

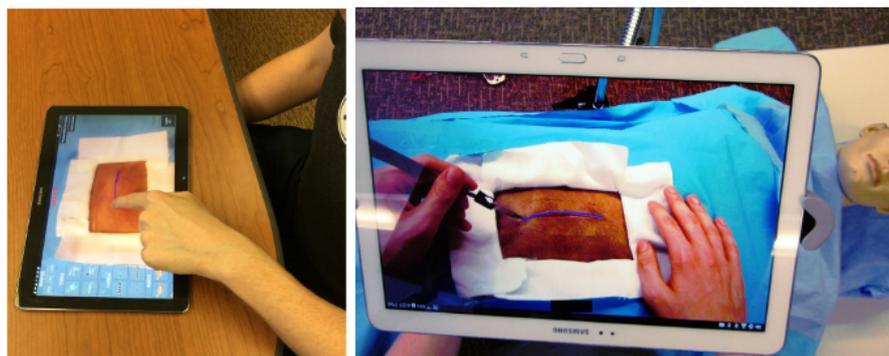


FIGURE 2.9 – Annotations virtuelles sur le corps du patient. Le mentor distant (à gauche) propose une incision. Le chirurgien sur place (à droite) voit la suggestion sur sa tablette et suit les instructions du mentor. Images issues de [6].

2.1.3 Limites de la réalité mixte en bloc opératoire

Nous venons de voir les multiples bénéfices que la RM peut apporter au travail des chirurgiens. Néanmoins, cette technologie présente également des contraintes et des limites. Les problèmes de perception liés à la RM constituent un sujet de recherche toujours actif à l'heure actuelle [116]. Dans les sections suivantes, nous détaillons en particulier deux facteurs clés pour le travail des chirurgiens : la taille du champ de vision augmenté et le besoin de précision.

Limites techniques

Champ de vision limité Les casques actuels de RM n'offrent qu'un champ de vision augmenté d'une taille limitée, loin de couvrir les capacités de la vision humaine [39, 66, 159]. Par exemple, le casque Hololens 2 [145] sorti en 2020 a un champ de

vision de 52° diagonalement, ce qui ne correspond qu'à une partie de la vision mi-périphérique [212]. Patterson et al. suggèrent que des champs de vision aussi limités ont un impact sur l'orientation spatiale et l'immersion [159]. Un champ de vision à taille limitée implique que la plupart de l'environnement mixte n'est pas visible (voir Figure 1.19). Tourner la tête pour déplacer le champ de vision augmenté est donc nécessaire pour rendre visibles les différents objets virtuels.

La conséquence majeure de cette limite est que l'utilisateur va perdre son contexte visuel initial en tournant la tête. Cela est particulièrement problématique dans des scénarios où l'utilisateur a besoin de concentrer son attention visuelle sur une partie précise de l'environnement. Pour les chirurgiens, se détourner du patient est contraignant car cela implique l'interruption de leur tâche chirurgicale. Devoir tourner la tête pour regarder une radio ou un menu virtuel éloigné du champ opératoire ajoute une charge cognitive non négligeable à des utilisateurs déjà très pris et concentrés [37]. Nous avons introduit à la Section 1.2.2 des techniques d'interaction qui permettent d'éviter ce changement de contexte visuel.

Manque de visibilité des hologrammes Plusieurs autres limitations techniques des casques de RM actuels sont à prendre en compte. Le chirurgien ayant besoin de beaucoup de lumière au niveau du patient, le champ opératoire est éclairé avec de puissantes lampes. Or, une trop grande exposition nuit à la visibilité des objets virtuels en RM. Il est possible sur de nombreux casques actuels de régler la luminosité des hologrammes. Néanmoins, cela peut augmenter la fatigue visuelle ressentie. Plusieurs critères techniques comme le contraste et la lisibilité du texte de plusieurs casques actuels ont été évalués dans des études comme celle de Qian et al [167].

Durée de fonctionnement et batterie Par ailleurs, Mitrasinovic et al. [147] citent également plusieurs limitations techniques comme la durée de la batterie et le temps de charge des casques actuels. La charge de ces équipements constitue une étape logistique supplémentaire au sein d'un processus pré-opératoire déjà complexe. De plus, une opération comme un remplacement de prothèse orthopédique peut prendre jusqu'à deux heures. Le dispositif doit donc autoriser un tel temps en fonctionnement normal, voir de fonctionnement intensif. En cas de manque de batterie, le chirurgien doit être capable de poursuivre et terminer l'opération sans RM.

Précision de la calibration du système et des gestes chirurgicaux

Plus encore que dans d'autres domaines d'application, la précision du système de RM est un facteur critique en chirurgie augmentée. Toute information virtuelle superposée au corps du patient doit être correctement positionnée à tout instant afin de ne pas fausser le jugement du chirurgien. Toute erreur peut avoir des conséquences graves pour le patient et tout contretemps augmente les risques infectieux. De plus, Cassel souligne l'importance du sentiment de contrôle pour le chirurgien [37]. Une opération est un processus long et complexe, dans lequel le chirurgien doit parfois prendre des décisions radicales en un temps très court. Tout élément perturbant sa concentration et sa sensation de maîtrise doit être évité. Une calibration imprécise du système de RM peut induire en erreur le chirurgien et introduire un doute sur les étapes de l'opération déjà réalisées. La précision du positionnement des objets

virtuels est donc cruciale. En particulier, la perception de la profondeur est un facteur clé pour le chirurgien.

Précision du geste chirurgical et conflits convergence-accommodation La RM soulève des questions de précision au niveau des gestes chirurgicaux. Prenons le cas d'un chirurgien voulant pratiquer une incision en RM. Même si le système est parfaitement calibré, il faut donc que le chirurgien soit capable de respecter le tracé virtuel superposé à la jambe du patient. En temps normal, une telle tâche est aisée puisque les chirurgiens ont une fine dextérité manuelle liée à leur pratique intensive. Cependant, une étude menée par Condino et al. suggère que pour une tâche manuelle, utiliser un casque de RM peut engendrer un manque de précision notable [47]. Dans leur expérimentation, les participants devaient tracer des lignes à la règle entre différents points d'une feuille de papier. Les auteurs ont comparé des conditions à l'œil nu, où les points étaient imprimés sur la feuille, avec des conditions où ces points étaient des objets virtuels affichés dans le casque Hololens [47] (Figure 2.10). Lorsque les participants ont utilisé le casque, Condino et al. ont observé des erreurs allant jusqu'à 5.9mm. Un manque de précision de cet ampleur au cours d'une tâche effectuée dans des conditions "idéales" (en terme de confort, de charge cognitive et de stress par rapport à une opération) soulève des questions pour la chirurgie augmentée. A titre de comparaison, l'OrthoPilot[®] permet de positionner une prothèse de genou ou de hanche avec une précision de l'ordre du millimètre.

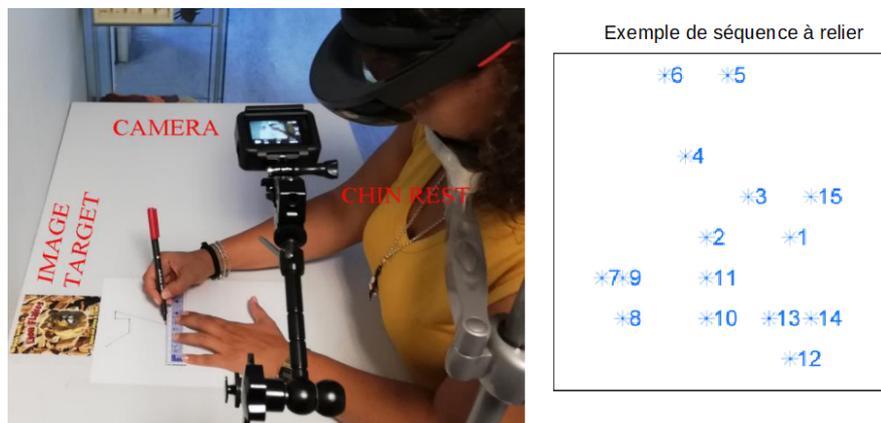


FIGURE 2.10 – Évaluation de la précision d'un système de réalité augmentée pour une tâche manuelle. Images issues de [47].

Cette imprécision peut s'expliquer par des problèmes optiques avec un casque basé sur la stéréoscopie. L'un des problèmes optiques les plus courants avec ce type de casque offrant une image légèrement différente pour chaque œil est appelé conflit accommodation-convergence (*vergence-accommodation conflict*) [116]. Pour expliquer ce phénomène, il est nécessaire de repartir du fonctionnement de la vision humaine.

Deux processus se combinent dans la vision humaine : l'accommodation et la convergence. L'accommodation est le processus qui vise à rendre nette l'image observée. Elle correspond à une adaptation automatique du cristallin des yeux en fonction de la distance entre l'utilisateur et l'objet observé. La convergence désigne

quant à elle la rotation des yeux qui doivent converger vers la cible [144]. Ainsi, pour observer un objet proche de soi, nos yeux vont converger l'un vers l'autre jusqu'à loucher si la cible est trop près. Si ces mises au point de la vision se font à des distances différentes, les indices de profondeur contradictoires vont entrer en conflit. Prenons l'exemple d'une information médicale virtuelle affichée à une profondeur différente : 5 cm derrière la jambe au lieu d'être superposée sur le tibia. En observant la scène, le chirurgien peut alors accommoder son regard au niveau de la jambe tout en faisant converger ses yeux vers l'information virtuelle, ou inversement. La scène observée deviendra alors en partie floue et inconfortable à regarder, ce qui risque d'augmenter la fatigue visuelle [92, 115]. La Figure 2.11 illustre le principe de ce conflit accommodation-convergence.

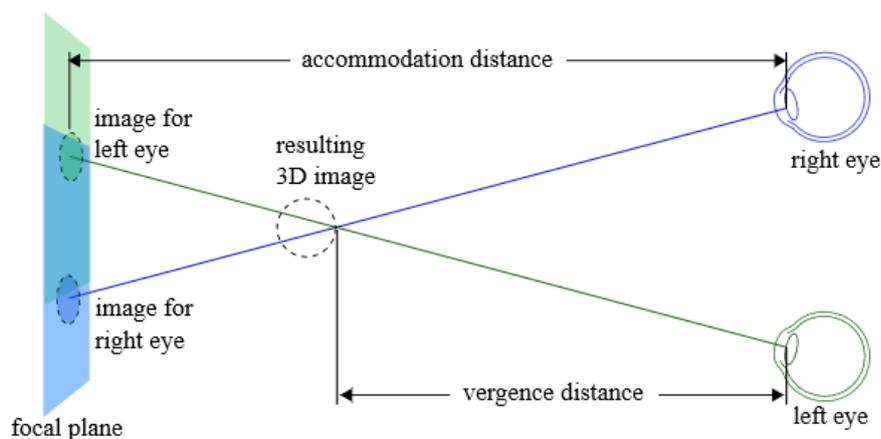


FIGURE 2.11 – Illustration du principe du conflit accommodation-convergence. La distance de convergence et d'accommodation sont différentes, ce qui perturbe la perception de la scène et brouille l'objet observé. Schéma issu de [115].

Ce phénomène est aujourd'hui bien connu et peut même avoir lieu à l'œil nu. Néanmoins, il est particulièrement courant en RA. Comme le détaillent Condino et al., la plupart des casques de RA actuels ont une distance focale proche de deux mètres [47]. En d'autres termes, la zone idéale pour voir les objets virtuels est située autour de deux mètres de l'utilisateur. A titre d'exemple, Microsoft conseille d'utiliser des hologrammes à au moins 1.25 mètre de distance de soi avec le casque HoloLens 2 [144]. En dessous de cette distance, l'écart entre les distances de convergence et d'accommodation risque d'être de plus en plus important et peut créer un inconfort visuel croissant. Cette limitation des casques actuels peut affecter les chirurgiens, qui travaillent à proximité immédiate du patient. L'influence des conflits accommodation-convergence sur la précision des gestes chirurgicaux reste à ce jour à étudier.

Perception de la profondeur La problématique des conflits accommodation-convergence en RM est liée à des questions de perception de la profondeur des objets virtuels. L'estimation des distances est souvent faussée lors de l'utilisation d'un casque [51, 192, 211]. Ce phénomène est lié à différents indices visuels de profondeurs qui entrent en conflit. La distance entre l'utilisateur et les objets virtuels joue également un rôle important dans la perception de la profondeur. Cutting propose une taxonomie d'indices visuels de profondeur découpant l'espace en trois zones

distinctes : la zone proche (moins de 1.5 mètre), la zone intermédiaire (entre 1.5 et 30 mètres environ) et la zone éloignée (au delà de 30 mètres) [51]. De nombreux travaux se concentrent sur la zone intermédiaire [88, 106, 192], mais certains étudient également la zone éloignée [211]. Des exemples de techniques de visualisation de la profondeur d'objets virtuels est disponible à la Figure 2.12.

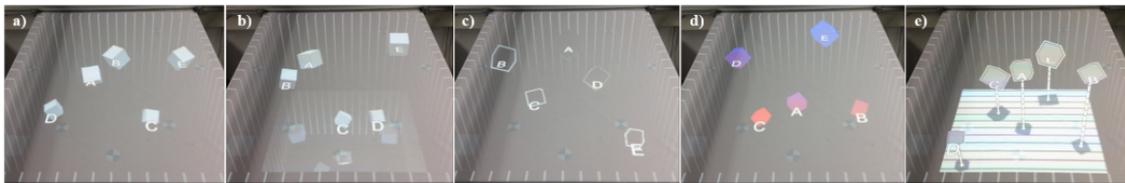


FIGURE 2.12 – Exemple de techniques de visualisation pour la perception de la profondeur en réalité mixte. De gauche à droite : *Phong Shading*, *Virtual Mirror*, *Depth-Encoding Silhouettes*, *Pseudo-Chromadepth* et *Supporting Lines*. Image issue de [88].

Dans le cadre de la chirurgie augmentée, nous nous intéressons essentiellement à la zone proche, qui correspond au champ opératoire. Dans leur étude expérimentale, Singh et al. ont évalué la perception de la profondeur d'objets virtuels proches (34-50 cm) via deux tâches [183]. La première tâche T1 consiste à déplacer un pointeur vers un objet cible (1 seul degré de liberté, déplacements vers l'avant ou l'arrière) jusqu'à ce que les deux éléments soient à la même distance (*perceptual matching*). La seconde tâche T2 suit le même principe, mais doit être effectuée avec le doigt au lieu d'un pointeur. Dans ce cas, le participant voit l'objet cible mais pas sa main. Il doit par conséquent évoluer à l'aveugle en utilisant la proprioception pour amener son doigt à la même profondeur que la cible (*blind reaching*). Les auteurs ont observé que la distance jusqu'aux objets virtuels proches (34-50 cm) étaient souvent sous-estimée [183]. Ce constat était plus marqué pour les tâches d'estimation à l'aveugle, pour lequel l'erreur était d'environ 4cm. Une étude ultérieure basée sur ces travaux a observé des résultats similaires pour la tâche T2, mais ont trouvé que les participants surévaluaient les distances avec la tâche T1 [191].

Ces résultats soulignent la complexité de la perception de la profondeur en RM. La bonne gestion des différents indices de profondeur en RM est encore un sujet de recherche actif à l'heure actuelle. Or, la tâche chirurgicale implique intrinsèquement un travail en trois dimensions sur le corps du patient. Percevoir correctement la profondeur des informations virtuelles superposées au champ opératoire est donc essentiel pour adopter le bon geste chirurgical. Une estimation faussée de ces distances de l'ordre de plusieurs millimètres voir centimètres constitue donc un frein pour l'utilisation des casques actuels de RM lors d'une opération.

D'autres indices visuels sont également importants pour la RM. Jansen et al. ont étudié l'impact de la physicalité d'un objet dans le domaine de la visualisation d'information [105]. Les auteurs ont observé que l'utilisation d'un objet physique pour la visualisation 3D conduisait à de meilleures performances pour les tâches de recherche d'informations que l'utilisation d'une visualisation à l'écran. Un facteur notable souligné par les auteurs pour expliquer cette différence de performance est le réalisme visuel des objets physiques. Plusieurs signaux visuels tels que les indices stéréoscopiques, les signaux d'accommodation et la résolution de l'affichage peuvent

avoir un impact sur l'interaction [105]. Cette notion de réalisme visuel soulève plusieurs questions pour la RM. En effet, des problèmes de perception liés à la fidélité des couleurs et la stéréoscopie soulèvent encore des défis techniques en RM [116].

Perturbation de l'attention Un autre frein à l'usage de la RM vient de la perturbation de l'attention visuelle qu'elle peut provoquer. En se concentrant trop sur les objets virtuels, l'utilisateur peut entrer dans un effet tunnel (*cognitive tunneling*) et ne pas détecter certains événements inattendus. Ce phénomène est particulièrement étudié dans des contextes applicatifs où l'attention de l'utilisateur est un facteur critique comme par exemple l'aviation [49]. Dans le cas de la chirurgie, le principal risque est de rater une complication inattendue au niveau du corps du patient [147]. Dixon et al. ont étudié ce phénomène de cécité d'attention en conduisant une étude intra-sujets (*between subjects*) sur un exercice d'endoscopie [60]. La Figure 2.13 illustre le dispositif expérimental utilisé.

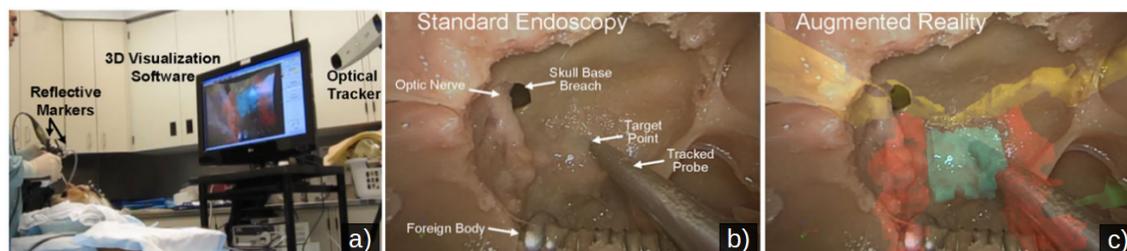


FIGURE 2.13 – a) Dispositif expérimental utilisé dans [60] pour étudier la cécité d'attention en chirurgie augmentée. b) et c) : Captures d'écran du retour vidéo affiché sur l'écran sans ou avec RA. Images issues de [60].

Le premier groupe de chirurgiens disposait d'un système de RA (groupe test) alors que le second devait effectuer une navigation classique sans RA (groupe contrôle). Les auteurs ont observé une tendance nette à propos du groupe test. Les chirurgiens utilisant le système de RA n'ont presque jamais repéré la complication ou le corps étranger situé à proximité de leur cible. Cette tendance était beaucoup moins marquée pour le groupe contrôle sans RA [60]. Bien que dans cette étude la réalité augmentée soit limitée à un retour vidéo augmenté sur un écran physique distant, de tels résultats montrent la nécessité d'étudier ce phénomène lors de l'utilisation de casques de réalité mixte. Par ailleurs, une autre étude menée auprès de chirurgiens suggère que cette cécité d'attention est étroitement liée à la charge cognitive [94], qui est élevée lors d'une opération [37].

Autres limites

Conduire des évaluations en conditions écologiques est une étape importante pour l'adoption de nouvelles techniques d'interaction. Or, la tâche chirurgicale complexifie grandement ce genre de tests. Le manque de temps des chirurgiens face à l'urgence des opérations et l'occupation des blocs pose souvent problème. Pour des questions de sécurité, les études préliminaires testant de nouvelles fonctionnalités logicielles ou de nouveaux outils sont souvent menées sur des cadavres. De telles évaluations restent néanmoins limitées à cause de leur coût financier important. Cette difficulté

à conduire des tests est une barrière importante au développement de nouvelles techniques d'interaction en bloc opératoire.

Ces problématiques de coût se reflètent aussi sur l'adoption de nouveaux équipements. Les machines médicales, outils spécialisés, protocoles de stérilisations et nombreux personnels de l'équipe médicale représentent un coût important. La tâche chirurgicale constitue donc une charge importante pour les hôpitaux. Tout nouvel équipement, que ce soit un casque de RM ou un système de reconnaissance de gestes, doit prouver sa valeur ajoutée sur le long terme et être soutenu par les chirurgiens afin d'avoir une chance d'être adopté. L'objectif reste qu'au bout de la chaîne, le contribuable ne paye que le juste prix pour les équipements hospitaliers utilisés. In fine, ces problématiques de coût concernent l'ensemble de système de santé, ce qui inclue aussi bien les industriels que la recherche académique.

Enfin, l'utilisation de casque de RM soulève également de nombreuses questions sur la sécurité des données et la vie privée des patients. Les données médicales font partie des informations personnelles les plus sensibles et doivent donc être protégées en conséquence. Cela peut inclure l'accès sécurisé aux données anatomisées stockées à des fins d'enseignement ou de recherche tout comme le chiffrement des communications. Nous citons à cet égard l'exemple récent de la première intervention chirurgicale menée avec une plateforme collaborative de réalité mixte dans un hôpital parisien [157]. Les fichiers contenant les données anatomiques du patient étant particulièrement lourds et complexes à traiter, les calculs ont été réalisés sur un serveur dédié afin de permettre au casque (Hololens) de conserver de bonnes performances lors de l'affichage des objets virtuels [143]. Cet exemple met en valeur la nécessité de protéger les données médicales et les communications afin d'empêcher toute récupération ou manipulation malveillante qui pourrait nuire au chirurgien et au patient.

La tâche chirurgicale est un travail crucial mais particulièrement complexe. Au travers de cette section, nous avons mis en valeur les multiples bénéfices de la RM pour le travail des chirurgiens. Plusieurs limitations de cette technologie ont également été présentées. La section suivante approfondit la RM appliquée à la chirurgie augmentée en étudiant les techniques d'interaction utilisables en bloc opératoire. Le contexte chirurgical impose en effet de fortes contraintes au chirurgien et restreint le champ des possibles parmi les techniques d'interaction présentées à la Section 1.2.

2.2 Interaction en bloc opératoire

Afin de compléter la présentation de notre contexte applicatif, nous nous concentrons dans cette section sur l'étude de l'interaction lors d'une opération chirurgicale en RM. Nous exposons les différentes contraintes existantes et de leur impact sur les techniques d'interaction utilisables en bloc opératoire. Nous mettons en valeur le fait que de nombreuses techniques d'interaction existantes présentées à la Section 1.2 ne sont pas adaptées à ces contraintes et que de nouvelles techniques sont par conséquent nécessaires.

2.2.1 Contraintes existantes en bloc opératoire

Asepsie et interaction avec les membres supérieurs

De part la nature même de son travail, le chirurgien doit respecter de nombreuses contraintes avant et pendant une opération. La première de ces contraintes et sans doute la plus importante est celle de l'*asepsie* liée aux risques infectieux. Afin de limiter au maximum la probabilité d'une contamination par des micro-organismes, l'équipe médicale doit suivre des protocoles très stricts au sein du bloc opératoire. Ces règles ont pour but de protéger le patient lui-même, mais également l'équipe médicale. Outre la désinfection corporelle des opérateurs, les outils chirurgicaux doivent suivre un long processus de stérilisation avant l'opération. Ce processus peut prendre jusqu'à deux heures car tout objet qui entre dans le champ opératoire (c'est à dire à proximité immédiate du patient) doit impérativement être stérile. Ainsi, dans le cas de l'OrthoPilot®), les différents rigid-bodies (Section 2.1.1) doivent donc être décontaminés avant chaque opération. Par ailleurs, la circulation de l'air de la salle doit être strictement contrôlée pour limiter la présence d'éléments infectieux.

L'équipement du chirurgien comprend des vêtements stériles, une cagoule pour entourer les cheveux et les oreilles, un masque, des sabots ainsi que des gants stériles (parfois deux paires) [184]. Cela force le chirurgien à ne toucher que des éléments stériles, sous peine de devoir enlever ses gants et recommencer le processus de désinfection. Or, les équipements électroniques ne peuvent pas subir le processus de stérilisation classique sans s'abîmer. Une procédure de nettoyage particulière et moins efficace est alors nécessaire. Tenir un écran à la main comme en RA sur dispositif mobile n'est donc pas envisageable lors d'une opération. Porter un casque de RM reste en revanche possible puisqu'il n'est pas en contact direct avec les membres supérieurs. Une fois le casque mis en place, le chirurgien ne peut néanmoins plus le toucher pour l'ajuster lui-même. Cela peut poser problème lors de l'opération si le casque est mal positionné ou fatigue trop le chirurgien. De plus, des projections de sang peuvent atteindre l'écran du casque et gêner la vision.

Même en cas de réelle nécessité, le chirurgien n'enlèvera pas ses gants stériles. Il sera en réalité contraint de forcer un autre membre de l'équipe médicale à effectuer la manipulation nécessaire à sa place. Ainsi, un chirurgien peut demander à un équipier d'aller sur l'ordinateur afin de vérifier une radio spécifique. La personne concernée devra alors :

1. Enlever ses gants stériles.
2. Suivre les instructions du chirurgien pour afficher à l'écran la disposition qui l'intéresse. Dans le cas d'une chirurgie sans RM, cela signifie utiliser des dispositifs classiques tels qu'un clavier/souris ou un écran tactile.
3. Se mettre à l'écart et recommencer le processus de stérilisations (mains + gants) lorsque c'est possible. Dans le cas contraire, la personne ne pourra plus remplir son rôle normal ni s'approcher du patient.

Bruit ambiant, encombrement et luminosité

Le bloc opératoire est un environnement bruyant. De nombreuses machines et équipements médicaux sont en effet présents dans la salle et créent un bruit de fond non négligeable. Certaines étapes d'une chirurgie peuvent également nécessiter

l'utilisation d'outils bruyants, par exemple pour découper une partie d'un os afin de pouvoir placer une prothèse. Il faut également ajouter à cela les communications orales entre les membres de l'équipe médicale. Un tel bruit ambiant pose problème pour les logiciels de reconnaissance vocale [91, 153]. Par ailleurs, tous les membres de l'équipe médicale doivent porter un masque chirurgical, ce qui complexifie l'utilisation de micros. Par ailleurs, comme l'ont souligné Mitrasinovic et al. [147], parler lentement et distinctement est un luxe lors d'une chirurgie puisque le temps est un facteur critique.

La présence de nombreuses machines médicales autour de la table d'opération limite également l'espace physique disponible. Les différents membres de l'équipe médicale doivent pourtant se déplacer pour remplir leur rôle, comme préparer les différents outils et les amener en temps utiles au chirurgien. Introduire de nouveaux dispositifs (par exemple, plusieurs écrans pour afficher des radios) posent donc non seulement un problème de temps (même éloignés du patient, ils doivent être nettoyés avant l'opération), mais aussi de place disponible. Les travaux prenant en compte ce type de contrainte sont relativement rares dans la littérature. Par exemple, Henderson et al. ont étudié un système de RA pour des tâches de maintenance à l'intérieur d'un véhicule militaire, dans lequel l'espace et la liberté de mouvement sont particulièrement contraints [89].

L'encombrement physique du bloc opératoire impacte également les interactions gestuelles. O'Hara et al. ont relevé l'effort d'adaptation des infirmières devant planifier leurs actions afin d'éviter toute interférence avec le chirurgien en train d'interagir avec le système [153] (Figure 2.14) en faisant des gestes dans l'air. L'encombrement, la portée et la robustesse du système de reconnaissance des gestes sont donc des problématiques clés en bloc opératoire.

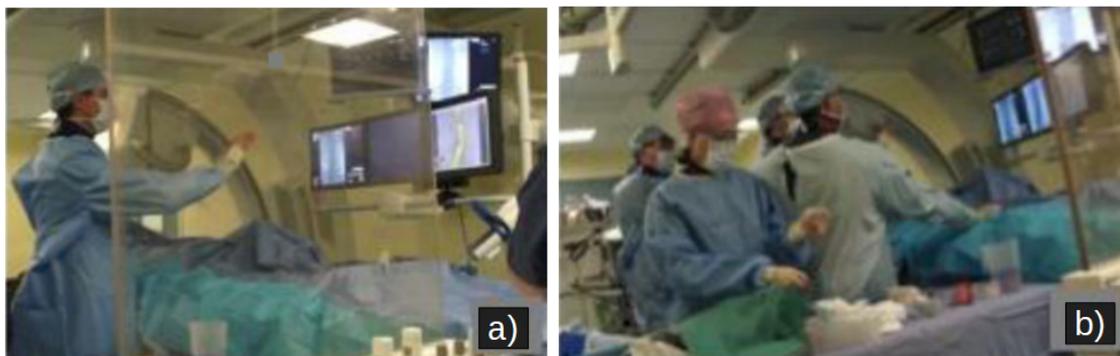


FIGURE 2.14 – Illustration de la gestion de l'espace disponible en bloc opératoire. a) Un chirurgien se plaçant pour que ses gestes soient détectés par le système (cas idéal). b) L'équipe médicale devant s'adapter à l'espace physique restreint au cours d'une opération. Images modifiées issues de [153].

Le bloc opératoire est également un environnement avec une forte luminosité artificielle. Une telle exposition peut poser problème pour la détection des gestes pour des techniques comme [52, 87]. Colaço et al. ont observé que le capteur qu'ils proposent pour l'interaction gestuelle maintenait de bonnes performances dans un environnement ensoleillé en extérieur [46], mais leur solution n'a pas été testée en bloc opératoire. Par ailleurs, il peut être plus difficile de détecter la position des

maines par rapport à la couleur de la peau ou des gants puisque le chirurgien porte une blouse et des gants qui peuvent être tâchés par des projections.

Tâche chirurgicale

La tâche chirurgicale elle-même a un fort impact sur le chirurgien et sur sa capacité à interagir. Une opération est une procédure complexe et critique, à la fois longue et sujette à une forte pression temporelle. Le chirurgien peut être forcé de changer ses plans en cas de complications et à prendre des décisions en un laps de temps très court. L'enjeu de ces décisions peut être la survie du patient. Le chirurgien a donc une charge cognitive très élevée au cours d'une opération. Devoir se rappeler d'un ensemble de gestes et de commandes [123, 136] est donc particulièrement contraignant pour un chirurgien déjà fortement occupé et stressé.

Les performances du chirurgien peuvent en effet être mises à mal par un stress élevé au cours de l'opération. Dans leur synthèse de l'état de l'art sur le sujet, Arora et al. ont observé différents facteurs de stress pouvant affecter la performance des chirurgiens [8]. Les auteurs évoquent non seulement des facteurs techniques (complexité de la procédure, pertes du sang, problèmes au niveau des équipements, etc), mais aussi des facteurs externes (distractions, pression temporelle, observateurs extérieurs) [8]. En plus de son savoir-faire technique, le chirurgien a également besoin de disposer de bonnes compétences inter-personnelles pour mener à bien l'opération [95]. La gestion du stress et de la fatigue, la communication et la capacité à travailler en équipe, l'organisation et la gestion des ressources sont autant de points clé en bloc opératoire [217, 218]. Une étude menée auprès de trois hôpitaux américains suggère que 43% des incidents reportés ont été causés par un problème de communication entre les membres de l'équipe médicale [77].

2.2.2 Techniques existantes utilisables en bloc opératoire

Les contraintes présentées ci-dessus restreignent les techniques d'interaction utilisables en bloc opératoire. Dans les sections suivantes, nous reprenons les classes de techniques d'interaction présentées en Section 1.2 et discutons de leur utilisabilité lors d'une opération.

Objet virtuel proche : Manipulation directe

Les contraintes d'asepsie sont impératives et ne peuvent être contournées sous aucun prétexte, en particulier à proximité du patient. Toute les techniques de main virtuelle [35, 123] ajoutant du matériel sur les mains du chirurgien posent donc problème. Utiliser d'épais gants électroniques lors d'une opération chirurgicale n'est tout simplement pas envisageable, aussi bien pour des raisons de stérilité que de dextérité manuelle.

Les techniques qui détectent directement la position des doigts et de la main ne souffrent pas de cette limitation [25, 177]. En revanche, la détection robuste de la main reste complexe en bloc opératoire en raison de l'encombrement, de la luminosité ambiante et des gants potentiellement tâchés par des projections. De plus, l'absence de retour haptique est une limitation importante pour la manipulation directe d'objets virtuels. Il n'est pas possible de pallier cette limitation grâce à la

RA tangible [31]. En effet, le problème est le même que pour les gants électroniques : le chirurgien ne peut pas manipuler d'objet non stérile à cause de l'asepsie. Les techniques basées sur ce type d'approche [121, 182] ne sont pas envisageables en chirurgie augmentée. Par ailleurs, la fatigue physique est également un facteur crucial pour le chirurgien qui impacte l'utilisabilité de techniques de manipulation directe en bloc opératoire.

Objet virtuel distant : Pointage

Les contraintes d'asepsie interdisent au chirurgien de toucher tout objet non stérile. Néanmoins, les outils chirurgicaux n'entrent pas dans cette catégorie et peuvent donc être utilisés pour interagir. Certains systèmes existants comme l'OrthoPilot®[3] comportent déjà des techniques de pointage basées sur des outils chirurgicaux existants. Par exemple, le palpeur est un outil semblable à un tournevis équipé d'un rigid-body : il est donc possible de connaître sa position en temps réel grâce à la caméra du système. Le chirurgien peut ainsi utiliser une technique de raycasting en utilisant ce palpeur [82], comme illustré précédemment à la Figure 2.5. Au final, de nombreuses techniques de raycasting existantes [66, 155, 209] peuvent potentiellement être adaptées au bloc opératoire. Certains outils chirurgicaux comme le *MultiTool* disposent d'une gâchette mécanique qui peut servir à valider une sélection [3]. En revanche, il est important de noter qu'aucun dispositif électronique n'est présent sur ces outils. Les approches basées sur des capteurs ou des surfaces tactiles [19] (voir Section 1.2.2) ne sont donc pas utilisables.

De manière similaire, les techniques d'interaction basées sur les mouvements de tête [10, 118, 155, 168] respectent les consignes d'asepsie. Il en est de même pour les techniques basées sur le regard [104, 118, 168]. Ces dernières souffrent néanmoins de la précision et de la stabilité moindres inhérentes au regard, qui posent particulièrement problème lors d'une opération chirurgicale. La sensation de contrôle et le sentiment de certitude sont en effet primordiaux pour le chirurgien [37]. De même, le temps et la fatigue sont également des facteurs clé en bloc opératoire. Les mécanismes tels que la poursuite de la cible [199] et les gestes du regard [96] ne sont donc pas adaptées au contexte chirurgical. En revanche, les techniques d'interaction permettant de sélectionner une cible sans tourner la tête pour la regarder [125, 213] (Section 1.2.2) sont particulièrement intéressantes en bloc opératoire. Elles permettent au chirurgien de concentrer leur attention visuelle sur le patient.

Sélection de commandes

Commandes gestuelles Les techniques basées sur des gestes "mains libres" [46, 52, 87] constituent une alternative compatible avec les contraintes d'asepsie. Comme pour les gestes de sélection d'un objet décrits ci-dessus, la détection robuste des gestes se heurte néanmoins à la forte luminosité ambiante. L'encombrement du bloc opératoire restreint la zone de détection possible et force l'équipe médicale à s'adapter pour ne pas perturber l'interaction en créant des faux positifs. De plus, le chirurgien doit mémoriser les gestes à effectuer et leur correspondance avec les actions associées, ce qui affecte sa charge cognitive. La fatigue physique des gestes manuels pose là encore problème pour une opération en chirurgie augmentée.

Augmenter le corps du chirurgien (Section 1.2.3) pour éviter cette fatigue physique n'est pas souhaitable. Ainsi, le menu virtuel affiché sur l'avant-bras proposé par Azai et al. [12] n'est pas adapté au contexte chirurgical. Le chirurgien doit limiter le plus possible tout contact direct au niveau des membres supérieurs, y compris avec sa propre blouse. En revanche, l'utilisation de menus invoqués et temporaires [52] reste possible.

Commandes au pied Les techniques d'interaction basées sur les pieds sont utilisables en bloc opératoire puisque les membres inférieurs du chirurgien n'entrent pas en contact direct avec le patient. Les consignes d'asepsie sont par conséquent légèrement moins strictes. Par ailleurs, cela permet au chirurgien d'interagir même si ses mains sont occupées. Pakkanen et al. ont comparé l'utilisation de la main et du pied pour interagir avec une interface graphique [156]. Les auteurs ont observé que l'interaction au pied était moins performante et jugée moins satisfaisante que l'interaction avec la main. En revanche, utiliser le pied semble acceptable pour des tâches secondaires ne demandant pas une grande précision ni une vitesse d'exécution élevée [156].



FIGURE 2.15 – Pédales médicales utilisables en bloc opératoire, filaire (a) ou sans fil (b). Images fournies par Aesculap.

Dans les systèmes tels que l'OrthoPilot®), la pédale est généralement installée à proximité de la table d'opération afin d'être facile d'accès. Cela peut néanmoins contraindre le chirurgien à se déplacer et à la chercher avec le pied lorsqu'elle se trouve hors de portée [93]. Hatscher et al. ont proposé un prototype attaché à une chaussure de bloc opératoire afin d'éviter cet écueil [86]. Au delà de simples actions du pied sur une pédale, offrir une plus grande liberté de mouvement au pied et une plus grande expressivité en associant des gestes du pied à des commandes a néanmoins un coût. De la même manière que les gestes manuels, de telles techniques doivent être robustes et éviter les faux positifs afin de ne pas perturber l'opération.

Commandes vocales De nombreux travaux ont exploré la combinaison de gestes avec des commandes vocales pour le bloc opératoire [65, 141, 153]. Ainsi, Ebert et al. ont proposé un prototype permettant au chirurgien de naviguer au sein d'images médicales grâce à des gestes des mains devant une Kinect et des commandes vocales [65]. Utiliser des commandes vocales en bloc opératoire est possible pour le chirurgien mais soulève encore des questions. Outre les problèmes techniques détaillés à la Section 2.2.1, il existe également un problème d'acceptation sociale de la part des chirurgiens. Il y a quelques années, l'entreprise Aesculap a mené une série d'interviews auprès de plusieurs hôpitaux au sujet de l'utilisation de la voix lors d'une

opération. Les résultats suggèrent que les chirurgiens interrogés étaient peu enclins à utiliser un système de reconnaissance vocale. Il semble qu'ils ne se voyaient pas "parler à une machine" devant le reste de l'équipe médicale.

Ces observations coïncident avec l'image du chirurgien présentée par Cassel [37]. L'auteur y souligne l'importance pour le chirurgien d'être le seul maître à bord, le meneur de l'équipe médicale. Cassel souligne qu'une opération est une procédure "publique", au sens où le chirurgien est observé, évalué, jugé par ses pairs et le reste de l'équipe médicale [37]. Il est possible que s'adresser vocalement à un système informatique et répéter des mots clés entre en conflit avec l'image de chef d'équipe que vise le chirurgien afin d'imposer ses décisions. Les récents progrès techniques du traitement de la parole [53, 203] et la démocratisation d'assistants vocaux à destination du grand public auront peut être raison de ces réticences dans les années à venir.

2.2.3 Bilan de l'interaction en bloc opératoire

Au final, quelles techniques d'interaction est-il possible de mettre en place en bloc opératoire ? D'un côté, de nombreux travaux en chirurgie augmentée se concentrent sur la visualisation d'informations médicales [14, 27, 109, 134]. Par exemple, les travaux de Mahmoud et al. soulignent les efforts de recherche en chirurgie laparoscopique augmentée pour positionner les informations médicales au bon endroit sur le patient [134]. Le système doit en effet tenir compte de multiples déformations des organes et autres tissus mous, par exemple avec la respiration du patient ou les gestes du chirurgien. Dans ce cas, l'étude de l'interaction en entrée et en sortie avec le système est alors mise de côté ou peu explorée. Or, de l'autre côté, de nombreuses techniques existantes dans le domaine de l'IHM n'ont pas été conçues pour un environnement aussi complexe que le bloc opératoire. Cette situation est schématisée à la Figure 2.16.

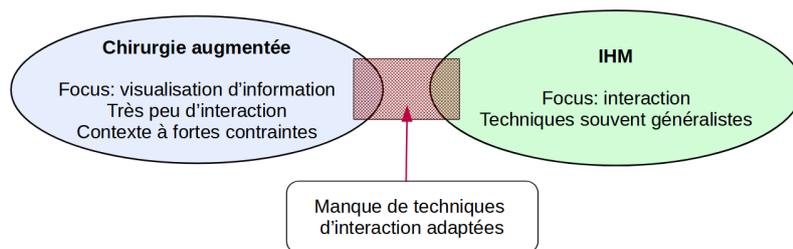


FIGURE 2.16 – Illustration du manque de techniques d'interaction adaptées à la chirurgie augmentée.

Les contraintes d'asepsie, de niveau de bruit et de charge cognitive limitent fortement les techniques d'interaction possibles. De plus, la charge cognitive et la disponibilité du chirurgien varient au cours de l'opération. Lors de certaines étapes de la chirurgie, le chirurgien a besoin de se concentrer exclusivement sur le patient. Au contraire, certains moments sont propices à l'interaction avec le système. L'utilisabilité des techniques d'interaction en bloc opératoire est donc liée aux caractéristiques des différentes tâches du chirurgien. La Figure 2.17 illustre les interactions possibles en fonction des différents types d'étapes d'une opération chirurgicale.

Étape	Criticité	Exemple en chirurgie orthopédique	Indisponible	Utilisable pour interagir
E1	Critique	Découpe de l'os pour placer la prothèse	Tout	Rien
E2	Importante	Placement de guides de coupe sur le corps du patient.	Mains occupées	Mouvements de tête Commandes vocales Commandes au pied
E3	Moindre	Plannification de la future position de la prothèse (étape de planning)		Mouvements de la main Mouvements de tête Commandes vocales Commandes au pied

FIGURE 2.17 – Techniques d'interaction disponibles en fonction de la criticité des étapes d'une chirurgie orthopédique.

Il est important de signaler que lors d'une étape de type E3, la charge cognitive du chirurgien reste élevée. Nous considérons néanmoins qu'il peut consacrer davantage de temps et d'attention à l'interaction avec le système qu'avec E2. Par exemple, cela peut être un temps consacré à la prise d'information et de décisions (*i. e.* consultation de radios ou inspection de l'anatomie augmentée du patient).

En guise de synthèse de la première partie de ce manuscrit (Chapitre 1 et 2), nous reprenons l'ensemble des techniques de la Section 1.2 au sein de la Figure 2.18. Nous y présentons les apports et les limites de chaque classe de technique pour la tâche chirurgicale en bloc opératoire.

Techniques basées sur	Utilisabilité en bloc opératoire	Avantages	Limites
Manipulation directe via des équipements électroniques	Interdit par les contraintes d'aseptie		
RA tangible, manipulation d'objets			
Interaction sur le corps du chirurgien			
Pédale et interaction au pied	E3	Mains libres Pas d'attention visuelle requise	Emplacement fixe de la pédale Expressivité limitée
Commandes gestuelles "mains libres"	E3	Respect de l'asepsie Pas d'attention visuelle requise	Mémorisation de la correspondance geste/commande nécessaire mais charge cognitive déjà élevée Nécessite une ou les deux mains Interférences de l'équipe médicale Luminosité ambiante, zone de détection
Pointage avec la main	E3	Respect de l'asepsie Pas d'attention visuelle requise	Nécessite une ou les deux mains Interférences de l'équipe médicale Luminosité ambiante, zone de détection
Raycasting avec un outil chirurgical tenu en main	E3	Compatible avec l'asepsie Déjà utilisé par certains systèmes	Une main nécessaire Attention visuelle requise + Perte du contexte visuel initial Tous les outils chirurgicaux ne sont pas adaptés
Commandes vocales	E2 et E3	Mains libres Pas d'attention visuelle requise	Bruit ambiant important, port du masque Pression temporelle Acceptation sociale Peu adapté aux manipulations continues
Pointer avec le regard	E2 et E3	Mains libres Plus rapide que la tête	Attention visuelle requise Précision limitée : conflit avec le besoin de contrôle Fatigue, confort moindre qu'avec la tête
Pointer avec la tête	E2 et E3	Ne nécessite pas les mains Meilleur confort et stabilité que le regard	Perte du contexte visuel initial Fatigue physique

FIGURE 2.18 – Caractéristiques des techniques d'interaction existantes en RM (Section 1.2) par rapport aux contraintes du bloc opératoire et de la tâche chirurgicale.

Conclusion

Une opération chirurgicale est donc une tâche particulièrement délicate qui impose une charge cognitive élevée au chirurgien. Cet aspect est primordial pour la conception de techniques d'interaction adaptées au bloc opératoire. Nous pensons qu'il est important d'insister sur ce point : le but du chirurgien est de mener à bien l'opération, pas d'interagir avec le système. Les techniques d'interaction doivent aider le travail du chirurgien et non alourdir sa tâche ou la remplacer. Le système de RM doit aider le chirurgien à prendre des décisions, à gagner du temps ou à mener l'opération de manière plus confortable tout en respectant les contraintes du bloc opératoire.

Nous avons expliqué à la Section 2.2.2 à quel point il est difficile d'introduire de nouveaux équipements en bloc opératoire. Notre approche est différente : pour concevoir des techniques d'interaction adaptées, nous partons des besoins et contraintes du chirurgien. Notre objectif est d'exploiter les éléments déjà présents en bloc opératoire tels que les outils chirurgicaux et les rigid-bodies. Ainsi, nous visons à ce que les nouvelles techniques d'interaction que nous proposons aident le chirurgien à mener à bien l'opération tout en étant les plus transparentes possibles pour l'ensemble de l'équipe médicale. Dans la partie suivante, nous détaillons nos contributions sur l'interaction avec des menus et sur la navigation en RM pour la chirurgie augmentée.

Deuxième partie

Interaction : techniques de menu et de navigation

Chapitre 3

Interaction avec la tête : techniques de menu

Sommaire

3.1	Espace de conception pour l'interaction avec les mouvements de la tête en RM dans le cas de menus	72
3.1.1	Élément d'interaction contrôlé par les mouvements de la tête	73
3.1.2	Physicalité de la cible	75
3.2	Conception de nouvelles techniques de menu	78
3.2.1	Principes de conception	78
3.2.2	Implémentation des techniques	79
3.3	Études expérimentales	81
3.3.1	Étude expérimentale 1 : HCC et HCM	81
3.3.2	Étude expérimentale 2 : transitions entre les modes RA et RV	89
3.3.3	Étude expérimentale 3 : physicalité des menus	91
3.4	Leçons apprises et discussion	102
3.4.1	HCC et HCM	103
3.4.2	Menus virtuels	103
3.4.3	Menus mixtes	103
3.4.4	Menus physiques	104

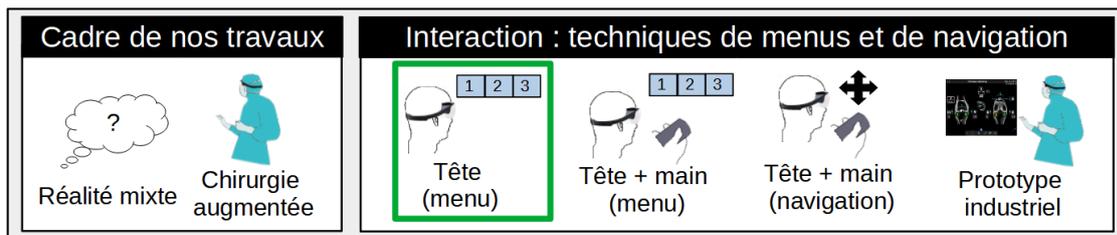


FIGURE 3.1 – Plan du mémoire (Figure B) - Focus sur le 3ème chapitre.

Introduction

Nous avons expliqué au chapitre précédent que les techniques d'interaction utilisables varient au cours d'une opération. Lors des différentes étapes d'une chirurgie augmentée, le chirurgien n'a pas forcément la même capacité à interagir avec le système. Dans ce chapitre, nous nous intéressons aux étapes de type E2 (Figure 3.2). Nous considérons que le chirurgien 1) a une charge cognitive importante, 2) a les mains occupées et 3) peut avoir besoin d'interagir de manière ponctuelle avec le système. Pour répondre à ce besoin, nous nous concentrons sur les techniques d'interaction basées sur les mouvements de tête. L'espace de conception, les techniques d'interaction et les deux premières études expérimentales exposés dans ce chapitre ont fait l'objet d'une publication à la conférence Interact 2019 [15].

Étape	Criticité	Exemple en chirurgie orthopédique	Indisponible	Utilisable pour interagir
E1	Critique	Découpe de l'os pour placer la prothèse	Tout	Rien
E2	Importante	Placement de guides de coupe sur le corps du patient.	Mains occupées	Mouvements de tête Commandes vocales Commandes au pied
E3	Moindre	Plannification de la future position de la prothèse (étape de planning)		Mouvements de la main Mouvements de tête Commandes vocales Commandes au pied

FIGURE 3.2 – Rappel de la Figure 2.17 : techniques d'interaction disponibles en fonction des différents types d'étapes d'une chirurgie orthopédique.

3.1 Espace de conception pour l'interaction avec les mouvements de la tête en RM dans le cas de menus

Les chirurgiens orthopédiques travaillant avec des systèmes de navigation chirurgicale ont l'habitude d'utiliser des techniques de raycasting pour interagir avec le système [82], comme expliqué à la Section 2.1.1. Avec les systèmes comme l'OrthoPilot (®), ces chirurgiens contrôlent un curseur virtuel sur un écran physique distant afin de sélectionner des commandes via des menus à une dimension [3]. Une illustration de ce type de menu est disponible à la Figure 3.3.

Notre objectif est de tirer parti de cette familiarité avec les curseurs virtuels et les menus. Nous avons vu en Section 2.2.2 que de nombreuses techniques d'interaction telles que les commandes vocales ou l'utilisation du regard posent problème en bloc opératoire. Parmi les techniques envisageables, les mouvements de la tête sont particulièrement intéressants pour l'interaction avec des menus et des curseurs. En effet, en RM avec un casque le champ de vision augmenté est directement lié à la position de la tête. Nous proposons un espace de conception axé sur les techniques de menu basées sur les mouvements de la tête. Les deux sections suivantes présentent les deux axes de cet espace de conception.

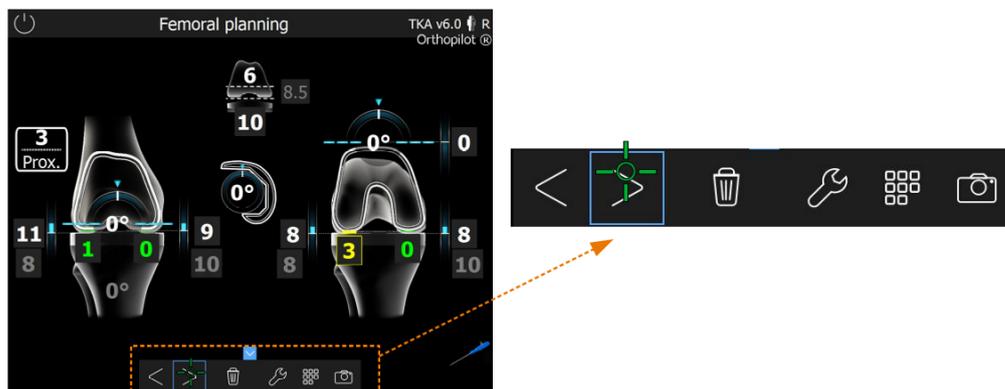


FIGURE 3.3 – Exemple de menu à 1 dimension utilisé dans plusieurs écrans de l’interface graphique de l’OrthoPilot®[3]. Image fournie par Aesculap.

3.1.1 Élément d’interaction contrôlé par les mouvements de la tête

Le premier facteur de conception que nous considérons est l’élément contrôlé par les utilisateurs lorsqu’ils bougent la tête : un curseur ou le menu lui-même. Lors de l’utilisation d’un casque, le champ de vision augmenté est directement lié à la position de la tête. Par conséquent, tout élément graphique affiché à une position fixe sur l’écran du casque se déplacera dans le cadre de référence du monde physique en fonction des mouvements de la tête. L’élément graphique est fixe dans le cadre de référence de l’écran et mobile dans le cadre de référence du monde physique. Lors de l’interaction avec un menu, ce principe conduit à deux solutions : déplacer le curseur ou déplacer le menu.

Avec un curseur contrôlé par la tête (HCC, pour *Head-Controlled Cursor*), les utilisateurs visent des cibles (c’est-à-dire des éléments de menu) qui sont fixées dans le cadre de référence du monde physique en orientant leur tête et en alignant un curseur sur les cibles (voir Figure 3.4). HCC minimise l’occultation visuelle sur le casque. Seul un curseur virtuel est affiché au centre de l’écran. Ceci est un avantage significatif au vu de la taille réduite du champ de vision augmenté sur les casques actuels. En outre, HCC reproduit la même métaphore qu’une interface WIMP classique avec une souris. Cette approche est donc familière pour la plupart des utilisateurs, y compris pour des novices en matière de RM.

Au lieu de déplacer un curseur vers un élément d’un menu, le menu peut être directement contrôlé avec la tête. Lors de l’utilisation d’un menu contrôlé par la tête (HCM, pour *Head-Controlled Menu*), l’élément du menu ciblé doit être amené à un emplacement fixe dans le cadre de référence du monde physique (c’est-à-dire une position de curseur) afin d’être sélectionné (voir Figure 3.5). Dans une certaine mesure, cette approche est similaire à la technique *Toolglass* [28]. *Toolglass* est basée sur une feuille transparente présentant un ensemble d’outils (*i.e.* une palette) qui peut être contrôlée avec une trackball. Cette palette peut ainsi être superposée à l’objet d’intérêt avec une main tandis que l’autre main contrôle la souris. Les utilisateurs peuvent alors sélectionner à la fois la commande et sa cible en même temps. HCM et *Toolglass* partagent une même idée : amener ce que Beaudouin-Lafon appelle des méta-instruments [22] dans une zone de sélection (c’est-à-dire le

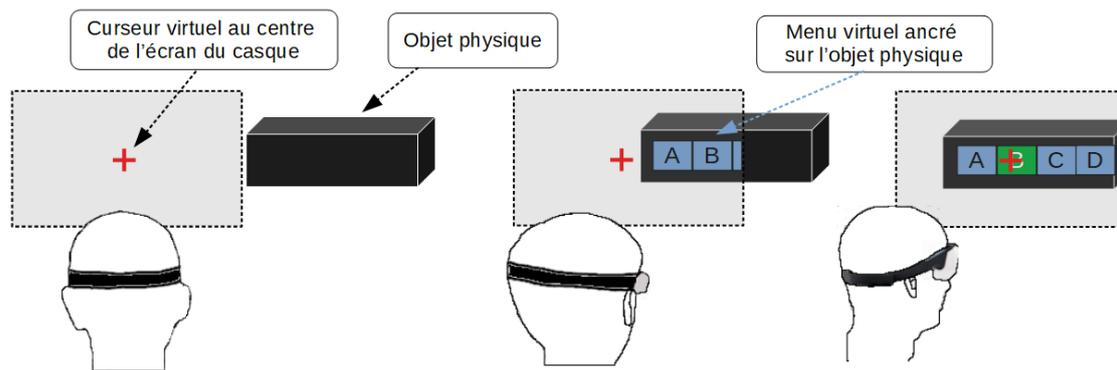


FIGURE 3.4 – Technique HCC : les mouvements de la tête permettent d’amener un curseur virtuel jusqu’au menu. Le curseur est au centre de l’écran du casque alors que le menu est inscrit spatialement dans l’environnement mixte.

curseur pour la technique HCM et l’objet d’intérêt pour la technique Toolglass).

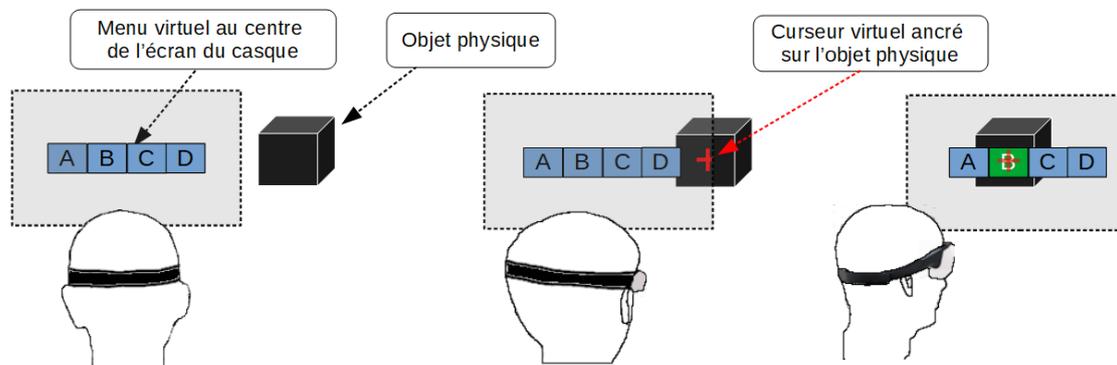


FIGURE 3.5 – Technique HCM : les mouvements de tête permettent d’amener un menu virtuel jusqu’au curseur. Le menu est au centre de l’écran du casque alors que le curseur est inscrit spatialement dans l’environnement mixte.

Bien sûr, un menu contrôlé par la tête implique une occultation visuelle importante en affichant le menu au centre de l’écran du casque. Cette occultation est bien plus importante que celle causée par un simple curseur. Cette limitation n’est pas acceptable pour les étapes chirurgicales E1 (Figure 3.2), lorsque les chirurgiens se concentrent uniquement sur les gestes complexes à effectuer. Aucune interaction ni occultation visuelle n’est alors envisageable. En revanche, lors des étapes E2 que nous considérons ici, l’intrusion visuelle est moins problématique tant qu’elle reste limitée aux moments d’interaction. Les menus ne nécessitent en effet pas d’espace d’écran permanent et peuvent être transitoires. Dans ce cas, ils apparaissent à la demande et se ferment immédiatement après la sélection d’un élément [17]. Par exemple, le système peut aussi afficher le menu uniquement lorsque les utilisateurs éloignent leur tête d’une zone d’intérêt principale (par exemple, le champ opératoire) dans une direction donnée ou après un appui sur une pédale. Ces actions pourraient être un déclencheur pour indiquer que l’objectif actuel de l’utilisateur est d’interagir avec le menu. Un avantage de HCM est que dès le début de la sélection, les utilisateurs perçoivent le menu et sont donc conscients des commandes actuellement disponibles.

L'élément d'interaction contrôlé par l'utilisateur, à savoir le curseur (HCC) ou le menu (HCM) définit le premier axe de notre espace de conception.

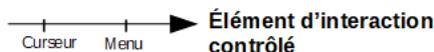


FIGURE 3.6 – Premier axe de l'espace de conception : l'élément d'interaction contrôlé par l'utilisateur.

3.1.2 Physicalité de la cible

Contrairement à la réalité virtuelle où chaque élément visible est virtuel, la RM ouvre plus de possibilités le long du continuum réalité-virtualité. Nous considérons 3 types de physicalité pour les cibles :

- Les cibles *virtuelles* ne sont pas spatialement inscrites dans l'environnement physique. Elles sont affichées soit 1) à une position donnée dans le repère de l'écran du casque, soit 2) à une position spatiale dans le repère de l'environnement mais sans attache avec l'environnement physique (un cible flottant dans les airs). Ces cibles ne sont visibles qu'au travers du casque de RM.
- Les cibles *mixtes* sont basées sur deux composantes : une composante virtuelle et une composante physique. Contrairement aux cibles purement virtuelles, les cibles mixtes sont ancrées dans l'environnement physique grâce à leur composante physique. Même sans casque de RM, cette ancre physique reste visible. Les objets augmentés, qui ont des objets physiques auxquels on superpose du contenu virtuel (Figure 2.6, entrent dans cette catégorie des cibles mixtes.
- Les cibles *physiques*. Ces cibles sont complètement visibles, même en dehors du champ de vision augmenté. En revanche, reposer entièrement sur un objet physique limite leur flexibilité. Dans cette catégorie, nous trouvons par exemple le menu tangible proposé par Ullmer et al. [197].

L'utilisation d'objets physiques comme ancres pour du contenu virtuel n'est pas nouvelle : Henze et al. ont mentionné cette approche comme un travail futur dans leur étude de la visualisation d'objets hors écran pour des cartes augmentées [90]. Cependant, cette approche doit encore être étudiée et évaluée. Ainsi, la position d'une cible mixte est directement visible dans la vision périphérique, ce qui peut aider les utilisateurs à mieux planifier leurs mouvements de la tête. Cela peut impacter en particulier la *phase préliminaire* de pointage dans le modèle de Meyer et al. [142] (voir Section 1.2.2). Dans le cas d'une cible physique, c'est l'ensemble de la cible qui est visible dès le début du pointage, comme le montre la Figure 3.7. Dans le cas d'un menu physique, tous les éléments du menu seront alors visibles. L'impact d'une telle différence sur une tâche de pointage a besoin d'être quantifié. Par ailleurs, comme expliqué en Section 2.1.3, le réalisme visuel des cibles virtuelles, mixtes ou physiques peut aussi influencer la performance des tâches de pointage.

En RM, il est donc possible d'utiliser des cibles physiques, mixtes ou virtuelles. Ces différentes physicalités définissent le second axe de notre espace de conception (voir Figure 3.8). Dans notre cas, l'élément d'interaction contrôlé par les mouvements de la tête doit cependant être virtuel. En effet, attacher un objet physique au casque

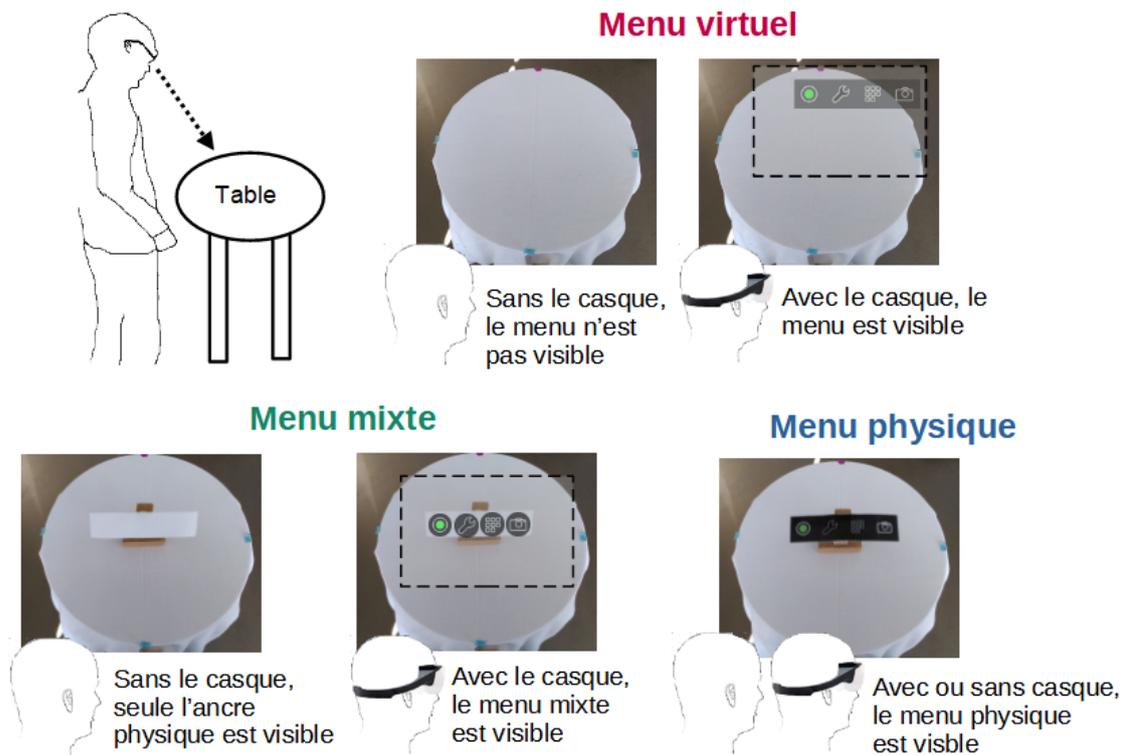


FIGURE 3.7 – Exemples de menu virtuel, de menu mixte (éléments virtuels attachés à une ancre physique) et de menu physique.

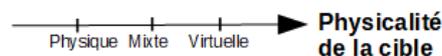


FIGURE 3.8 – Second axe de l'espace de conception : la physicalité de la cible

de RM dans le seul but de pouvoir interagir serait encombrant et malavisé, surtout en bloc opératoire. Ainsi, en déplaçant la tête, les utilisateurs contrôlent soit un curseur virtuel (curseur contrôlé par la tête) soit un menu virtuel (menu contrôlé par la tête). En synthèse, nous obtenons un espace de conception défini par deux axes, comme présenté à la Figure 3.9.

Ces axes sont importants car ancrer du contenu virtuel sur des objets physiques permet de visualiser l'emplacement de ce contenu dans l'environnement, et ce même en dehors du champ de vision augmenté. Nous pouvons alors étudier si de telles ancrages physiques ont un impact sur l'apprentissage de la localisation des cibles. Par ailleurs, la vision périphérique peut également être exploitée lors du pointage, ce qui a peut être une incidence sur les performances.

Nous avons expliqué dans la Section 2.2.1 que les règles d'asepsie limitent l'introduction de matériel dans la salle d'opération, en particulier à proximité du patient. Il n'est donc pas envisageable d'ajouter des objets au sein du champ opératoire dans le seul but de disposer d'ancres physiques pour les objets virtuels. Cependant, certains objets déjà présents dans cet espace pourraient être utilisés comme ancres. Par exemple, les *rigid-bodies* utilisés par l'OrthoPilot® sont des candidats pertinents. Ces branches métalliques sont fixées à la jambe du patient au début de

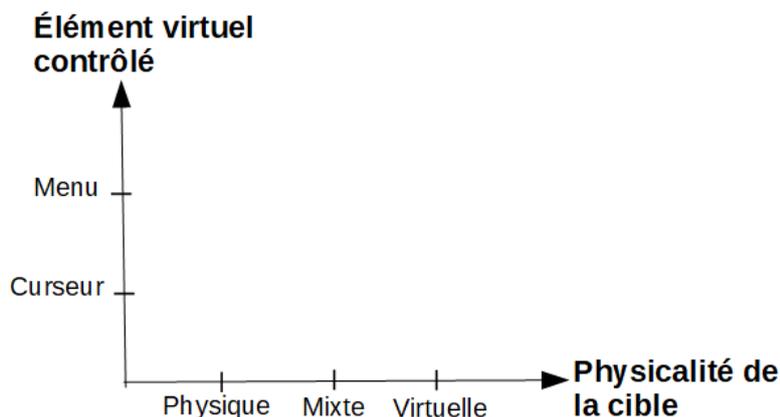


FIGURE 3.9 – Espace de conception complet

l'opération et regroupent différents marqueurs suivis par le système. Il est donc aisé de connaître leur position à tout instant. De plus, ces objets sont déjà inclus dans la procédure standard de stérilisation du matériel. La Figure 3.10 illustre l'une des utilisations possibles de rigid-body en tant qu'ancre physique.

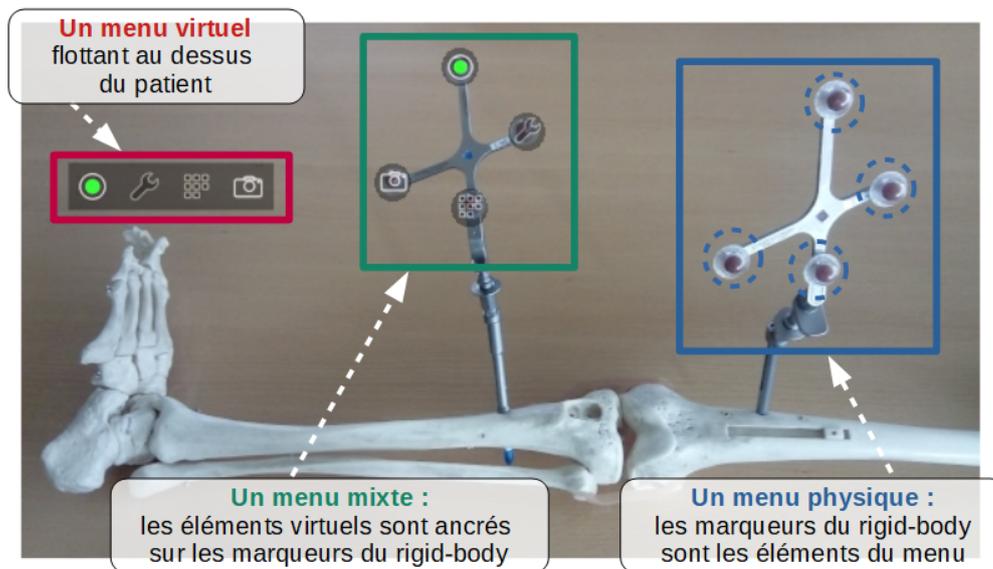


FIGURE 3.10 – Exemple de rigid-bodies utilisés comme ancres physiques pour du contenu virtuel. Un élément de menu virtuel est lié à chaque marqueur des rigid-bodies.

L'exploration de cet espace de conception a été réalisée en trois temps. Nous nous sommes d'abord concentrés sur la conception et l'évaluation de nouvelles techniques de menu impliquant des cibles virtuelles et mixtes (Section 3.3.1). Les résultats de cette première évaluation expérimentale nous ont amené à concevoir et évaluer des variantes des techniques conçues (Section 3.3.2). Enfin, une troisième étude expérimentale a inclus les menus physiques afin d'évaluer l'ensemble des possibilités de l'axe *Physicalité de la cible* de l'espace de conception (Section 3.3.3).

3.2 Conception de nouvelles techniques de menu

3.2.1 Principes de conception

L'espace de conception peut être divisé en six cases, comme illustré à la Figure 3.11. Les deux premières cases concernent les techniques basées sur des cibles physiques. Dans ce cas précis, une différence importante entre les techniques HCC et HCM doit être prise en compte. En effet, avec une technique HCM, la forme du curseur physique est secondaire. Le but de ce curseur physique est avant tout de matérialiser une zone précise servant à l'interaction. Par exemple, un marqueur de rigid-body servant de curseur physique (Figure 3.10) pourrait aussi bien être rectangulaire ou circulaire sans que cela ait une influence significative sur l'interaction.

Cette propriété n'est en revanche pas valable pour les techniques HCC. En effet, lorsque l'utilisateur contrôle un curseur virtuel, la conception du menu physique cible a une grande importance. Deux approches sont possibles : 1) matérialiser complètement chacun des éléments du menu (icône, label, etc) et 2) simplement indiquer l'emplacement de chaque élément avec une ancre physique. Dans ce second cas, l'utilisateur doit alors apprendre la correspondance entre chaque ancre physique et l'élément qui lui est associé. Ainsi, dans l'exemple présenté à la Figure 3.10, le chirurgien ne voit pas quel élément du menu est associé à quel marqueur. Autrement dit, il doit mémoriser cette correspondance pour chaque rigid-body, ce qui risque d'accroître sa charge cognitive déjà élevée.

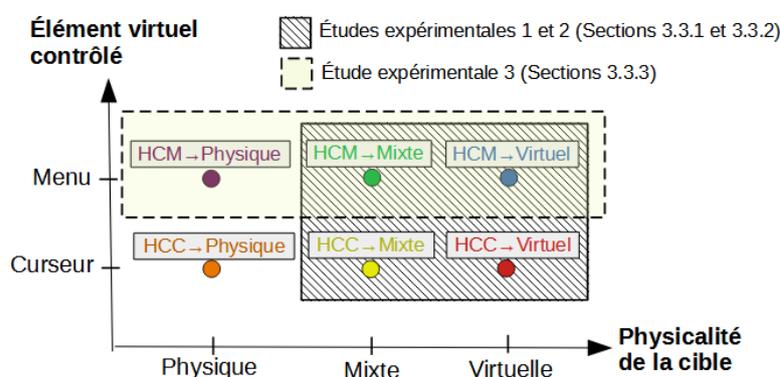


FIGURE 3.11 – 6 nouvelles techniques d'interaction induites de l'espace de conception. Exploration de cet espace par trois études expérimentales.

Pour l'exploration expérimentale de cet espace de conception, nous n'avons d'abord considéré que deux types de cibles : les cibles virtuelles et les cibles mixtes. Ces deux catégories ne souffrent en effet pas de cette différence entre HCC et HCM sur la matérialisation de la cible. La catégorie des cibles physiques fait l'objet d'une étude présentée à la Section 3.3.3.

Dans les deux premières études, la partie de l'espace de conception considérée est donc composée de quatre cases. Chacune de ces cases définit une technique de menu : $HCC \rightarrow Virtuel$, $HCC \rightarrow Mixte$, $HCM \rightarrow Virtuel$ et $HCM \rightarrow Mixte$. Parmi ces 4 techniques de menu, nous nous concentrons particulièrement sur les 2 techniques avec un menu contrôlé avec la tête (HCM). Ces deux techniques sont en effet les plus innovantes. Contrôler un curseur virtuel n'est en effet pas une nouvelle approche

(voir Section 1.2.2). Des techniques d'interaction basées sur ce principe existent déjà dans la littérature [103, 126] et dans certains casques de RV comme la première version de l'Oculus Rift [70]. La technique $HCC \rightarrow Virtuel$ est donc ici considérée comme une technique de référence. Cependant, l'influence de pointer vers un menu *Mixte* au lieu d'un menu *Virtuel* doit encore être étudiée (technique $HCC \rightarrow Mixte$).

Comme première exploration de l'espace de conception, nous considérons un menu à un niveau, proche de celui utilisé dans l'OrthoPilot [®][3] (Figure 3.3). Nous choisissons néanmoins une conception de menu simple et générique, non liée à un groupe d'utilisateurs ou à des conditions écologiques particulières. L'objectif est de pouvoir généraliser les résultats expérimentaux obtenus tout en utilisant un menu proche de ce que connaissent les chirurgiens. Un menu horizontal à un niveau est donc utilisé et chaque élément de menu est étiqueté avec un numéro ou une lettre. Ce type de menu sera utilisé lors des trois évaluations expérimentales présentées.

3.2.2 Implémentation des techniques

Les 4 techniques de menu considérées ont été implémentées avec le casque de RA Epson Moverio BT-300 [68] et le logiciel de suivi des mouvements *Tracking Tools*.

Choix du casque de RM

Les 4 techniques de menu ont été développées avec le casque Epson Moverio BT-300 [68]. Nous avons choisi ce casque en raison de sa légèreté. Les lunettes ne pèsent en effet que 70g. Ce poids réduit est lié au fait que la plupart des composants électroniques (dont la batterie) sont regroupés dans un petit boîtier relié en filaire aux lunettes. Ce boîtier peut être tenu à la main ou accroché à la taille. La légèreté du casque est un atout pour une tâche chirurgicale fatigante physiquement et mentalement. De plus, son encombrement réduit par rapport à d'autres casques existants plus volumineux est également un argument pour convaincre un chirurgien. Un aperçu de ce casque de RA est disponible à la Figure 3.12.

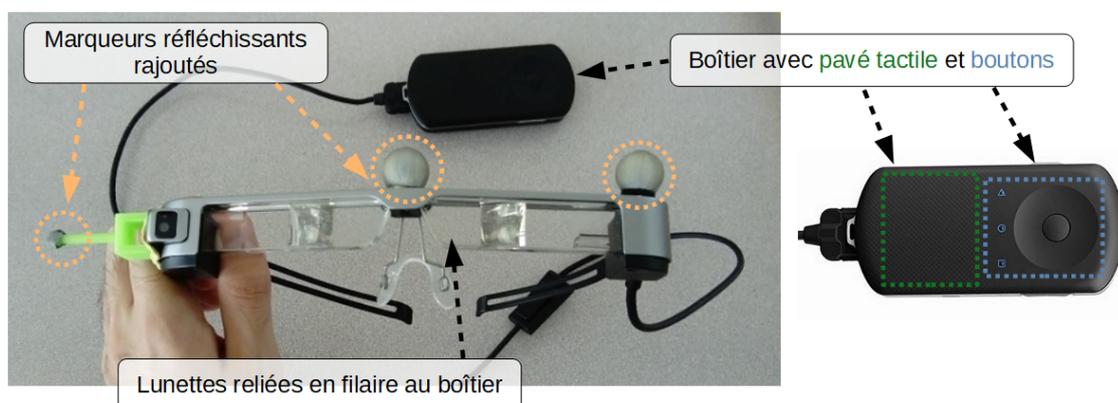


FIGURE 3.12 – Le casque de RA Epson Moverio BT-300 [68] utilisé lors de nos études expérimentales.

Choisir un tel casque de RA n'est cependant pas sans inconvénient. Les lunettes BT-300 disponibles dès novembre 2016 sont vendues par le constructeur pour des applications de RA comme le pilotage de drones [68]. L'accent a été mis sur la qualité

d'image et la légèreté des lunettes, qui doivent être utilisables dans un contexte mobile. En revanche, les performances techniques sont limitées, surtout en comparaison des capacités des casques les plus récents [145]. Pour conserver des performances optimales, nous tirons parti du fait que ce casque fonctionne sous Android 5.1. En effet, cela nous permet de développer en utilisant OpenGL ES 2.0 (*Open Graphics Library for Embedded System*), une API dédiée à l'affichage 2D et 3D pour des dispositifs embarqués (<https://www.khronos.org/opengles/>).

Système de suivi des mouvements

Une autre limitation du casque BT-300 est qu'il requière un système externe de suivi des mouvements de la tête. Le casque dispose d'une caméra et de ses propres capteurs de position et d'orientation tels qu'un gyroscope et un magnétomètre. La précision de ces capteurs et la dérive de cette précision au cours du temps posent néanmoins problème, surtout pour des contextes applicatifs comme la chirurgie augmentée. Par conséquent, un système externe de caméras est plus adapté afin de savoir en temps réel où l'utilisateur regarde et adapter l'affichage en conséquence.

Pour cela, nous avons choisi d'utiliser le logiciel spécialisé *Tracking Tools*, version 2.3.1. Ce logiciel se base sur un ensemble de caméras Flex3 pour détecter des marqueurs passifs. Comme illustré à la Figure 3.12, de simples marqueurs sphériques réfléchissant la lumière infra-rouge sont suffisants. Les caméras détectent donc ces marqueurs et envoient leur position au logiciel *Tracking Tools* qui s'exécute sur un ordinateur distant. Le logiciel identifie chaque marqueur et reconstruit la pose (position et orientation dans l'espace) de chaque élément suivi (ici, les lunettes de RA). Cette position est ensuite envoyée au casque BT-300 grâce à un script Python de notre conception. Cela permettra lors du prochain cycle de dessin de mettre à jour l'affichage en fonction de la pose de la tête de l'utilisateur. Ce processus est illustré à la Figure 3.13. Une moyenne de 24 ms est nécessaire pour obtenir l'estimation de la pose de la tête : 10 ms pour les caméras, 4 ms pour les calculs de pose et 10 ms pour envoyer le résultat au casque via Wifi.

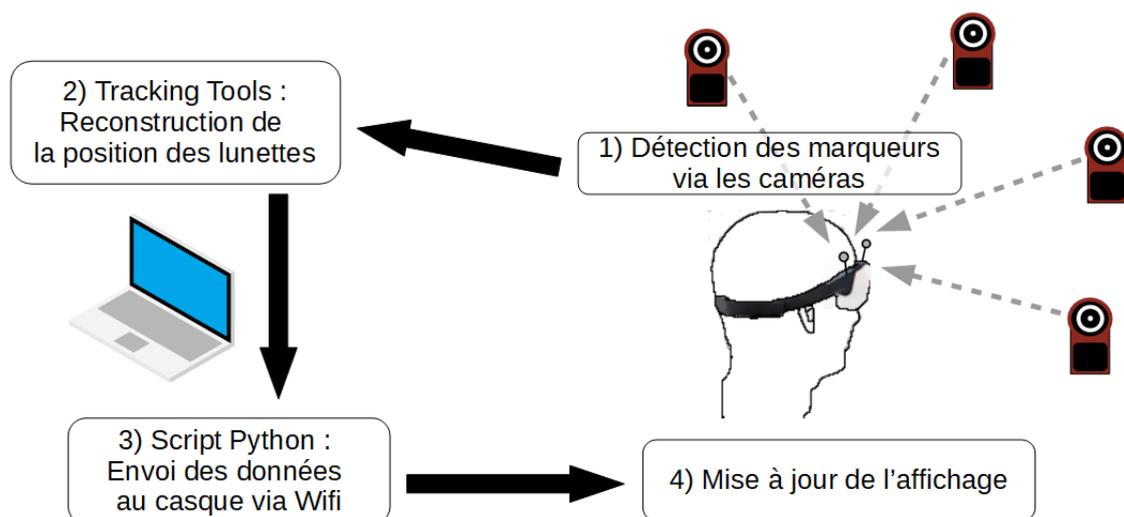


FIGURE 3.13 – Fonctionnement du suivi des mouvements de la tête de l'utilisateur grâce à un jeu de caméras externes.

3.3 Études expérimentales

Nous avons détaillé au sein de la section précédente notre espace de conception pour l'interaction avec des menus basée sur les mouvements de la tête (Figure 3.9). Nous avons également décrit la réalisation matérielle et logicielle des techniques de menu issues de cet espace. Cette section présente l'exploration de cet espace de conception grâce à trois évaluations expérimentales que nous avons positionné au sein de notre espace de conception à la Figure 3.11. La première étude se concentre sur les cibles mixtes et virtuelles en considérant les 4 techniques d'interaction présentées ci-dessus. La seconde étude poursuit ces travaux en évaluant des variantes de ces techniques. L'objectif de ces variantes est d'étudier si une transition vers un mode RV temporaire permet d'éliminer tout problème de perception au niveau des cibles mixtes. Enfin, une troisième étude se concentre sur la physicalité des menus, l'un des deux axes de notre espace de conception.

3.3.1 Étude expérimentale 1 : HCC et HCM

Cette première étude expérimentale avait deux objectifs. Notre premier objectif était de comparer les approches *Head-Controlled Cursor* (HCC) et *Head-Controlled Menu* (HCM). Étant donné que le HCC est plus proche d'une interaction WIMP classique que le HCM, nous nous attendions à ce que les participants préfèrent HCC et obtiennent de meilleures performances avec cette approche (**H1**). Notre deuxième objectif était d'évaluer les avantages des cibles mixtes par rapport aux cibles virtuelles. Nous avons émis l'hypothèse que les cibles mixtes auraient un impact positif sur l'interaction avec les menus. Deux raisons ont motivé cette hypothèse : 1) les cibles mixtes peuvent faciliter l'apprentissage des emplacements des cibles ; 2) elles peuvent changer la façon dont les participants planifient leurs mouvements de tête, conduisant à de meilleures performances (**H2**).

Conception de l'étude

Le but de notre installation expérimentale était de recréer les conditions d'un scénario général lorsque les utilisateurs ont les deux mains occupées. Pour ce faire, nous nous sommes inspirés d'un scénario générique de chirurgie augmentée. Notre but n'était cependant pas de recréer les conditions écologiques du bloc opératoire. Douze participants (3 femmes, 9 hommes) de 22 à 39 ans ($sd = 5,3$) ont été recrutés pour cette étude. Comme indiqué par notre questionnaire préliminaire, ils étaient tous novices en réalité mixte. Les participants avaient une vision normale ou corrigée à la normale et pouvaient utiliser des lunettes en dessous du casque si nécessaire.

Les participants se tenaient derrière une table carrée grise (81 cm de long, 73 cm de haut) représentant la zone de travail. Il leur été demandé de tenir deux os légers en plastique afin d'occuper leurs deux mains. Les ancrs physiques pour les deux conditions incluant des cibles mixtes étaient représentées par des balles de ping-pong orange. Le but était d'être sûr que les ancrs physiques soient clairement visibles sur la table grise en termes de taille (diamètre : 3,7 cm) et de couleur (voir Figure 3.14).

La temporisation (*dwell time*) est une approche classique pour valider la sélection d'une cible et a souvent été utilisée pour des sélections à l'aide du regard [102, 180].

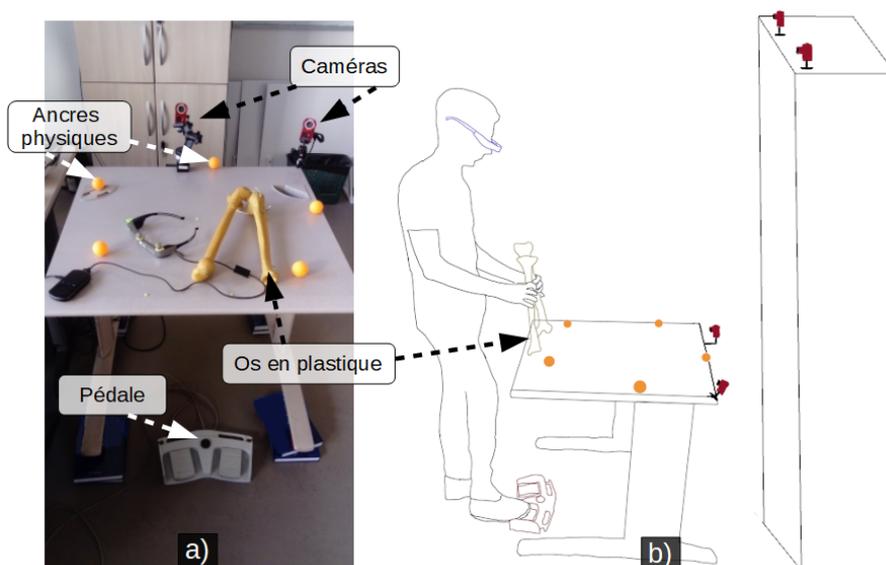


FIGURE 3.14 – (a) Installation expérimentale annotée. (b) Position d'un participant au cours de l'expérimentation.

Cependant, plusieurs auteurs l'ont également critiqué car cette méthode ralentit l'interaction afin d'éviter les erreurs et nécessite une concentration importante de la part de l'utilisateur [186]. Or, nous avons vu que le temps et la charge cognitive du chirurgien sont des facteurs clés lors d'une opération. Nous préférons par conséquent utiliser une pédale pour confirmer la sélection, comme pour l'OrthoPilot[®]. Utiliser l'un des boutons du boîtier du casque impliquerait de se servir d'une main lors de l'interaction, ce qui n'est pas envisageable lors des étapes chirurgicales E2 que nous considérons (voir Figure 3.2).

Tâche de pointage Les participants avaient pour consigne de sélectionner rapidement et avec précision un élément donné dans un menu virtuel en tournant leur tête. Ils devaient tenir les os en plastique avec les deux mains et les maintenir dans une position donnée. Un superviseur veillait à ce qu'ils gardent une pose correcte au cours de l'expérimentation. L'objectif était de s'assurer qu'ils ne pouvaient pas utiliser leurs mains pour interagir. Un menu horizontal à un niveau avec quatre éléments de 100*100 pixels était utilisé.

De manière similaire à l'étude réalisée par Louis et Berard [129], nous avons choisi de garder un petit nombre de cinq cibles (cinq menus ou cinq curseurs selon la technique) pour permettre aux participants de mémoriser rapidement les différents emplacements. Nous voulions être sûrs que les participants aient une connaissance préalable de l'emplacement des cibles pour éviter toute phase d'exploration lors des tâches de pointage. Au regard de notre contexte applicatif, cela était motivé par le fait que les chirurgiens ne peuvent pas se permettre de perdre du temps à chercher des cibles lors de l'exécution de leurs tâches principales.

Position des cibles Inspiré par l'approche adoptée par Guillon et al. [83], nous avons appliqué une version modifiée de la tâche de pointage ISO 9241-9 [100]. Comme indiqué précédemment, cinq cibles étaient placées sur la table d'opération afin de

former un cercle. Nous avons demandé aux participants de retourner au centre de ce cercle (qui était également le centre de la table) au commencement de chaque tâche de pointage. Pour être sûr qu'ils étaient correctement placés, il leur était demandé d'aligner un petit cercle affiché au centre de l'écran du casque avec une marque placée au centre de la table. Une fois correctement placé, le cercle affiché devenait vert pour indiquer aux participants qu'ils pouvaient valider afin d'obtenir la prochaine cible.

L'ordre des cibles était aléatoire. Pour identifier la cible actuelle, une flèche verte pointant vers l'emplacement cible était affichée en bas de l'écran du casque. Cette flèche de guidage disparaissait lorsque le menu cible (dans le cas HCC) ou le curseur cible (dans le cas HCM) devenait visible dans le casque. Toutes les cibles se trouvaient en dehors du champ de vision augmenté initial. En revanche, les ancrs physiques étaient visibles dans la vision périphérique dans les deux conditions basées sur des cibles mixtes. Une illustration de la tâche de pointage est disponible à la Figure 3.15.

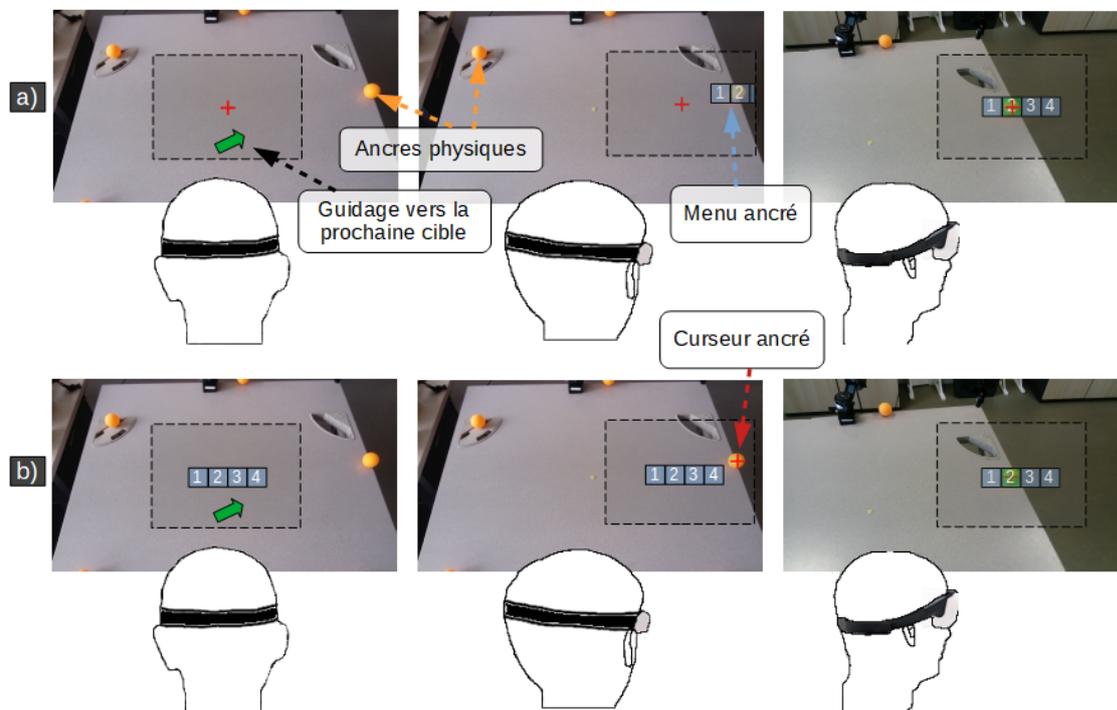


FIGURE 3.15 – Illustration de la tâche de pointage de l'étude 1. a) Technique HCC. b) Technique HCM.

Nous avons choisi de fixer l'indice de difficulté (ID) de toutes les cibles à 3,2. Selon Soukoreff et Mackenzie, cela correspond à des tâches relativement faciles [185]. L'ID a été calculé en utilisant la formulation de Wingrave et Bowman de la loi de Fitts [209] (voir Section 1.2.2 équation 1.3). Notre but était d'obtenir des tâches avec une difficulté comparable à celles de l'étude de Guillon et al. [82]. En effet, les auteurs de ces travaux ont étudié les techniques d'expansion des cibles pour le pointage à distance dans le contexte de la chirurgie augmentée avec l'OrthoPilot [®][3].

Sélection des cibles Lorsque la partie mobile de l'interaction était correctement positionnée, l'élément du menu ciblé était surligné en vert. Pour confirmer la sélection,

tion, les participants pouvaient utiliser l'un des deux boutons de la pédale placée sous la table. Si la sélection était correcte, la cible disparaissait et le participant était invité à revenir à la position initiale. En cas d'erreur lors d'une tâche, le participant devait continuer avec la cible actuelle jusqu'à finir par l'atteindre.

Conception de l'étude L'expérience était divisée en 4 blocs de 30 pointages chacun, avec un bloc par technique. L'ordre des blocs était contrebalancé selon un carré latin. Par conséquent, nous avons 4 techniques de menu (2 *éléments d'interaction contrôlés* x 2 *physicalité de la cible*) x 5 cibles x 6 répétitions, conduisant à 120 pointages par participant. Une session d'entraînement était imposée avant chaque bloc. Ces essais d'entraînement étaient identiques aux tâches réelles mais rien n'était enregistré à l'exception du temps d'entraînement total. Au début de l'expérience, nous avons expliqué aux participants qu'ils devaient apprendre la position des cinq cibles autour de la table d'opération et que ces cibles resteraient les mêmes pendant toute l'expérience. Les participants étaient invités à s'entraîner avec la technique de menu actuelle jusqu'à ce qu'ils (1) connaissent les cinq emplacements cibles et (2) qu'ils se sentent efficaces et à l'aise. Un minimum de 12 pointages d'entraînement était imposé, comme illustré à la Figure 3.16.

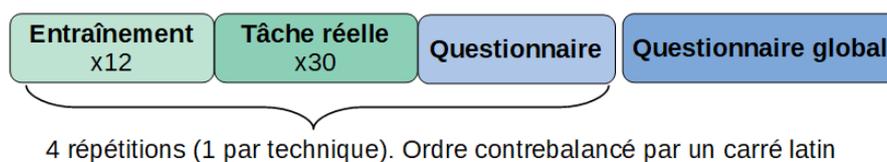


FIGURE 3.16 – Déroulement de l'étude 1.

Nous avons observé au cours d'une première étude pilote que certains participants ont ressenti une fatigue oculaire après avoir effectué plus de 60 pointages consécutifs. Par conséquent, entre chaque bloc de 30 pointages, les participants étaient encouragés à faire une pause. Après la pause, les participants ont rempli un questionnaire pour recueillir leurs commentaires sur la technique de menu qu'ils venaient de découvrir. Un questionnaire final a également été remis à la fin de l'étude pour obtenir un retour sur la comparaison de toutes les techniques.

Mesures Pour chaque tâche de pointage, le temps pour effectuer la sélection et le nombre d'erreurs commises par les participants étaient enregistrés. Lorsqu'une erreur était commise, la distance entre la position actuelle et le centre de la cible était également sauvegardée. Ci-après, cette distance est appelée *distance d'erreur*. De plus, la trajectoire de la tête était également enregistrée afin de pouvoir la comparer à la trajectoire idéale. Pour ce faire, la position de la tête était enregistrée toutes les 100 ms lors des tâches de pointage. L'objectif était de mieux comprendre les performances du participant et de distinguer les différentes phases de mouvement de la tête.

Les retours qualitatifs étaient divisés en deux types de questionnaires. Tout d'abord, les participants ont répondu après chaque bloc à un questionnaire dont les réponses étaient selon une échelle de Likert à 7 éléments. Les questions portaient sur la performance subjective (vitesse, précision) et le confort subjectif (facilité à comprendre la technique, charges physiques et cognitives). Puis, à la fin de l'étude, un

questionnaire global demandait aux participants de classer les 4 techniques de menu selon les mêmes critères (performance subjective et confort). La fatigue, l'utilité perçue des cibles mixtes et les préférences étaient également discutées lors d'entretiens semi-structurés.

Résultats

Pour analyser les résultats, nous avons utilisé une ANOVA à mesures répétées (avec $\alpha = 0,05$). Des tests post hoc ont été effectués avec des t-tests pairés et des corrections de Bonferroni. En suivant les principes de Dragicevic [61], nous avons appliqué une transformation logarithmique aux données lorsqu'une telle opération était pertinente (c'est-à-dire pour le cas d'une asymétrie positive dans les mesures de temps). Lorsque utiliser une ANOVA n'était pas approprié, nous avons utilisé un test de Friedman suivi d'un test post-hoc de Wilcoxon avec corrections de Bonferroni. Les barres d'erreur des différents graphiques représentent les Intervalles de Confiance (IC) à 95%.

13 pointages (0,9 % des 1440 pointages) ont été retirés de l'analyse en raison d'une latence due au système de suivi de la tête. Cette latence a eu lieu lorsque les marqueurs réfléchissants sur les lunettes ont été partiellement masqués par des cheveux ou lorsque certains participants ont eu besoin d'ajuster le casque sur leur tête au milieu d'une tâche de pointage.

Temps de sélection Sur l'ensemble de l'étude, nous n'avons observé aucune évidence de différence au niveau des temps de sélection avec les techniques évaluées ($F(3, 33) = 0,11, p = 0,95$), comme le montre la figure 3.17a.

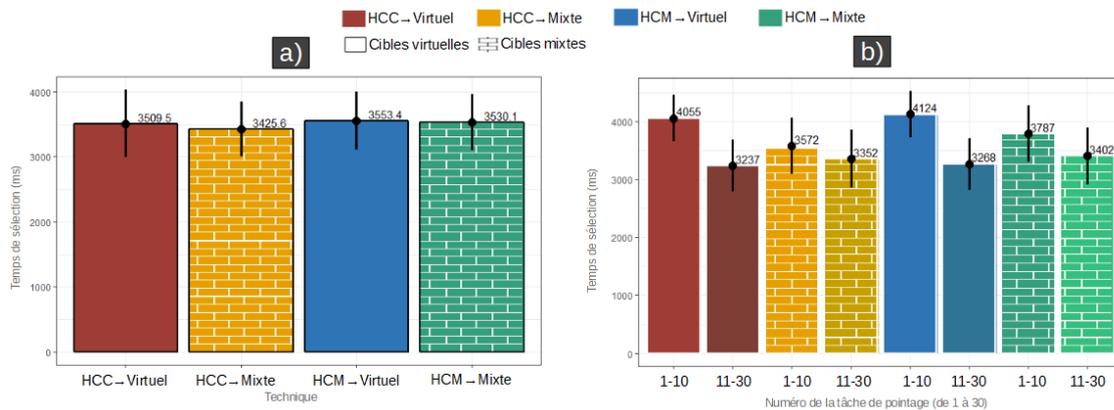


FIGURE 3.17 – Étude 1. a) Temps de sélection des cibles pour les quatre techniques de menu. b) Comparaison des temps de sélection au sein de chaque bloc : les 10 premiers pointages contre les 20 derniers. Les barres d'erreurs représentent les intervalles de confiance de 95%.

Le temps de sélection moyen était proche de 3,5 secondes pour toutes les techniques de menu. Aucune différence n'a été observée entre les techniques HCC et HCM. Cependant, nous avons observé une tendance différente en comparant les 10 premiers essais de chaque bloc avec les 20 essais restants (Figure 3.17 b). Dans ce cas, nous avons observé de bonnes preuves de différence entre les techniques

avec cibles virtuelles et les techniques avec cibles mixtes. Les preuves étaient fortes pour $HCM \rightarrow Virtuel$ ($p = 0,01$) et $HCC \rightarrow Virtuel$ ($p = 0,02$). En comparaison, nous n'avons observé aucun effet notable concernant $HCM \rightarrow Mixte$ ($p = 0,66$) et aucun effet pour $HCC \rightarrow Mixte$ ($p = 1$).

Lors de l'utilisation de cibles virtuelles, les participants ont tout de même progressé après la phase de formation en réduisant leur temps de sélection moyen de 0,84 secondes. Lors de l'utilisation de cibles mixtes, l'amélioration était limitée à 0,3 s. De plus, les temps de formation des participants étaient similaires pour toutes les techniques (aucune différence trouvée, $p = 0,48$). Les participants ont donc encore beaucoup progressé lors des tâches avec les deux techniques comportant des cibles virtuelles. Afin de contrebalancer cet effet d'apprentissage, nous avons également vérifié si les performances étaient les mêmes une fois la courbe d'apprentissage stabilisée pour toutes les techniques. Pour ce faire, nous n'avons analysé que les 20 derniers essais. Encore une fois, nous n'avons trouvé aucune preuve de différence entre les techniques ($F(3,33) = 0,2$, $p = 0,91$).

Erreurs Comme pour le temps de sélection, les résultats ne sont pas concluants sur les 4 techniques en termes d'erreurs. Plus précisément, le nombre moyen d'erreurs ($\chi^2 = 2,9$, $p = 0,41$) et les distances d'erreur moyennes ($F(3,33) = 1,73$, $p = 0,18$) n'ont mis en évidence aucune preuve de différence entre les techniques (Figure 3.18). Dans les deux cas, les écarts entre les participants nous ont également empêché de tirer des conclusions. Nous avons observé des résultats similaires en comparant le début et la fin de chaque bloc enregistré. Aucune différence n'a été trouvée entre les 10 premiers essais et les 20 derniers essais pour le nombre d'erreurs ($\chi^2 = 1,9$, $p = 0,2$) et les distances d'erreur ($F(7,77) = 0,9$, $p = 0,51$).

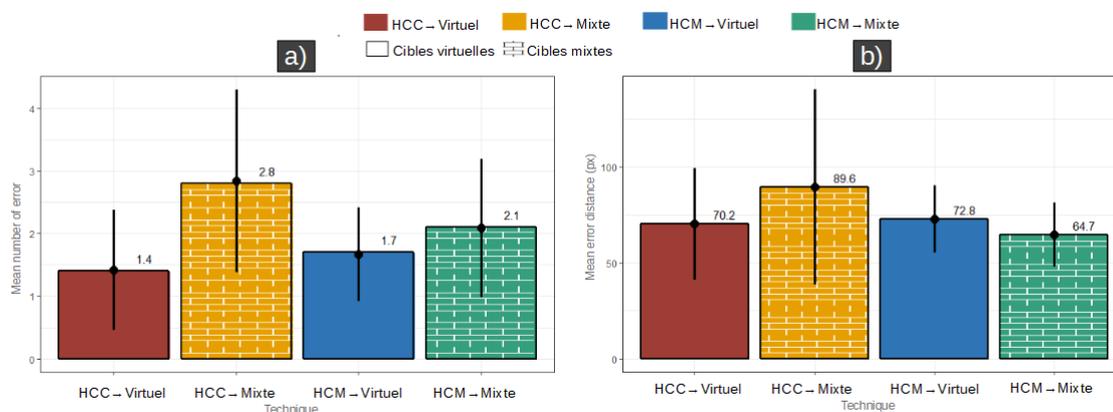


FIGURE 3.18 – Étude 1. a) Erreurs moyennes pour les quatre techniques de menu. b) Distances d'erreur moyennes pour les quatre techniques de menu. Les barres d'erreurs représentent les intervalles de confiance de 95%.

Commentaires qualitatifs Sur la base du questionnaire global à la fin de l'étude, $HCM \rightarrow Virtuel$ a été classé premier (globalement préféré). La plupart des participants (8/12) ont préféré contrôler le menu avec la tête plutôt que de contrôler le curseur. Ces participants ont déclaré préférer HCM qu'ils estimaient plus facile, plus rapide et/ou plus précis. La Figure 3.19 présente les principaux résultats qualitatifs pour chaque technique.

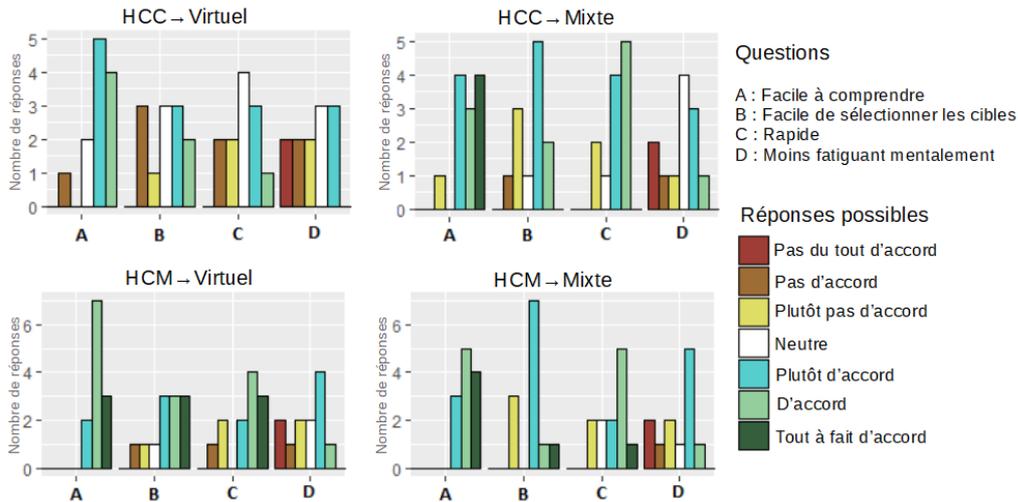


FIGURE 3.19 – Étude 1 : retours qualitatifs collectés auprès des participants.

Fait intéressant, P12 a explicitement mentionné l'avantage de voir le menu depuis le début : *"J'ai préféré déplacer le menu avec la tête (HCM) parce qu'on le voit de base et on cherche simplement la croix. C'est un peu comme si les deux autres (HCC) demandaient deux fois plus d'effort parce qu'on cherche d'abord le menu, puis l'élément"*. P11 a également souligné qu'il y avait moins d'éléments en mouvement à l'écran avec HCM (puisque le menu était fixé au centre de l'écran) et a donc trouvé HCM moins déroutant. Les 4 participants qui ont préféré les techniques HCC ont donné le même type d'explication : ils ont trouvé les techniques HCC plus faciles à utiliser, plus rapide ou se sentaient davantage en contrôle. P3 a également déclaré être plus familier avec le fait de contrôler un curseur.

Cinq participants n'ont pas perçu la différence entre HCC et HCM. Ils ont déclaré pendant la phase d'entraînement qu'il s'agissait exactement de la même chose (par exemple, faire *HCC*→*Virtual* après *HCM*→*Virtual*), et ont eu besoin de plusieurs pointages pour comprendre la différence. Pour les 7 autres participants, la différence était évidente dès les premiers essais. À la fin des tâches de pointage, 2 participants n'étaient toujours pas sûrs de la différence entre les techniques HCC et HCM. Ils ont finalement compris cette différence lors du questionnaire final grâce aux dessins de rappel représentant chaque technique fournie avec le questionnaire.

Les opinions sur les cibles mixtes et leurs ancrages physiques étaient plus diverses. D'une part, 4 participants ont déclaré qu'ils étaient vraiment utiles. En revanche, 2 participants les ont catégoriquement écartés. Trois autres participants ont indiqué qu'ils avaient utilisé des ancres au début pour connaître l'emplacement des cibles, mais pas après car ils savaient déjà où les cibles étaient positionnées. Fait intéressant, les 3 derniers participants ont déclaré qu'ils trouvaient les ancres utiles, mais que la superposition de l'ancre et des cibles créait une certaine gêne visuelle.

Discussion

Toutes les techniques de menu conduisent à des performances similaires en termes de temps de sélection des cibles, nombre d'erreurs et distances d'erreur. Ce résultat est en contradiction avec notre hypothèse H1 car nous pensions que les techniques

HCC seraient plus rapides que leurs homologues HCM. En outre, la condition HCM a été préférée par la plupart des participants. Ces résultats mettent en évidence le potentiel de HCM lors de l'utilisation de menus simples. Nous avons délibérément choisi une conception de menu simple à un niveau, avec seulement 4 éléments pour être proche du menu principal existant du système de navigation OrthoPilot®[3]. Comme l'a commenté P10, nos résultats peuvent ne pas être valables avec des menus plus complexes et à plusieurs niveaux hiérarchiques.

Partiellement en accord avec notre hypothèse H2, marquer l'emplacement des cibles virtuelles avec des objets physiques a eu un impact sur le processus d'apprentissage. Les participants ont atteint des performances optimales plus rapidement avec des cibles mixtes qu'avec des cibles virtuelles. Or, tous les participants se sont considérés à l'aise avec les techniques au bout des 12 pointages d'entraînement minimum. Nous n'avons pas observé de preuve de différence au niveau des temps totaux d'entraînement des participants. De plus, ces derniers devaient confirmer qu'ils se sentaient prêts avant chaque bloc. Il semble donc que davantage de répétitions soient nécessaires avec des cibles virtuelles pour atteindre des performances optimales. Une étude plus approfondie est requise pour déterminer l'influence des cibles mixtes sur les différentes phases du processus d'apprentissage. Scarr et al. ont proposé en 2011 un cadre d'expertise d'interface [178]. Ce cadre caractérise l'évolution des performances des utilisateurs du mode novice au mode expert et peut être appliqué aux menus. Un axe de recherche prometteur serait d'évaluer dans quelle mesure des cibles mixtes ont un impact sur les phases de *performance initiale* et *d'apprentissage étendu* au niveau des performances utilisateur [178].

Cependant, il est important de noter qu'une fois que les participants ont été suffisamment entraînés, les ancres physiques n'ont pas amélioré les performances, ce qui est en contradiction avec H2. Ce résultat peut s'expliquer par le fait que les participants connaissaient parfaitement les emplacements cibles. Ainsi, les ancres n'étaient potentiellement plus utiles. Cette explication est appuyée par les commentaires de 3 participants.

L'inconfort visuel signalé par certains autres participants peut également avoir une influence sur les performances lors de l'utilisation de cibles mixtes. L'origine de cette gêne visuelle peut être liée à un conflit de vergence-accommodation [116] exposé à la Section 2.1.3 (Figure 2.11). Ce conflit s'est potentiellement produit lorsque les participants ont fixé une des deux composantes de la cible mixte tout en accommodant sur la seconde. Par exemple, il est possible qu'ils aient concentré leur attention visuelle sur l'ancre physique tout en accommodant leur yeux vis à vis d'une distance différente (la profondeur de l'écran du casque où la cible virtuelle est affichée). De tels conflits de vergence-accommodation ont un impact négatif sur les performances et augmentent la fatigue visuelle [92]. Or, nous avons observé que les 3 participants qui ont signalé une gêne visuelle ont tous déclaré ressentir une fatigue visuelle à la fin de l'expérience. Nous avons mené une deuxième étude afin d'approfondir cette question.

3.3.2 Étude expérimentale 2 : transitions entre les modes RA et RV

Le but de cette deuxième étude est de se concentrer sur l'inconfort visuel signalé par certains participants lors de l'étude 1. Pour ce faire, nous avons conçu et mis en œuvre des versions modifiées des deux techniques de menu avec des cibles mixtes qui suppriment cette gêne visuelle. Nous avons ensuite comparé les 2 nouvelles techniques de menu avec les deux techniques de menu de l'étude 1 avec des cibles virtuelles. Le but était d'évaluer si la suppression de tout inconfort visuel permet à ces nouvelles techniques de menu avec des cibles mixtes de surpasser les techniques de menu avec des cibles virtuelles. Le protocole expérimental de cette deuxième étude était similaire à celui de la première étude. Seules les différences entre les deux protocoles sont rapportées dans cette section. 9 nouveaux participants (5 femmes, 4 hommes) de 20 à 30 ans ($sd = 3.3$) ont été recrutés pour cette deuxième évaluation expérimentale.

Techniques de menu modifiées : transitions entre mode RA et mode RV

Lors de l'utilisation de cibles mixtes, les utilisateurs doivent se concentrer d'abord sur la flèche verte virtuelle pour savoir quelle cible doit être sélectionnée, puis sur l'ancre physique, et enfin sur les éléments virtuels lorsque le menu est apparu. Cela force une accommodation supplémentaire pour les yeux et crée potentiellement des conflits de vergence-accommodation. De plus, étant donné que le menu et le curseur sont partiellement transparents, il est toujours possible de percevoir le monde réel derrière eux, ce qui pourrait gêner visuellement les utilisateurs. Lorsqu'aucune ancre physique n'est utilisée, les utilisateurs peuvent uniquement se concentrer sur les éléments virtuels et ignorer les fonctionnalités du monde réel.

Pour résoudre ce problème, nous proposons de masquer progressivement le monde réel. Plus les utilisateurs sont proches de la cible, plus la couleur d'arrière-plan du champ de vision augmenté deviendra opaque, comme illustré dans la figure 3.20. Le changement d'opacité commence dès que la cible apparaît dans le champ de vision augmenté. Ainsi, lorsque les utilisateurs ont atteint la cible virtuelle, ils ne peuvent plus percevoir aucune caractéristique du monde réel dans le champ de vision augmenté et peuvent se concentrer entièrement sur les éléments virtuels. Le but de cette approche est d'éliminer à la fois les transitions de mise au point oculaire requises et la perturbation des caractéristiques du monde réel. Conceptuellement, cela correspond à une transition progressive d'un mode de réalité augmentée (RA) vers un mode de réalité virtuelle (RV). Cependant, les utilisateurs sont toujours conscients de leur environnement physique car le monde réel reste visible dans leur vision périphérique. La séparation avec l'environnement physique est ainsi limitée dans le temps (uniquement à proximité de la cible) et dans l'espace (uniquement dans le champ de vision augmenté).

Le concept de ces nouvelles techniques peut également être vu comme un changement dans l'espace de conception d'une cible mixte à une cible virtuelle (Figure 3.11). Un aspect particulièrement intéressant de ce changement est qu'il se produit pendant l'interaction. Les utilisateurs commencent par une cible mixte pour pouvoir voir l'emplacement de la cible dans la vision périphérique. Puis, lorsqu'ils n'ont plus besoin du guidage de l'ancre physique, ils basculent vers une cible virtuelle pour

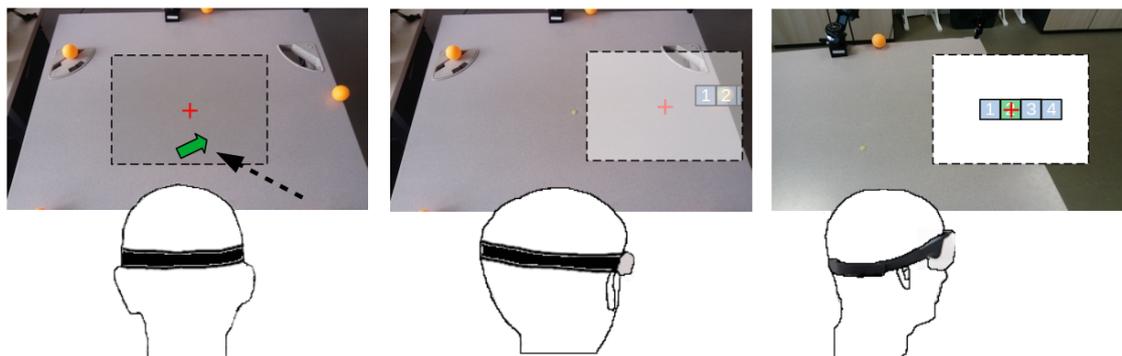


FIGURE 3.20 – Nouvelles techniques de menu avec un changement d’opacité progressif. L’utilisateur passe de la RA à un mode RV temporaire. Mécanisme illustré avec la technique $HCC \rightarrow Mixte$.

limiter les conflits d’accommodation-vergence.

Nous avons émis l’hypothèse suivante : masquer le monde réel à la fin de la tâche de pointage aiderait les participants en minimisant les conflits vergence-accommodation. Ainsi, les techniques de menu avec changement d’opacité surpasseront les techniques de menu sans ancrage physique (**H3**).

Résultats

Temps de sélection Comme le montre la Figure 3.21, nous n’avons pas observé de différence de temps de sélection des cibles pour les 4 techniques étudiées ($F(3,24) = 0.35$, $p = 0.81$). En appliquant le même processus d’analyse que pour l’étude 1, nous avons vérifié si les participants avaient atteint des performances optimales après la phase d’entraînement. Comme dans l’étude 1, nous avons comparé chaque paire des 8 conditions (4 techniques divisées en deux parties chacune, 10 premiers essais et 20 derniers essais). Cependant, nous n’avons observé qu’une faible preuve de différence ($F(7,56) = 1.8$, $p = 0.11$). Lors de l’exécution d’un test post-hoc, la seule tendance identifiée se situait entre 1) les 20 derniers essais de $HCC \rightarrow Mixte$ et 2) les 10 premiers essais de $HCM \rightarrow Virtuel$, ce qui n’est pas pertinent pour la comparaison.

Cependant, la tendance de $HCM \rightarrow Virtuel$ semble différente de celle de l’étude 1, où l’amélioration des utilisateurs pendant les essais a été clairement observée. De plus, les temps de sélection des cibles pour les deux techniques HCM semblent un peu plus élevés que pour l’étude 1 (environ 3,9s au lieu de 3,5s).

Erreurs De façon similaire à l’étude 1, nous n’avons observé aucune preuve de différence entre les techniques en ce qui concerne les erreurs. Le nombre moyen d’erreurs ($\chi^2 = 9.7$, $p = 0.2$) et les distances d’erreur moyennes ($F(3,24) = 0.46$, $p = 0.72$) étaient similaires pour toutes les techniques. De même, nous n’avons observé aucune évolution notable en comparant les 10 premiers pointages avec les 20 derniers.

Retours qualitatifs 6 des 9 participants ont perçu le changement d’opacité de l’arrière-plan comme une aide pour le pointage. Ils ont indiqué que ce changement les avait aidés à ajuster leur vitesse lorsqu’ils se sont approchés de la cible. Il est

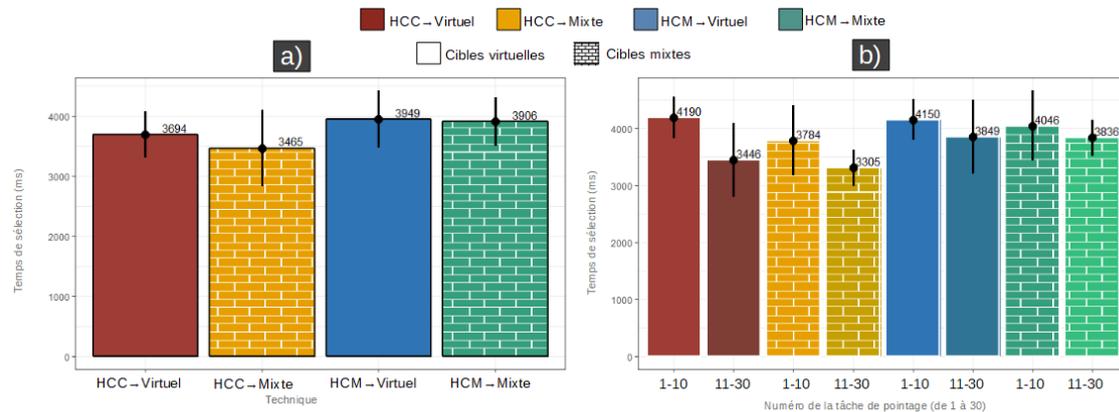


FIGURE 3.21 – Étude 2. a) Temps de sélection pour les quatre techniques de menu. b) Comparaison des temps de sélection au sein de chaque bloc : les 10 premiers pointages versus les 20 derniers. Les barres d'erreurs représentent les intervalles de confiance de 95%.

intéressant de noter que c'était le seul avantage rapporté par les participants. Le passage d'une cible mixte à une cible virtuelle n'a jamais été mentionné. Les 3 autres participants étaient plus sceptiques quant à l'impact du changement d'opacité et se sentaient neutres à ce sujet.

Les cibles mixtes ont été jugées moins utiles que le changement d'opacité. 5 participants ont indiqué que des ancrages physiques pourraient être utiles au début pour connaître l'emplacement des cibles. Ces participants ont rapporté qu'après un certain temps, ils n'utilisaient plus les ancrages physiques. Seuls 2 participants ont trouvé les ancrages vraiment utiles pendant toute l'expérience, et les 2 derniers participants ne les ont pas trouvées utiles. Cependant, aucun participant n'a signalé d'inconfort visuel.

L'hypothèse H3 ne semble pas être soutenue par les résultats. En masquant le monde réel et donc en minimisant l'inconfort visuel rapporté dans l'étude 1, nous voulions améliorer les performances des techniques de menu avec des cibles mixtes. Cependant, nous n'avons trouvé aucune preuve d'une telle amélioration. Une explication possible est que les participants connaissaient parfaitement la position des cibles virtuelles et ne considéraient plus les ancrages physiques. Les conflits de vergence-accommodation qui ont pu arriver au cours de la première étude semblent avoir disparu grâce au passage au mode RV.

3.3.3 Étude expérimentale 3 : physicalité des menus

Lors des deux premières études, nous nous sommes concentrés sur la comparaison des cibles virtuelles et des cibles mixtes. Les résultats expérimentaux suggèrent que malgré des performances similaires, les cibles mixtes ont permis aux participants d'atteindre plus rapidement des performances optimales. Mais qu'en est-il des cibles purement physiques ? Celles-ci sont toujours visibles même en dehors du champ de vision augmenté et ne souffrent d'aucun biais d'implémentation possible (i.e. aucune latence perceptible). L'objectif de cette troisième étude expérimentale est donc de s'intéresser aux cibles physiques afin d'explorer l'ensemble du second axe de l'espace

de conception basé sur la physicalité des cibles. Ce second axe est présenté à la Section 3.1.2.

L'espace de conception exploré dans cette étude 3 (Figure 3.11) met de côté une partie du premier axe de l'espace de conception. En effet, nous avons considéré dans les études 1 et 2 les deux types d'élément d'interaction contrôlable par l'utilisateur : le curseur et le menu. Contrôler un menu virtuel avec la tête (HCM) voudrait dire utiliser des curseurs comme cibles. Or, la différence entre un curseur mixte et un curseur physique est mince, comme expliqué à la Section 3.1.1. Puisque l'objectif de cette 3ème étude est de se concentrer sur la physicalité des cibles, il est plus judicieux de comparer des menus virtuels, mixtes et physiques comme cibles. Nous nous intéressons par conséquent uniquement aux techniques HCC dans cette étude.

Physicalité des menus

Plusieurs menus mixtes basés sur les principes de la RA tangible [31] (voir Section 1.2.1) ont été proposés. Par exemple, le menu étudié par Lee et al. correspond à un menu mixte puisqu'il combine des éléments virtuels liés aux faces d'un objet physique [122]. En revanche, les menus complètement physiques sont beaucoup moins courants. Ullmer et al. ont proposé le concept de *menus tangibles* : des liens numériques (comme des commandes "ouvrir" et "sauvegarder") sur des surfaces physiques (papier, bois, plastique) [197]. Des exemples de menus virtuels, mixtes et physiques sont présentés à la Figure 3.7 et rappelés à la Figure 3.22.

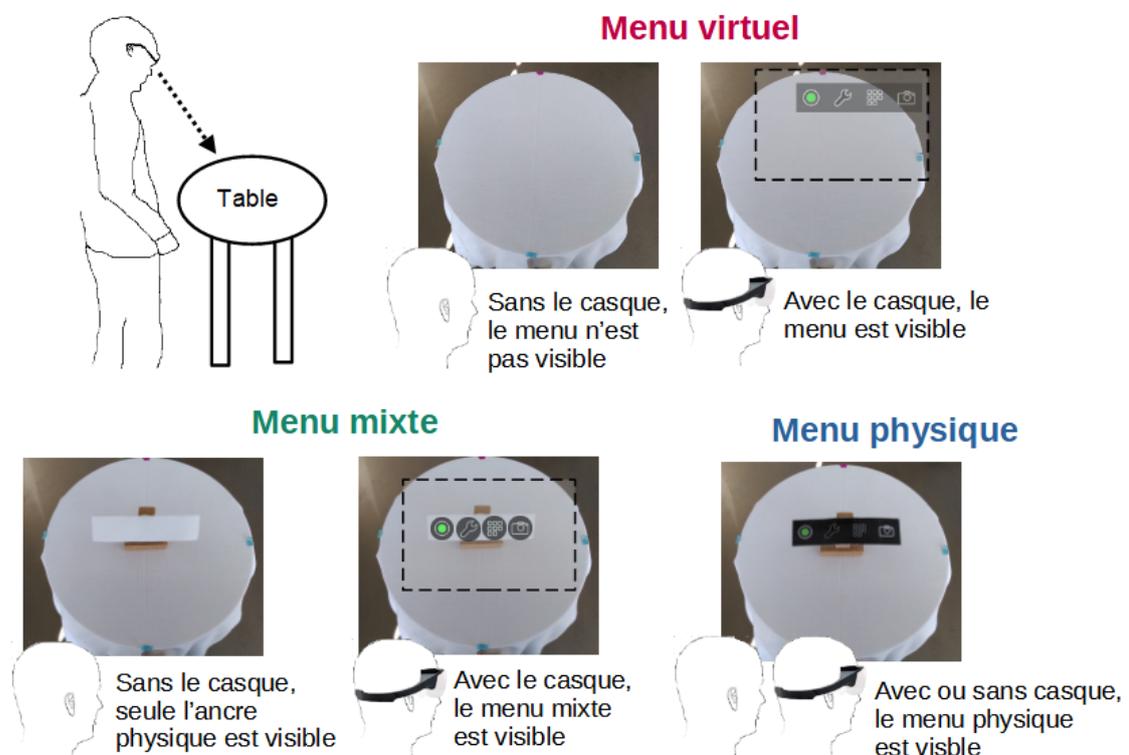


FIGURE 3.22 – Étude 3 : Les trois types de menus considérés dans notre étude : menus virtuels, menus mixtes (éléments virtuels attachés à une ancre physique) et menus physiques.

Notre approche diffère de ces travaux et des interfaces utilisateur tangibles (TUIs) [99] car nous nous concentrons uniquement sur l'incarnation physique des menus. Ainsi, nous ne nous intéressons pas ici aux modalités d'interaction en entrée impliquant la manipulation d'objets physiques. Le terme *tangibilité* n'est donc pas adapté à notre étude. Nous utilisons le terme *physicalité* pour la *physicalité* des widgets et nous désignons par *menu physique* un menu avec une représentation physique.

Utiliser des menus physiques plutôt que des menus mixtes ou virtuels peut sembler évident au premier abord. Les menus physiques ne souffrent pas de limitations techniques telles qu'un faible champ de vision, la latence ou un manque de précision du système de suivi des mouvements. Ils semblent donc idéaux comme cibles de pointage. Par ailleurs, les menus physiques sont directement et entièrement visibles dans la vision périphérique, ce qui peut aider les utilisateurs à mieux planifier leurs mouvements de tête (voir Section 3.1.2).

Cependant, les menus physiques présentent aussi des limitations principalement en terme de flexibilité par rapport aux menus virtuels ou mixtes :

- Les menus physiques reposent sur des objets physiques, dont l'ajout peut ne pas convenir à tous les contextes. Par exemple, ils ne sont pas adaptés aux environnements encombrés tels que les salles des machines ou les véhicules [89] et impliquent un environnement prédéfini n'autorisant pas l'interaction mobile [162].
- Même lorsqu'ils conviennent au contexte d'usage, les menus physiques ont un impact sur la spatialité de l'interface utilisateur car ils reposent sur des positions fixes spécifiques. Contrairement aux menus virtuels, ils ne peuvent pas être placés n'importe où. Le contrôle de l'utilisateur sur ces menus peut être limité.
- Les menus physiques ont un coût en ce qui concerne leur fabrication, leur mise en place et leur mise à jour.

Comme observé avec les menus mixtes lors de l'étude 1, les transitions fréquentes de focus visuel peuvent avoir un impact sur l'interaction avec des menus physiques en RM. Les conflits de vergence-accommodation ont un impact négatif sur les performances et augmentent la fatigue visuelle [92, 116]. La nature de la tâche à effectuer (par exemple, la fréquence des transitions visuelles, la durée de la tâche et la charge cognitive) et les caractéristiques de l'environnement (par exemple la complexité, la congestion, la physicalité et la position des objets) peuvent influencer les performances et le confort. Par conséquent, supposer que l'utilisation de menus physiques est toujours la meilleure approche n'est au final pas si évident.

Choix de conception

Nous nous concentrons sur l'aspect perceptuel de la physicalité des menus. Les aspects cognitifs tels que les métaphores d'interaction [131] et la mémorisation de commandes [75] sont en dehors du cadre de nos travaux. Ainsi, nous faisons le choix d'utiliser des ancrages physiques le plus neutres possibles, sans lien sémantique avec le menu ni métaphore particulière. Contrairement aux deux premières études, les ancrages physiques des menus mixtes ont donc cette fois la même forme que leur partie virtuelle.

Les menus virtuels, mixtes et physiques ont été implémentés de la même manière. Pour chaque condition, le participant pointait des menus virtuels dont les positions

prédéfinies étaient connues du système. Dans la condition *Menus physiques*, les menus virtuels n'étaient pas affichés à l'écran. Ils étaient néanmoins présents et alignés avec les emplacements des menus physiques. Lors des tâches de pointage, le système se basait toujours sur ces menus virtuels (masqués ou non) pour vérifier si le participant avait correctement sélectionné l'élément cible. Ainsi, les trois types de menus étaient implémentés de la même manière puisque seule la visibilité des menus virtuels différait.

Enfin, notre protocole expérimental implique également une tâche de distraction. Cette tâche a été créée : 1) pour augmenter la fréquence des transitions de focus visuel et 2) pour augmenter la charge cognitive des participants et refléter la charge cognitive d'utilisateurs mentalement occupés. Notre objectif est de quantifier la différence de performance entre les menus virtuels, mixtes et physiques car cet aspect est absent des deux études précédentes.

Conception de l'étude

L'installation expérimentale est proche de celle utilisée dans les deux premières études : le casque de RM et le système de suivi des mouvements était le même. Les participants se tenaient derrière une table carrée grise (81 cm de long, 73 cm de haut) représentant une zone de travail. Un panneau blanc (102 cm de long, 125 cm de haut) était placé verticalement derrière la table et surélevé du sol. La zone de panneau résultante au-dessus de la table (102 cm de long, 74 cm de haut) était utilisée comme zone principale pour les tâches expérimentales (c'est-à-dire où les menus apparaissent). Un aperçu de la configuration est disponible à la Figure 3.23.

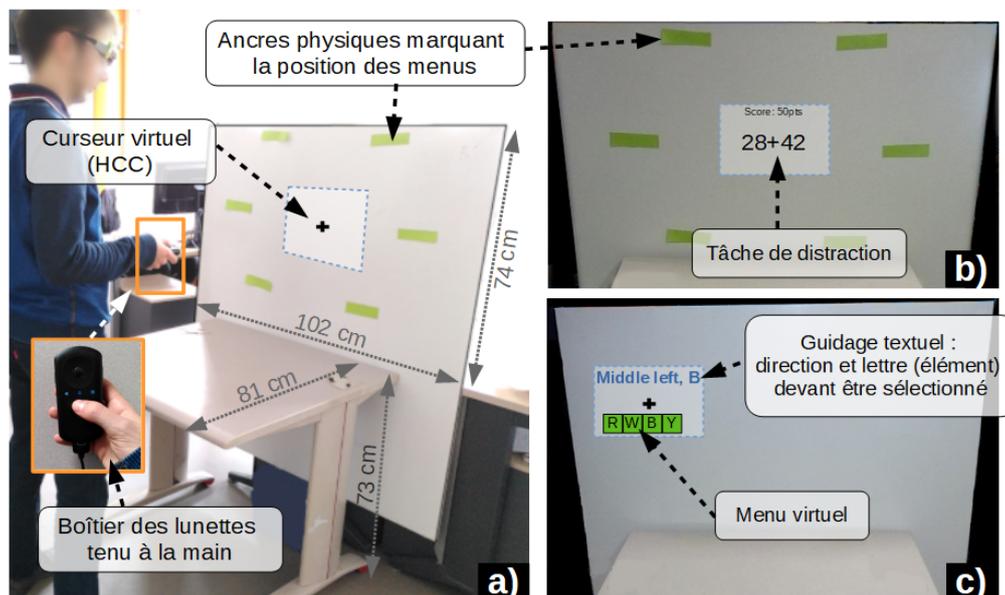


FIGURE 3.23 – Installation expérimentale pour l'étude 3. a) Participant devant l'espace de travail. b) Tâche de distraction. c) Tâche de pointage.

Des rectangles de papier (12,5 cm de long, 3,5 cm de haut) ont été utilisés comme ancrs physiques pour indiquer les emplacements des menus. Les ancrs étaient fixées au panneau blanc à l'aide de ruban adhésif transparent et fabriquées à partir de papier vert pour s'assurer qu'elles soient clairement visibles. Les ancrs physiques

correspondaient ainsi à la forme rectangulaire des menus virtuels. Les étiquettes des éléments de menu étaient des lettres. Pour les menus physiques, les lettres étaient écrites sur des feuilles de plastique transparentes à l'aide d'un marqueur noir. Ces feuilles de plastique transparentes étaient fixées et superposées à des ancrs. Les menus physiques étaient un peu plus petits que les ancrs physiques (12 cm de long, 3 cm de haut). Pour les menus virtuels, le panneau blanc était retourné afin de masquer les ancrs physiques attachées de l'autre côté. La Figure 3.24 montre une comparaison des 3 types de menus résultants.

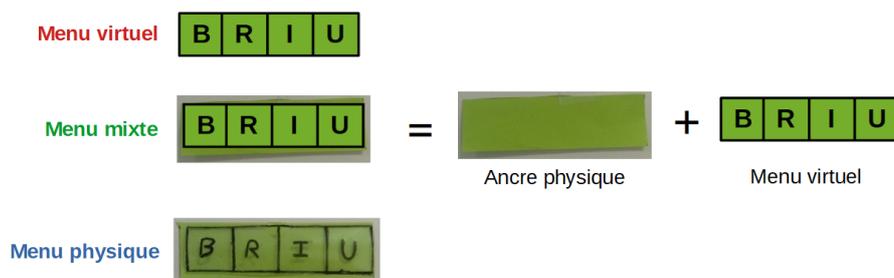


FIGURE 3.24 – Les trois types de menus comparés lors de l'étude 3.

Tâches L'expérimentation était divisée en deux types de tâches : les tâches de distraction et les tâches de pointage. Le but des tâches de distraction était d'augmenter la charge cognitive des participants pendant les essais. Cela a été fait pour refléter notre domaine d'application et la charge cognitive importante des chirurgiens. Les tâches de distraction étaient basées sur le raisonnement mathématique et la mémorisation. Pour les tâches de pointage, les participants devaient sélectionner un élément de menu. Au cours de l'expérience, chaque étape comprenait deux tâches de distraction suivies d'une tâche de pointage.

Les tâches de distraction et les tâches de pointage exigeaient que les participants utilisent le boîtier du casque BT-300 pour la validation. Au départ, nous avons prévu d'utiliser une pédale pour la validation, comme pour les deux premières études. Cependant, nous nous sommes aperçus que la pédale souffrait par moment d'une latence gênante pour l'interaction (jusqu'à 1.5s pour obtenir un retour visuel après un clic). Cette latence passagère semblait principalement due à de mauvaises performances des communications Bluetooth avec le casque BT-300 déjà occupé à gérer l'affichage 3D. Pour se prémunir de tout biais, nous avons préféré ne plus utiliser la pédale. Pour valider une sélection, les participants utilisaient à la place le bouton central du boîtier du casque qu'ils tenaient dans leur main directrice. Par conséquent, la validité de notre étude est restreinte aux tâches de pointage (excluant ainsi le mécanisme de validation), de la même manière que d'autres études précédentes où la validation n'est pas étudiée [42].

Tâche de distraction Le raisonnement mathématique est une tâche couramment utilisée pour augmenter la charge cognitive des participants lors des études expérimentales [40, 84, 97]. Plusieurs variantes de ce type de tâche peuvent être trouvées dans la littérature, y compris le calcul mental [40, 97] et la comparaison des nombres [84]. La mémorisation peut également être trouvée dans d'autres types

de tâches comme la lettre N-game dans une étude récente d'Evain et al. [69]. Dans notre étude, la tâche de distraction combinait addition, comparaison et mémorisation de nombres. Notre objectif était d'augmenter uniquement la charge cognitive des participants. Aucune habileté motrice n'était requise. En effet, nous ne considérons ici que les étapes chirurgicales de type E2 (Figure 3.2), au cours desquelles le chirurgien a une charge cognitive importante mais n'effectue pas forcément de gestes chirurgicaux sur le patient qui impliquent une habileté motrice.

Les participants ont été informés que leur objectif était de marquer autant de points que possible. Ces points étaient gagnés en répondant correctement à une succession de tâches arithmétiques. Les participants devaient calculer mentalement l'addition de nombres à 2 chiffres. Ce résultat devait être comparé au résultat de l'addition précédente et la réponse devait être donnée grâce au pavé tactile du boîtier du casque BT-300. Si le résultat R_N de l'addition courante était supérieur au résultat de l'addition précédente R_{N-1} , il fallait appuyer sur la flèche supérieure du pavé tactile. Sinon, il fallait appuyer sur la flèche du bas. Deux additions consécutives ne donnaient jamais le même résultat. Les participants devaient mémoriser le résultat courant d'une fois sur l'autre, surtout lorsque l'addition suivante devait être faite après une tâche de pointage. Ce processus est illustré à la Figure 3.25.

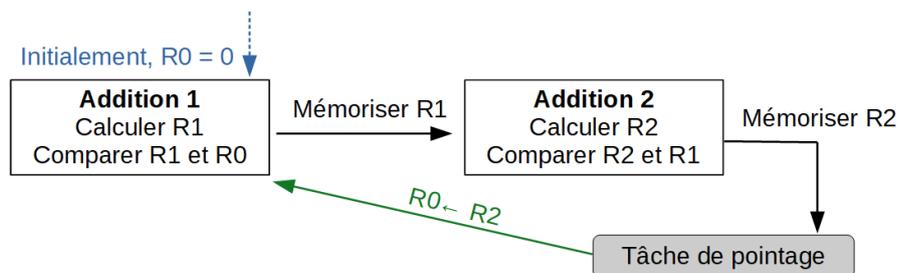


FIGURE 3.25 – Étude 3 : Déroulement d'une étape. Deux tâches de distraction sont effectuées à la suite avant chaque tâche de pointage. Le participant doit mémoriser le résultat R_N de chaque addition pour pouvoir le comparer ultérieurement au résultat R_{N+1} .

Un court message était affiché après avoir appuyé sur une des flèches afin de fournir un retour visuel aux participants. Les bonnes réponses valaient 5 points tandis que les erreurs enlevaient 30 points. Nous avons mené une première étude pilote où la tâche de distraction utilisait des nombres compris entre 10 et 99 pour les additions. Nous avons néanmoins observé que la charge cognitive induite par les tâches de distraction était trop importante. Par conséquent, nous avons décidé d'utiliser des nombres aléatoires entre 10 et 50 seulement pour les tâches de distraction.

Tâche de pointage Pour les tâches de pointage, les participants devaient sélectionner un élément dans un menu à un niveau de 4 éléments comme dans les études 1 et 2 (technique HCC, voir Section 3.1.1). Chaque élément de menu était identifié par une lettre aléatoire unique. Pour savoir quel élément sélectionner, les participants devaient suivre les instructions textuelles apparaissant en haut de l'écran du casque. Ces informations affichées comprenaient la position du menu et l'élément de menu (par exemple, milieu gauche, B), comme le montre la Figure 3.26. Les participants devaient ensuite bouger la tête pour aligner un curseur virtuel au centre de l'écran

avec l'élément de menu correspondant (technique HCC). La sélection était validée à l'aide du bouton central du boîtier du casque. Les participants avaient pour consigne d'aller le plus vite possible sans faire d'erreur. En cas de sélection incorrecte, les participants devaient continuer avec la cible actuelle jusqu'à la sélectionner.

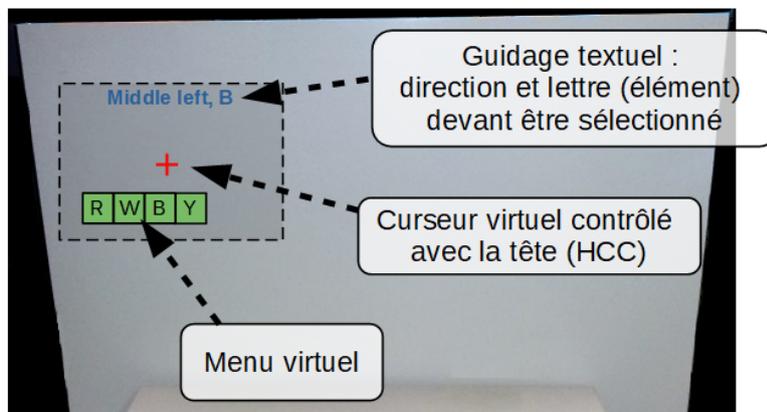


FIGURE 3.26 – Guidage textuel indiquant aux participants quelle cible sélectionner lors des tâches de pointage de l'étude 3.

Les positions des menus suivaient une version modifiée de la tâche de pointage ISO 9241-9 [100]. Six menus étaient placés sur le panneau au-dessus de la table pour former un motif circulaire, comme le montre la figure 3.23. Les participants commençaient par deux tâches de distraction au centre du cercle, puis devaient pointer vers un élément de l'un des 6 menus du cercle. Tous les menus se trouvaient en dehors du champ de vision augmenté initial au centre du cercle. La distance angulaire pour atteindre le centre des menus à partir de la position centrale était constante et proche de 33 degrés. L'ordre des cibles (menus et éléments de menu) était aléatoire. Au début de la condition *Menus mixtes*, les participants étaient informés que l'emplacement des cibles possibles était indiqué par les repères verts. Pour la condition *Menus physiques*, les participants étaient encouragés à cibler les lettres physiques. Pour les menus virtuels, le panneau blanc était retourné par le superviseur afin de masquer les ancrés et menus physiques. L'indice de difficulté des cibles a été fixé entre 4,1 et 4,2 selon l'élément de menu ciblé. Cet ID a été calculé en utilisant la formulation de Wingrave et Bowman de la loi de Fitts [209] (voir Section 1.2.2 équation 1.3). D'après Soukoreff et Mackenzie, nos tâches de pointage sont de difficulté moyenne [185], donc légèrement plus difficiles que celles des études 1 et 2.

Conception de l'étude 15 participants volontaires (2 femmes, 13 hommes) âgés de 19 à 45 ans ($sd = 6,6$) ont été recrutés pour participer à notre étude sur la comparaison des menus virtuels, mixtes en physiques en RM. Comme indiqué dans notre questionnaire préliminaire, ils étaient tous novices en RM et seuls 3 d'entre eux avaient déjà essayé de porter un casque de RM auparavant. Les participants pouvaient porter des lunettes sous le casque si nécessaire. Comme notre protocole impliquait une tâche de calcul mental (tâche de distraction), nous nous sommes assurés que tous les participants avaient une bonne connaissance des mathématiques usuelles.

L'étude était divisée en 3 blocs de 30 pointages, un bloc par condition : *Menus virtuels*, *Menus mixtes* et *Menus physiques*. Au total, il y avait 3 conditions x 6 positions de menu x 5 répétitions, soit 90 tâches de pointage par participant. L'ordre des blocs était équilibré grâce à un carré latin standard. Au début de l'étude, les participants pouvaient d'abord ajuster l'effet stéréo du casque si nécessaire. Ensuite, ils pouvaient répéter la tâche de distraction autant qu'ils le souhaitent jusqu'à être certain de bien avoir compris le principe. Avant chaque bloc, 12 étapes d'entraînement étaient imposées pour permettre aux participants de se familiariser avec la physicalité du menu. Cette phase d'entraînement était similaire aux phases enregistrées : les participants alternaient tâches de distraction et tâches de pointage. À la fin d'un bloc, les participants pouvaient faire une pause avant de répondre à un questionnaire Raw-TLX [85] pour recueillir leurs opinions sur la session effectuée. Un questionnaire final était remis à la fin de l'expérience pour obtenir des commentaires sur la comparaison de toutes les conditions.

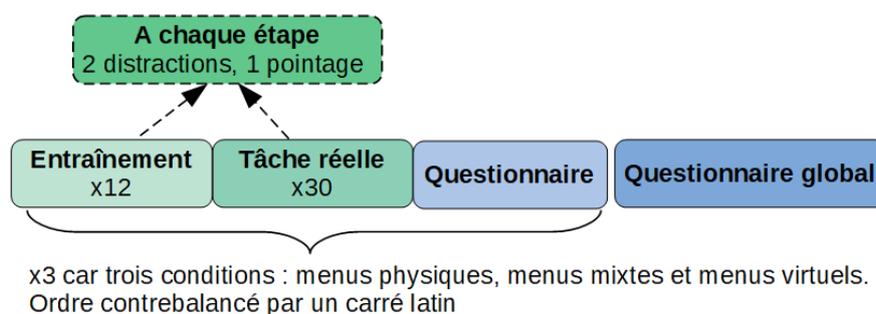


FIGURE 3.27 – Déroulement de l'étude 3

Notre objectif était d'éviter tout effet d'apprentissage sur les positions des menus. Pour se faire, un guidage textuel était affiché dès le début de chaque tâche de pointage, comme expliqué auparavant. De plus, l'expérience comprend un petit nombre de menus et une session d'entraînement conséquente. L'objectif était de permettre aux participants de se familiariser avec les tâches et les 6 menus, ce qui nous a permis d'utiliser une conception expérimentale intra-sujet [79, 132] comme dans les deux premières études.

Le temps de sélection et les nombres d'erreur ont été enregistrés à chaque pointage. En cas de sélection invalide, la distance jusqu'à l'élément de menu actuel était également enregistrée. De plus, les questionnaires Raw-TLX donnés étaient basés sur une échelle de Likert à 5 éléments (voir Annexe A). Le questionnaire final comprenait des questions sur les performances subjectives et les préférences concernant les 3 conditions physiques du menu. Les préférences du participant et l'utilité perçue des ancres étaient également discutées lors d'un entretien semi-structuré.

Résultats

L'analyse statistique a été conduite de la même façon que pour les deux premières études. Une ANOVA à mesures répétées (avec $\alpha = 0,05$) a été le principal outil statistique que nous avons utilisé pour l'analyse. Des tests post hoc ont été réalisés en utilisant des t-tests pairés et des corrections de Bonferroni. Comme suggéré par Dragicevic [61], une transformation logarithmique a été appliquée aux données

lorsque cela était pertinent (c'est-à-dire pour le cas d'une asymétrie positive dans les mesures de temps). Lorsqu'une ANOVA n'était pas appropriée, nous avons utilisé un test de Friedman à la place suivi d'un test post hoc de Wilcoxon avec des corrections de Bonferroni. Dans toutes les figures, les barres d'erreur sont des intervalles de confiance (IC) à 95 %.

Temps de sélection du menu Le temps de sélection moyen pour les menus virtuels et les menus mixtes est proche de 6s. Pour les menus physiques ce temps est pratiquement réduit de 2s, comme illustré à la Figure 3.28. Cette tendance est soutenue par l'analyse statistique, qui a fourni de bonnes preuves que la condition *Menus physiques* était différente de la condition *Menus virtuels* ($p = 0.0021$) et de la condition *Menus mixte* ($p = 0.0027$). Aucune preuve de différence n'a été trouvée entre les menus virtuels et mixtes ($p = 1$).

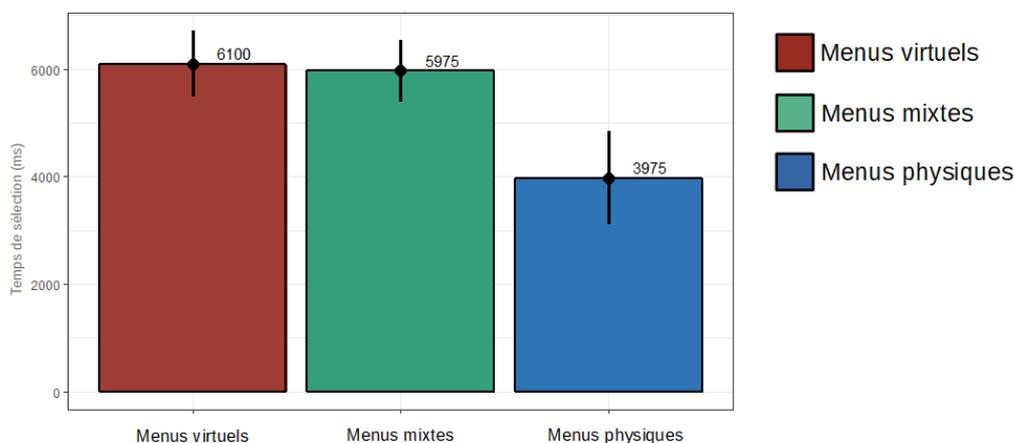


FIGURE 3.28 – Étude 3 : Temps de sélection moyens (en ms) pour chaque type de menu. Les barres d'erreur représentent les CI à 95 %.

6 secondes est un temps relativement long pour sélectionner un élément de menu avec un curseur contrôlé avec la tête. Dans les deux études précédentes, le temps de sélection moyen ne dépassait pas 4 secondes. Néanmoins, les participants n'avaient pas à réaliser une tâche secondaire pour augmenter leur charge cognitive contrairement à cette 3ème étude. Comme détaillé dans la section sur les retours qualitatifs, la plupart de nos participants ont jugé les tâches mentalement exigeantes. La charge cognitive due à la difficulté d'une tâche mathématique peut avoir un impact important sur les performances [63, 84]. Par ailleurs, l'indice de difficulté de la tâche de pointage était légèrement plus élevé dans cette étude (entre 4.1 et 4.2 au lieu d'environ 3.2).

Transfert asymétrique des compétences Les résultats ci-dessus sur les temps de sélection moyens doivent être nuancés par le transfert asymétrique de compétences observé dans notre étude. Nous avons vérifié si l'ordre des conditions pouvait affecter les performances. Comme expliqué ci-dessus, nous nous sommes assurés que la position des menus était bien comprise par les participants avant qu'ils ne commencent les essais et nous fournissons en outre des indications textuelles sur

l'emplacement du menu. Cependant, même avec des positions de menu parfaitement connues au début des essais, nous avons tout de même observé un transfert de compétences asymétrique, comme illustré à la Figure 3.29. Ce phénomène a été partiellement corroboré par notre analyse statistique. Comme suggéré par Cockburn et al. [43], nous avons vérifié l'effet principal de *l'ordre des conditions* sur les temps de sélection, et avons trouvé une preuve d'effet moyenne ($p = 0,076$).

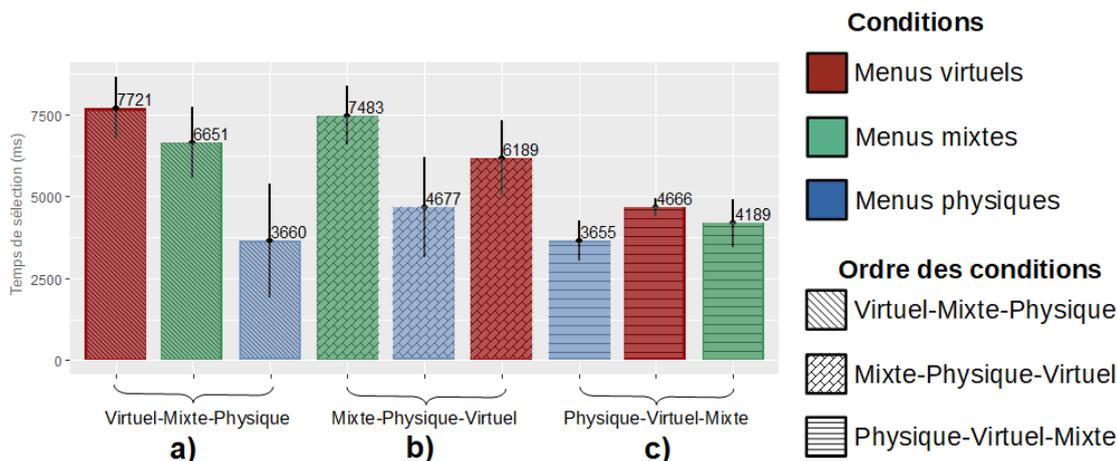


FIGURE 3.29 – Étude 3 : Transfert asymétrique de compétence. Les barres d'erreur représentent les CI à 95 %.

Nous avons trouvé des preuves notables que les participants qui ont commencé avec des menus virtuels se sont améliorés lorsqu'ils ont utilisé des menus mixtes par la suite (Figure 3.29a, $p = 0,077$). Cette tendance était beaucoup plus claire des menus virtuels aux menus physiques (Figure 3.29a, $p = 0,0026$). Lorsque les participants ont commencé avec des menus mixtes, le phénomène est plus subtil (Figure 3.29b), et nous n'avons observé qu'une faible preuve de différence entre la première condition (*Menus mixtes*) et la seconde condition (*Condition Menus physiques*, $p = 0,14$). Cependant, nous avons observé une tendance complètement différente pour les participants qui ont commencé avec des menus physiques. Dans ce cas, la performance globale des participants avec des menus virtuels et mixtes était beaucoup plus stable une fois qu'ils ont utilisé les menus physiques (Figure 3.29c, aucune preuve de différence trouvée, $p = 1$).

Par conséquent, il semble qu'il y ait eu un transfert asymétrique de compétences entre les conditions. Les participants ont pu transférer les compétences acquises à l'aide de menus physiques vers d'autres conditions. Nous avons observé des tendances similaires lors de la suppression des 10 premiers essais de chaque condition pour nous assurer qu'aucun effet d'apprentissage sur les positions de menu n'était impliqué. Néanmoins, malgré le transfert apparent de compétences, les menus physiques semblent toujours plus rapides à utiliser.

Erreurs Très peu d'erreurs ont été commises lors de l'étude, indépendamment de la physicalité des menus. Nous avons trouvé de bonnes preuves de différence pour les menus physiques, avec lesquels les participants ont commis un peu plus d'erreurs ($\chi^2 = 13,849$, $p = 0,001$). Plus précisément, le test post hoc de Wilcoxon

corrigé par Bonferroni a suggéré que la condition *Menus Physiques* était différente de la condition *Menus Mixtes* ($p = 0.011$) et de la condition *Menus Virtuels* ($p = 0,024$). Cependant, l'ordre de grandeur de cette différence reste néanmoins très faible. Comme le montre la Figure 3.30, le taux d'erreur moyen pour les menus physiques est aux alentours de 0.75%.

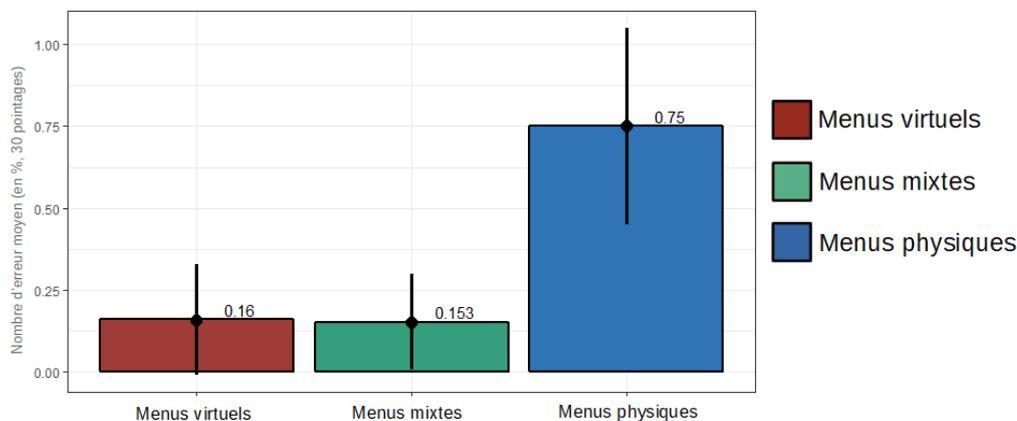


FIGURE 3.30 – Étude 3 : Taux d'erreur moyen (en %) sur 30 pointages. Les barres d'erreur représentent les CI à 95 %.

Retours qualitatifs Les questionnaires Raw-TLX donnés à la fin de chaque condition ont confirmé que les participants ressentaient une charge cognitive importante pendant les tâches. Pour chaque condition, une majorité de participants ont répondu que la tâche était au moins *Plutôt exigeante* sur le plan mental : la charge mentale était généralement classée 4/5 ou 5/5 sur l'échelle de Likert à 5 points. Des réponses similaires ont été données pour l'effort subjectif requis pour réaliser les tâches, qui a été perçu comme au moins assez élevé par la plupart des participants (indépendamment de la condition). Ces résultats ont confirmé que les tâches de distraction affectaient efficacement la charge cognitive des participants pendant les essais.

Le questionnaire final (Annexe C) demandait aux participants de classer les 3 conditions selon 5 critères : facilité d'utilisation, rapidité, précision, fatigue et condition favorite. Les résultats de ce classement sont particulièrement clairs car la condition *Menus physiques* a été systématiquement préférée aux deux autres conditions, comme le montre la figure 3.31. Ce résultat est cohérent avec les résultats quantitatifs. La condition *Menus mixtes* a été largement classée au second rang, et la condition *Menus virtuels* a été largement jugée comme la pire. Cette tendance claire entre *Menus mixtes* et *Menus virtuels* est également intéressante puisqu'elle n'est pas reflétée par les résultats quantitatifs.

Les entretiens semi-structurés indiquent que les menus physiques ont été globalement jugés plus faciles à utiliser car les éléments de menu étaient toujours visibles et parfaitement stables. 10 participants sur 15 ont trouvé que les ancrages physiques (la partie physique des menus mixtes) étaient au moins *Plutôt utiles*. La plupart de ces 10 participants ont indiqué que les ancrages les ont aidés à voir les emplacements des menus avant de commencer la tâche de pointage. 4 autres participants n'ont pas jugé les ancrages physiques utiles, expliquant que l'information textuelle affichée était suffisante pour connaître l'emplacement du menu.

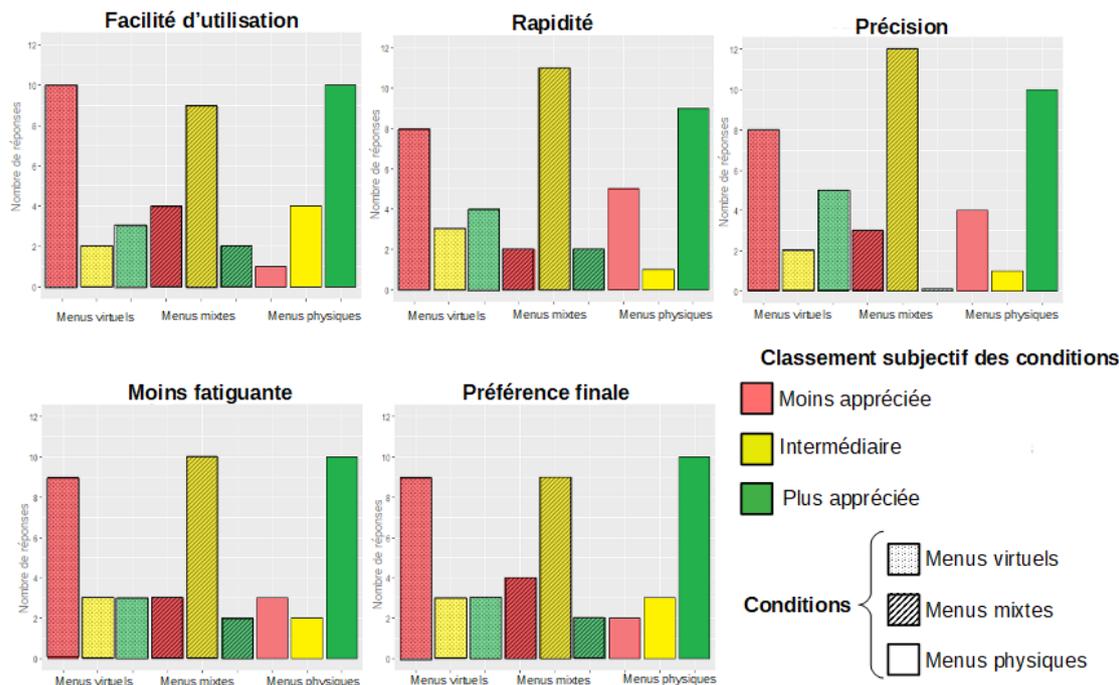


FIGURE 3.31 – Étude 3 : Retours qualitatifs pour chaque condition.

Enfin, deux opinions contradictoires sur les ancres physiques étaient particulièrement intéressantes. Un participant a expliqué *"Pour Mixte, j'ai utilisé les ancres pour aller dans le menu, j'ai ensuite cherché la lettre (élément de menu), et enfin j'ai regardé l'ancre lors de la validation de la sélection"*. Autrement dit, ce participant s'est concentré sur l'ancre physique avant de valider la sélection de l'élément virtuel du menu. Cela a été rendu possible grâce à la forme des ancres, qui était similaire à la forme des éléments virtuels. Au contraire, un autre participant n'a pas apprécié et a jugé dérangement la similitude entre l'ancre et les éléments du menu, expliquant que les ancres des menus mixtes étaient comme des "doublons". Le participant a ensuite évoqué l'idée d'ancres physiques différentes des éléments virtuels du menu, sans superposition complète. Ce participant aurait préféré de petits repères physiques montrant simplement la position des menus comme dans les études 1 et 2 au lieu d'une ancre couvrant l'ensemble des éléments du menu.

3.4 Leçons apprises et discussion

L'objectif de ces travaux était d'étudier l'utilisation des mouvements de tête pour l'interaction avec un menu en réalité Mixte. Dans un premier temps, nous avons proposé un espace de conception qui se base sur deux facteurs. Le premier de ces facteurs est l'élément d'interaction contrôlé par l'utilisateur : un curseur contrôlé avec la tête (HCC) ou un menu contrôlé avec la tête (HCM). Le second facteur de l'espace de conception repose sur la physicalité des cibles. Sur la base de l'espace de conception, nous avons ensuite conçu et évalué différentes techniques basées sur les mouvements de tête.

3.4.1 HCC et HCM

Dans l'étude 1, le curseur contrôlé par la tête (HCC) et le menu contrôlé par la tête (HCM) ont eu des performances similaires, mais HCM a été préféré par la plupart des participants. Tout en provoquant une intrusion visuelle importante, HCM permet aux utilisateurs de voir directement les éléments de menu. Cela a été possible grâce à une conception de menu simple où tous les éléments de menu étaient visibles à l'écran depuis le début de la tâche.

De tels résultats ouvrent de nouvelles perspectives sur les techniques de menu basées sur la tête. Nos travaux suggèrent qu'il peut être intéressant d'étudier des approches allant au-delà du déplacement d'un curseur affiché au centre de l'écran du casque. Deux axes de recherche se dégagent à ce sujet. Le premier consiste à étudier plus avant les techniques HCM en utilisant des menus hiérarchiques plus complexes. Le second axe de recherche concerne le déclenchement d'une technique HCM. Il pourrait en effet être intéressant d'approfondir quelle approche est la plus propice pour faire apparaître le menu à l'écran pour limiter l'occultation visuelle tout en laissant l'utilisateur voir les éléments du menu dès que nécessaire.

De plus amples études en conditions écologiques sont nécessaires pour évaluer pleinement les techniques HCM pour la chirurgie augmentée. Nos travaux sont une première étape pour mettre en valeur le potentiel de cette approche. Voir les éléments d'un menu dès le début de l'interaction est en effet intéressant pour limiter la charge cognitive de l'utilisateur. Les retours qualitatifs encourageants des études 1 et 2 laissent penser que les techniques HCM pourraient être bénéfiques pour le travail des chirurgien et méritent d'être approfondies.

3.4.2 Menus virtuels

Les menus virtuels permettent une entrée multiplexée dans le temps : ils peuvent changer leurs éléments (et donc, les commandes disponibles) au fil du temps. De plus, les menus virtuels permettent également un multiplexage spatial complet car ils ne reposent pas sur des objets physiques. Ils peuvent être placés à n'importe quelle position dans l'environnement sans créer d'encombrement physique. Les menus virtuels offrent ainsi une grande flexibilité à l'interface utilisateur. Cependant, aucun composant du menu n'est visible sans un casque de RM ni dans la vision périphérique. Les études 1 à 3 suggèrent également que l'utilisation des menus virtuels est l'approche la moins préférée. De plus, nous avons observé lors de l'étude 3 que les menus virtuels ne permettent pas aux participants de transférer les compétences acquises vers d'autres types de menus.

3.4.3 Menus mixtes

Les menus mixtes proposent un bon compromis entre les menus virtuels et physiques. Ils permettent une entrée multiplexée dans le temps car les éléments virtuels attachés aux ancres physiques peuvent évoluer dans le temps. En termes de performances, nous n'avons pas trouvé de preuve de différence entre les menus mixtes et les menus virtuels dans aucune des trois études. Cependant, sur la base des questionnaires Raw-TLX, la plupart des participants ont préféré utiliser des menus mixtes plutôt que des menus virtuels. Les menus mixtes peuvent donc être prometteurs par

rapport aux menus virtuels pour rendre l'interaction plus confortable pour les utilisateurs. Ceci est cohérent avec les travaux précédents malgré le fait que dans l'étude 3, les participants avaient une charge cognitive importante pendant les différentes tâches, comme en témoignent les questionnaires Raw-TLX. Néanmoins, la flexibilité de l'interface utilisateur des menus mixtes est limitée par leurs ancrages physiques. Les menus mixtes ne prennent donc en charge qu'un multiplexage spatial limité.

Ces résultats soulèvent aussi plusieurs questions de recherche sur le choix des ancrages physiques pour les menus mixtes, leur conception et leur impact précis sur les différentes phases du processus d'apprentissage. Ces premiers résultats expérimentaux sont néanmoins encourageants dans le cadre de notre contexte applicatif. En effet, faciliter le travail du chirurgien en lui permettant d'être efficace plus rapidement lorsqu'il interagit avec l'environnement mixte est une perspective prometteuse.

Lors de l'étude 2, les techniques étendues comprenaient une transition progressive du mode RA au mode RV masquant le monde réel. Ce masquage temporaire n'a pas eu d'effet visible sur les performances. En revanche, il a été apprécié par les participants comme un indicateur de distance restante à parcourir. De plus, il semble avoir permis d'éviter les conflits de vergence-accommodation mentionnés par certains participants lors de l'étude 1. Une autre explication possible de l'absence d'effet des ancrages physiques est qu'après suffisamment de répétitions, les participants connaissaient parfaitement les positions cibles virtuelles. Dans ce cas, les participants n'avaient plus besoin de se concentrer sur les ancrages physiques, éliminant ainsi les conflits de vergence-accommodation et leur impact sur la performance. Quoi qu'il en soit, les transitions du mode RA au mode RV semblent avoir eu un impact positif sur l'interaction.

3.4.4 Menus physiques

Les résultats expérimentaux suggèrent que le fait de pouvoir voir le menu complet dans la vision périphérique peut avoir un impact positif clair sur les performances et les préférences. En outre, le transfert asymétrique des compétences entre les conditions observées dans l'étude 3 ouvre des perspectives intéressantes pour la conception d'applications en RM. L'utilisation de menus physiques peut aider les utilisateurs à acquérir rapidement des compétences transférables sur l'interaction avec des menus via la tête. Par exemple, des menus physiques pourraient être utilisés lors de la familiarisation avec un nouveau système afin de permettre aux utilisateurs d'atteindre des performances optimales plus rapidement lorsqu'ils interagissent avec des menus virtuels. Il s'agit d'une considération importante pour des contextes spécifiques comme la chirurgie augmentée, où le temps est un facteur critique. Les chirurgiens doivent devenir efficaces avec un système le plus rapidement possible. La réduction du temps d'apprentissage pourrait ainsi favoriser l'adoption et le développement de systèmes de RM pour la chirurgie.

Outre la capacité à voir l'ensemble du menu dans la vision périphérique, un autre facteur possible expliquant les bonnes performances des menus physiques pourrait être leur réalisme visuel. Dans leurs travaux, Jansen et al. ont observé qu'une visualisation 3D physique conduisait à de meilleures performances pour la tâche de recherche d'informations qu'une visualisation à l'écran [105]. Les auteurs suggèrent que le réalisme visuel de la visualisation physique peut avoir eu un impact positif sur

les performances d'interaction. Comme le montre la figure 3.24, nos menus ont été implémentés pour être autant que possible visuellement similaires les uns aux autres. Cependant, les participants devaient effectuer des tâches de pointage 2D dans notre étude et non des tâches de recherche d'informations sur des visualisations 3D. Les indices visuels tels que les indices stéréoscopiques, d'accommodation et de texture constituent une piste prometteuse nécessitant plus de recherche pour évaluer leur impact sur l'interaction avec des menus en RM.

Les menus physiques reposent entièrement sur des objets physiques et ne permettent pas une entrée multiplexée dans le temps ou multiplexée dans l'espace. En outre, l'introduction d'objets physiques dans des environnements limités en espace est gênante, et même lorsque cela est possible, les utilisateurs peuvent ne pas avoir le contrôle sur la position des menus physiques. Ils sont également coûteux à créer et à entretenir. Par conséquent, les bonnes performances et les propriétés des menus physiques ont un coût non négligeable.

Au-delà des performances et des préférences, la physicalité de la représentation des widgets implique de considérer la flexibilité de l'interface utilisateur qui en résulte. Dans la littérature sur les interfaces utilisateur tangibles (TUI), la flexibilité est souvent déclinée sous deux aspects : le multiplexage temporel et le multiplexage spatial [73, 113, 196]. Le multiplexage temporel est basé sur un dispositif contrôlant différentes fonctions au cours du temps. Au contraire, le multiplexage spatial implique que chaque fonction a un transducteur (*transducer*) dédié à une position spatiale donnée dans l'environnement [73]. Les TUI suivent une conception multiplexée dans l'espace, ce qui les rend intuitifs à utiliser [29]. Le multiplexage temporel et le multiplexage spatial font référence aux dispositifs d'entrée manipulés. Dans le cadre de notre étude, nous interrogeons les aspects de multiplexage temporel et spatial par rapport aux menus comme méta-instruments d'entrée [22].

Sur la base de ces concepts et des résultats de notre étude expérimentale (étude 3), nous proposons un ensemble d'avantages et de limites des menus virtuels, mixtes et physiques en MR. La figure 3.32 résume ces avantages et limitations pour chaque type de menu. Notre objectif est d'inspirer de nouvelles directions de conception qui peuvent être établies en considérant des applications particulières et des environnements écologiques.

Physicalité du menu	Avantages	Limitations
Menu virtuel	Multiplexage temporel Multiplexage spatial complet Pas d'encombrement physique	Casque nécessaire pour voir le menu virtuel Ne peut pas bénéficier du monde physique Plus lent, moins apprécié que les deux autres
Menu mixte	Multiplexage temporel Préféré aux menus virtuels	Casque nécessaire pour voir les éléments du menu Multiplexage spatial limité aux ancrages physiques Mêmes performances que les menus virtuels
Menu physique	Meilleure performance Type de menu préféré Favorise l'apprentissage (transfert de compétence)	Pas de multiplexage spatial ni temporel Pas de mobilité Coût de mise en place et de mise à jour élevé

FIGURE 3.32 – Étude 3 : Récapitulatif des avantages et limitations des menus virtuels, mixtes et physiques.

Conclusion

Dans ce chapitre, nous avons présenté nos travaux sur l'interaction avec un menu reposant sur les mouvements de la tête en RM. Nous avons proposé un espace de conception sur ce type d'interaction qui nous a permis de concevoir de nouvelles techniques d'interaction basées sur un curseur (HCC) ou un menu contrôlé avec la tête (HCM). Nous avons également étudié l'impact de l'utilisation de menus virtuels, mixtes ou physiques sur l'interaction avec les mouvements de la tête.

Jusqu'à présent, nos contributions se concentrent sur les étapes chirurgicales de type E2. Le chirurgien n'a alors pas la possibilité de se servir de ses mains pour interagir. Certaines étapes E3 d'une opération laissent néanmoins une main libre au chirurgien. Dans la partie suivante, nous présentons nos contributions sur l'interaction tête + main en RM.

Chapitre 4

Interaction tête + main : techniques de menu *Bring2Me*

Sommaire

4.1	Éléments de conception	108
4.1.1	Sélection d'un menu	109
4.1.2	Déplacement du menu	110
4.1.3	Sélection d'un élément du menu	110
4.2	Conception de nouvelles techniques de menu : <i>Bring2Me</i>	110
4.3	Étude expérimentale	112
4.3.1	Protocole expérimental	112
4.3.2	Résultats	115
4.4	Leçons apprises et discussion	119
4.4.1	Tourner la tête	119
4.4.2	Techniques <i>Bring2Me</i>	119
4.4.3	Perspectives	121

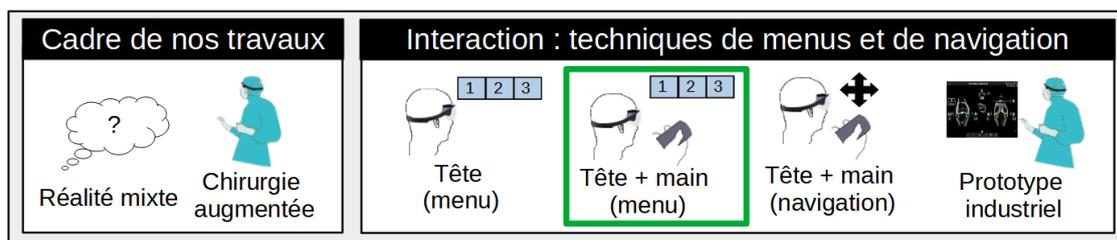


FIGURE 4.1 – Plan du mémoire (Figure B) - Focus sur le 4ème chapitre.

Introduction

Toutes les étapes d'une chirurgie n'ont pas la même criticité. Lorsqu'un chirurgien effectue des gestes médicaux sur le corps du patient, il lui est impossible d'interagir avec le système (étapes E1). Au contraire, certaines étapes laissent plus de marge de manœuvre au chirurgien (étapes E3). Ce dernier peut alors se servir de différentes techniques d'interaction basées sur les mouvements de la tête ou des mains. Dans ce

chapitre, nous nous intéressons à l'interaction en chirurgie augmentée impliquant la tête et les mains du chirurgien, comme illustré à la Figure 4.2. A l'instar du chapitre précédent, nous nous intéressons au cas des menus en Réalité Mixte (RM).

Étape	Criticité	Exemple en chirurgie orthopédique	Indisponible	Utilisable pour interagir
E1	Critique	Découpe de l'os pour placer la prothèse	Tout	Rien
E2	Importante	Placement de guides de coupe sur le corps du patient.	Mains occupées	Mouvements de tête Commandes vocales Commandes au pied
E3	Moindre	Plannification de la future position de la prothèse (étape de planning)		Mouvements de la main Mouvements de tête Commandes vocales Commandes au pied

FIGURE 4.2 – Techniques d'interaction disponibles en fonction des différents types d'étapes d'une chirurgie orthopédique.

Le bloc opératoire peut comporter de nombreux objets virtuels comme des radios et divers menus répartis dans potentiellement toute la pièce. Néanmoins, la priorité du chirurgien reste toujours le patient. Détourner son attention du patient pour regarder ailleurs est problématique car cela affecte le déroulement de l'opération et la concentration du chirurgien. Or, la taille réduite du champ de vision augmenté des casques de RM actuels force souvent les utilisateurs à tourner la tête pour pouvoir voir les objets virtuels présents dans l'environnement. Cette problématique a été détaillée dans la section 1.2.2.

Les contributions présentées dans ce chapitre visent à pallier cette problématique. Plutôt que de tourner la tête pour trouver un objet virtuel éloigné de la table d'opération, nous explorons une nouvelle approche que nous désignons par le nom *Bring2Me*. Cette approche consiste à ramener l'objet virtuel désiré dans le champ de vision augmenté. Dans ce chapitre, nous présentons la conception de 3 techniques de menu *Bring2Me* et leur évaluation expérimentale. Comme pour les trois études précédentes, nous nous concentrons sur l'interaction avec des menus. Ces contributions ont fait l'objet d'une publication à la conférence AVI'2020 [16].

4.1 Éléments de conception

Toutes les techniques de menu conçues se basent sur un même principe : ramener temporairement le menu virtuel désiré dans le champ de vision augmenté des utilisateurs. Les utilisateurs n'ont ainsi plus à tourner la tête jusqu'à l'emplacement de ce menu dans l'environnement comme avec les techniques du Chapitre 3. Seul le menu cible est temporairement déplacé pendant la durée de l'interaction (pseudo-mode) avant de revenir à son emplacement initial. Le contexte visuel initial des utilisateurs est donc préservé. Pour la conception des techniques nous considérons l'ensemble de l'interaction avec un menu, qui comprend trois étapes :

1. Sélection d'un menu donné parmi les menus placés dans l'environnement
2. Déplacement du menu sélectionné dans le champ de vision augmenté
3. Sélection d'un élément du menu déplacé maintenant visible dans le champ de vision augmenté.

Le tableau 4.1 annonce les caractéristiques de chaque étape par rapport à une technique de référence (*HeadPointing*) du chapitre précédent (technique *Head-Controlled Cursor*, ou HCC, Section 3.1.1). Les termes *Raycasting (main)* désignent une technique de raycasting basée sur le boîtier du casque Moverio BT-300 tenu à la main. Enfin, *FOV* renvoie au champ de vision augmenté de l'utilisateur (*Field Of View*). Nous détaillons ces caractéristiques de conception dans les sections suivantes.

Technique	Paramètres de conception			Propriétés	
	Sélection d'un menu	Déplacement du menu	Sélection de l'élément	Contexte visuel	Mains libres
TeleHead	Raycasting (main)	Discret (saut→FOV)	HCC	Préservé	Non
TelePad	Raycasting (main)	Discret (saut→FOV)	Raycasting (main)	Préservé	Non
Grab	Raycasting (main)	Continu	HCC + main	Préservé	Non
Head Pointing	-	-	HCC	Perdu	Oui

TABLE 4.1 – Comparaison des caractéristiques des trois techniques *Bring2Me* (*TeleHead*, *TelePad* et *Grab*) avec la technique de référence *HeadPointing* du chapitre précédent.

4.1.1 Sélection d'un menu

Afin de permettre aux utilisateurs d'apporter un menu donné dans leur champ de vision, nous nous appuyons sur la mémoire spatiale tout en préservant le contexte visuel des utilisateurs. Nous nous inspirons des travaux antérieurs sur l'interaction bimanuelle utilisant un stylet [50] et sur les techniques de pointage n'impliquant pas le regard présentées en Section 1.2.2 (Figure 1.21) [214]. Ainsi, par rapport au chapitre précédent, nous introduisons une autre modalité en entrée basée sur un outil tenu à la main. Cette modalité reflète également notre contexte applicatif avec les différents outils du chirurgien comportant un rigid-body détecté par une caméra externe [3] (voir Figure 2.3). Dans notre cas, nous utilisons le boîtier portable du casque BT-300 comme base d'une technique de raycasting. Les utilisateurs tiennent le boîtier dans leur main directrice et peuvent pointer vers l'emplacement d'un menu virtuel sans tourner la tête (et donc, sans voir le menu). La sélection du menu est validée à l'aide du bouton du boîtier. Une illustration de cette approche est disponible à la Figure 4.3.

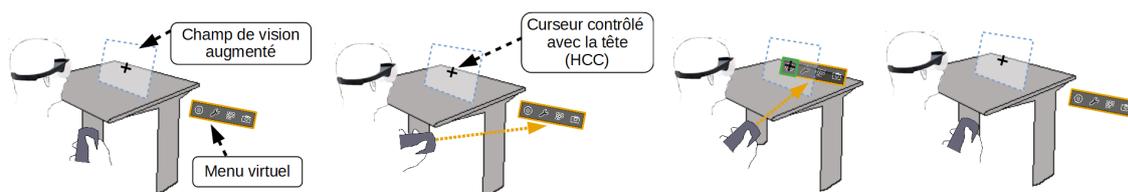


FIGURE 4.3 – Principe des techniques *Bring2Me*.

4.1.2 Déplacement du menu

Nous distinguons deux types de déplacement du menu : le contrôle *discret* et le contrôle *continu*.

Avec un contrôle discret, le menu sélectionné sera transporté instantanément jusqu'au champ de vision augmenté. Nous assimilons cette approche à un saut ou une téléportation instantanée. Des travaux antérieurs suggèrent que la perte de contrôle pendant l'interaction peut avoir un impact sur les performances et le confort de l'utilisateur en générant des problèmes de désorientation [161]. Par exemple, les systèmes de prédiction des mouvements de curseur comme le bureau Delphian [9] peuvent parfois être inexacts ou erronés, perturbant la trajectoire initiale planifiée par les utilisateurs. Cependant, dans notre cas, le comportement du menu est déterministe car il apparaîtra toujours près de la bordure supérieure du champ de vision augmenté. Nous émettons l'hypothèse qu'un tel comportement permet de réduire la désorientation des utilisateurs.

Le contrôle continu de l'emplacement du menu implique que les utilisateurs contrôlent librement la position du menu au fil du temps une fois que le menu est sélectionné. Dans ce cas, les utilisateurs doivent déplacer le menu vers leur champ de vision (geste de glisser-déposer dans les airs).

4.1.3 Sélection d'un élément du menu

Enfin, notre dernier paramètre de conception concerne la technique pour sélectionner l'élément souhaité lorsque le menu est visible dans le champ de vision. Nous nous concentrons sur trois techniques : 1) la tête en utilisant uniquement un curseur de tête 2) la main en utilisant uniquement le raycasting avec le boîtier des BT-300 et 3) une approche mixte (tête + main).

Il est important de noter que nous ne considérons dans ce chapitre que des menus virtuels. En effet, l'objectif de notre approche est de pouvoir déplacer à distance des menus cibles pour les ramener dans le champ de vision augmenté. Un tel déplacement n'est pas faisable avec des cibles physiques distantes, à moins de forcer l'utilisateur à se déplacer physiquement ou de motoriser les cibles. Utiliser des cibles mixtes est envisageable : l'utilisateur ramènerait dans ce cas uniquement la composante virtuelle du menu mixte. Cette approche n'est néanmoins pas étudiée dans ces travaux. En effet, nous concentrons sur du pointage hors écran pour aller sélectionner un menu : ancrer physiquement ces menus ne présenterait donc qu'un intérêt limité dans le cadre de notre étude. Au final, la technique *HCC*→*Virtuel* présentée au chapitre précédent est celle qui se rapproche le plus des techniques étudiées dans ce chapitre. En effet, ces deux types de techniques impliquent l'utilisation d'un curseur virtuel contrôlé par l'utilisateur et de menus virtuels utilisés comme cibles.

4.2 Conception de nouvelles techniques de menu : *Bring2Me*

Les trois techniques *Bring2Me* que nous proposons sont basées sur une combinaison des paramètres de conception exposés ci-dessus pour les trois étapes de l'interaction avec le menu. Leurs caractéristiques sont résumées dans le Tableau

4.1. Les trois techniques conçues partagent le même premier paramètre de conception : seuls les deux autres paramètres de conception varient. Les trois techniques *Bring2Me* sont décrites à la Figure 4.4.

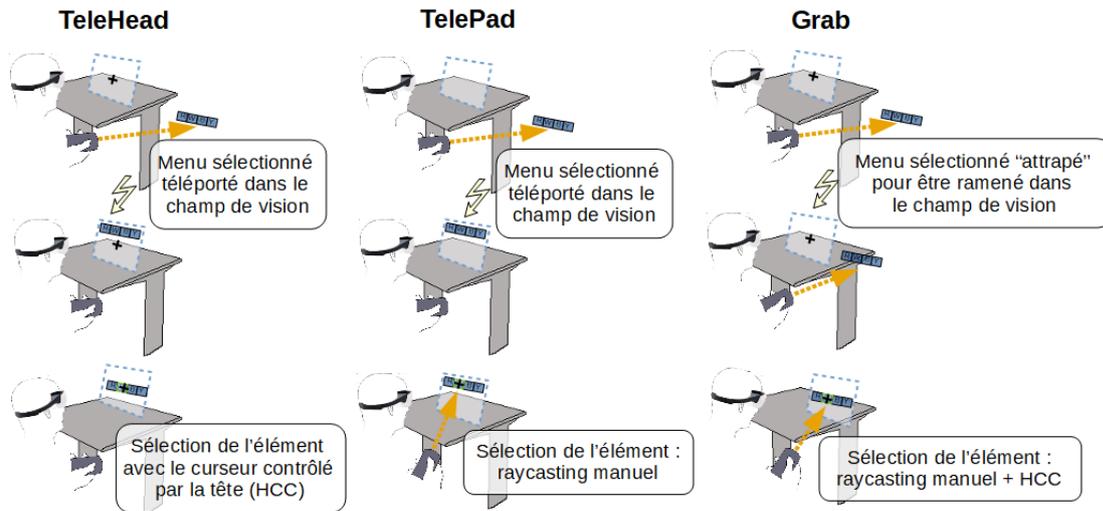


FIGURE 4.4 – Fonctionnement des techniques *Bring2Me*.

La technique *TeleHead* est basée sur un contrôle discret du menu sélectionné et sur un curseur contrôlé avec la tête (HCC) pour sélectionner l'élément de menu souhaité. Son nom traduit cette combinaison de paramètres de conception : la téléportation du menu (saut instantané) et la technique de sélection de l'élément (curseur contrôlé avec la tête, soit *head*). De même, la technique *TelePad* repose également sur un contrôle discret de la position du menu, mais elle utilise le raycasting avec le boîtier du casque pour sélectionner l'élément de menu. Par conséquent, toute l'interaction avec le menu se fait en utilisant uniquement la main dans ce cas.

La technique *Grab* est une approche mixte basée sur un geste de glisser-déposer. Comme les techniques *TeleHead* et *TelePad*, les utilisateurs doivent d'abord sélectionner le menu souhaité situé hors du champ de vision augmenté par une technique de raycasting. Cependant, ils doivent dans ce cas maintenir le bouton du pad enfoncé au lieu de simplement cliquer dessus. Tant que le bouton est enfoncé, le menu est lié au curseur du rayon du pad et les utilisateurs contrôlent librement sa position (contrôle continu). Une fois que le menu est affiché dans le champ de vision et que le curseur de la tête survole l'élément de menu souhaité, le bouton du pavé peut être relâché pour confirmer la sélection.

Nous avons éliminé la technique basée sur un contrôle continu de la position du menu et un curseur contrôlé par la tête pour sélectionner l'élément de menu. Un tel changement de modalité d'entrée briserait la métaphore unique du geste de glisser-déposer que nous voulions étudier. Contrairement à *TeleHead* et *TelePad* basés sur deux étapes de pointage distinctes (sélection du menu, puis de l'élément), *Grab* repose donc sur un geste continu pour sélectionner et déplacer le menu, puis sélectionner l'élément. Par ailleurs, les deux modalités (mouvement de la main et de la tête) peuvent être utilisées en parallèle avec cette technique.

Par ailleurs, nous avons voulu comparer ces trois techniques *Bring2Me* à une technique de référence obligeant l'utilisateur à tourner la tête pour regarder sa cible.

Pour cette technique de référence, nous avons dans un premier temps considéré une technique de raycasting standard avec un dispositif tenu à la main. Néanmoins, une telle technique impose deux contraintes : 1) perdre le contexte visuel initial en tournant la tête et 2) avoir une main disponible pour interagir. Comme nous étudions le compromis de conception entre perdre le contexte visuel et impliquer les mouvements d’au moins une main, nous avons considéré que 1) si un mouvement d’une main est possible alors l’objectif est de préserver le contexte visuel (objet d’étude de ce chapitre) 2) si le contexte visuel n’est pas préservé alors l’objectif est de ne pas utiliser des mouvements de la main (objet d’étude du chapitre précédent). Par conséquent, nous avons choisi d’utiliser une technique de raycasting basée sur les mouvements de la tête [118, 155] comme technique de référence : technique *Head-Controlled Cursor* ou HCC dans le chapitre précédent Section 3.1.1 et que nous désignons ici par *HeadPointing* en opposition aux trois autres techniques. Un curseur en forme de croix est affiché à une position fixe au centre de l’écran des lunettes et doit être aligné avec la cible en tournant la tête. Lorsqu’il est correctement placé, l’élément cible est mis en surbrillance et la sélection peut être confirmée.

Comme pour l’étude 3 (Section 3.3.3) nous avons choisi de ne pas utiliser la pédale pour confirmer la sélection des cibles avec la technique de référence *HeadPointing*. La latence occasionnelle constatée aurait en effet pu introduire un biais dans l’étude. Nous avons donc retiré la pédale et laissé les participants confirmer la sélection des cibles en appuyant sur un bouton du boîtier des lunettes tenu dans la main directrice. Par conséquent, la validité de la technique de référence *HeadPointing* est limitée aux tâches de pointage (excluant ainsi le mécanisme de confirmation).

4.3 Étude expérimentale

Nous avons conduit une évaluation expérimentale afin de comparer les trois techniques *Bring2Me* à la technique de référence *HeadPointing*. Le matériel utilisé est semblable aux études précédentes du Chapitre 3 : le casque de RA Epson Moverio BT-300 et le logiciel de suivi des mouvements *TrackingTools* (Section 3.2.2). Afin d’améliorer la détection de la position de la tête avec les caméras externes, nous avons ajouté un quatrième marqueur réfléchissant sur les lunettes BT-300. Un nouveau support avec trois marqueurs a également été conçu avec une imprimante 3D afin de pouvoir détecter la position du boîtier des lunettes.

4.3.1 Protocole expérimental

Espace de travail

Comme le montre la figure 4.5a, les participants se tenaient derrière une table carrée grise (81 cm de long, 73 cm de haut) représentant une zone de travail. Un panneau blanc (102 cm de long, 125 cm de haut) a été placé verticalement derrière la table et surélevé du sol (+ 35 cm). Près du haut de ce panneau blanc, un carré vert de 30 cm de long appelé *Contexte visuel* était affiché dans les lunettes (Figure 4.5 b). Un objet physique (ici, un os en plastique) matérialisait le centre de cette zone utilisée comme espace visuel principal lors de l’expérimentation. Notre objectif lors du positionnement du contexte visuel était de permettre aux participants de regarder

devant eux. Garder la tête baissée est en effet moins confortable et pourrait créer une fatigue supplémentaire au niveau du cou.

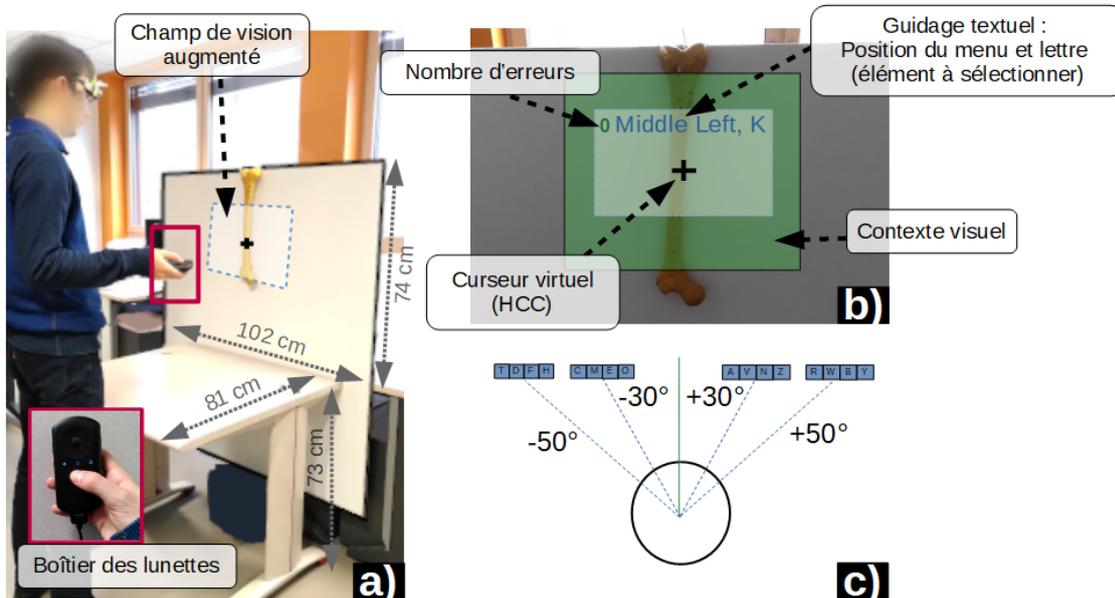


FIGURE 4.5 – Étude 4 : installation expérimentale. a) Participant devant l'espace de travail. b) Champ de vision au début d'une tâche de pointage c) Vue aérienne des distances angulaires (angles horizontaux) entre le champ de vision initial et les 4 menus.

Positions du menu

Deux critères ont guidé le positionnement des menus virtuels. Tout d'abord, nous voulions obliger les utilisateurs à changer complètement leur contexte visuel initial lorsqu'ils tournent la tête avec la technique *HeadPointing*. Cependant, nous voulions également rester dans une plage de distances angulaires permettant une précision acceptable lors de pointages sans le regard, comme l'ont étudié Yan et al. [214] (voir Section 1.2.2). Les auteurs ont observé que les meilleurs résultats de précision étaient concentrés dans une plage d'angles horizontaux de -60 à $+60$ degrés. Nous avons donc choisi 30 et 50 degrés pour les deux distances angulaires possibles entre le centre du contexte visuel et les menus. Ces distances angulaires sont incluses dans la vision mi-périphérique. De plus, entre 30 et 60 degrés (horizontalement), la vision est toujours binoculaire et les couleurs peuvent être détectées efficacement [212].

Nous avons ainsi obtenu quatre positions horizontales pour les menus : -50 , -30 , $+30$ et $+50$ degrés, comme le montre la figure 4.5c. Les menus étaient tous alignés verticalement au-dessus du panneau blanc à une distance angulaire proche de 20 degrés du centre du contexte visuel. Les quatre menus étaient désignés aux participants en fonction de leur emplacement : respectivement à gauche, milieu-gauche, milieu-droite et à droite. Pour faciliter la phase de pointage hors écran, les cibles du menu étaient étendues à l'aide de zones de "hitbox" invisibles de 80 cm de long. Aucun menu virtuel n'était visible dans le contexte visuel initial.

Tâche

Pour démarrer une tâche, les participants devaient regarder et sélectionner le contexte visuel au centre de la configuration en utilisant le raycasting avec le pad. Le curseur au centre de l'écran et le curseur du raycasting manuel étaient tous deux visibles dès le départ. Ces deux curseurs devaient être placés dans les limites du contexte visuel pour pouvoir lancer la tâche via un bouton du boîtier. Notre objectif était de nous assurer que les participants commencent dans des positions neutres au niveau de la tête et des mains. Pour les techniques *Bring2Me*, les participants devaient également continuer à regarder la zone *Contexte visuel*.

Comme lors des études du chapitre précédent, un guidage textuel indiquait le menu actuel et l'élément à sélectionner au début d'une tâche (par exemple "Milieu gauche, K" sur la figure 4.5b). Ce retour textuel visuel était affiché en haut de l'écran du casque.

Les participants devaient tourner la tête ou apporter le menu ciblé dans leur champ de vision en fonction de la technique d'interaction courante. L'élément de menu ciblé était surligné en vert lorsqu'il était survolé par le curseur au centre de l'écran ou par le curseur du boîtier. Après une sélection d'élément réussie, les participants devaient revenir à la position initiale au centre du contexte visuel avant de commencer un autre essai. En cas d'erreur, la tâche était interrompue et reprogrammée à la fin du bloc. Le nombre actuel d'erreurs commises était toujours affiché dans le coin supérieur gauche de l'écran. Vert au début (aucune erreur commise), ce nombre d'erreur devenait orange puis rouge lorsque trop d'erreurs étaient commises (rouge pour plus de 3 erreurs). Les participants étaient invités à être aussi rapides que possible tout en évitant les erreurs. Ils avaient pour consigne d'ajuster leur vitesse si ce nombre d'erreur devenait rouge.

Conception de l'étude

12 participants volontaires (5 femmes, 7 hommes) âgés de 23 à 43 ans ($sd = 5,3$) ont été recrutés pour cette étude. Comme indiqué dans notre questionnaire préliminaire, les participants étaient tous novices au sujet de l'interaction en RM. Seulement 2 d'entre eux avaient déjà essayé un système de RM auparavant. Les participants pouvaient porter des lunettes ou des lentilles de contact sous le casque si nécessaire.

L'expérimentation était divisée en 4 sections, une section par technique d'interaction : *HeadPointing*, *TeleHead*, *TelePad* et *Grab*. Chaque section était divisée en deux blocs de 16 essais, conduisant à 32 tâches de pointage par technique. Au cours de ces 32 tâches de pointage, chacun des 16 éléments de menu a été sélectionné deux fois, comme le montre la figure 4.6. Nous obtenons donc un total de 128 tâches de pointage par participant au cours de l'étude. L'ordre des sections était équilibré grâce à un carré latin tandis que l'ordre des éléments de menu était aléatoire. Après chaque section, les participants étaient invités à faire une courte pause s'ils le souhaitaient, puis devaient remplir un questionnaire Raw-TLX (Annexe A) pour recueillir leurs commentaires sur la technique qu'ils venaient d'utiliser.

L'expérience commençait par les explications de l'instructeur sur les tâches de pointage et la position des objets virtuels dans l'environnement, illustrées par des images. Après avoir ajusté le casque et l'effet stéréo si nécessaire, les participants

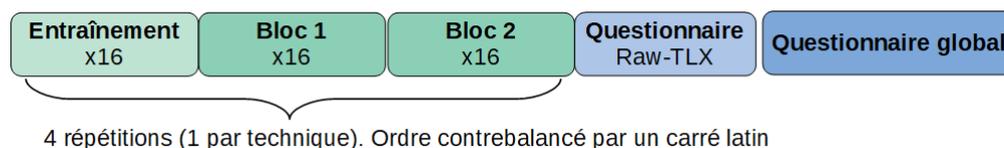


FIGURE 4.6 – Étude 4 : déroulement de l'expérimentation.

étaient invités à regarder chaque zone virtuelle (le contexte visuel et les 4 menus) et à mémoriser leurs positions. Les participants étaient informés que les emplacements des menus resteraient les mêmes pendant toute l'étude. Ils étaient libres d'explorer l'environnement mixte jusqu'à ce qu'ils se sentent en confiance. Notre objectif était d'être comparable à un scénario de *peephole pointing* avec une connaissance préalable des emplacements cibles [110] (voir Section 1.2.2). Un bloc d'entraînement de 16 essais était imposé avant chaque condition. Le but de cette longue phase de formation était de s'assurer que les participants étaient habitués à la technique d'interaction et avaient mémorisé les positions des quatre menus.

Mesures

Nous avons enregistré le temps de chaque étape des tâches de pointage. Pour la technique *HeadPointing*, nous avons enregistré le temps pour tourner la tête vers le menu jusqu'à ce qu'il devienne visible (temps de recherche avec la tête) et le temps pour sélectionner l'élément de menu (temps de sélection de l'élément). Pour les autres techniques, nous avons enregistré le temps de sélection du menu (temps de raycasting pour atteindre le menu) et le temps de sélection de l'élément. Trois types d'erreurs ont également été enregistrés :

- Erreur d'attention visuelle : détourner le regard du contexte visuel au lieu de se concentrer dessus ;
- Erreur de sélection du menu : valider une sélection de menu hors écran incorrecte ;
- Erreur de sélection de l'élément du menu.

Les trajectoires de la tête et des mains ont également été enregistrées toutes les 100 ms pour avoir une compréhension précise du comportement des participants lors des tâches de pointage. Pour l'analyse qualitative, nous avons utilisé des questionnaires Raw-TLX basés sur une échelle de Likert à 5 items. Les préférences subjectives concernant les techniques d'interaction et leur classement ont été discutées lors d'entretiens semi-structurés (Annexe D).

4.3.2 Résultats

L'analyse statistique a été menée de la même manière que pour les études du chapitre précédent. Nous avons principalement utilisé une ANOVA à mesures répétées (avec $\alpha = 0,05$). Des tests post-hoc ont été réalisés en utilisant des t-tests pairs et des corrections de Bonferroni. Lorsque l'utilisation de l'ANOVA n'était pas appropriée, nous avons utilisé un test de Friedman à la place suivi d'un test post-hoc de Wilcoxon avec des corrections de Bonferroni. Les résultats de ces tests statistiques sont rapportés ci-dessous. Cependant, comme suggéré par Dragicevic [61], nous mettons davantage l'accent sur les intervalles de confiance et la taille des

effets. Dans toutes les figures, les barres d'erreur sont des intervalles de confiance (IC) à 95 %. En cas d'asymétrie positive dans les mesures de temps, nous avons appliqué une transformation logarithmique aux données avant analyse [61].

Temps de sélection

Nous avons observé une nette évidence de différence au niveau des temps de sélection totaux entre les 4 techniques. Tourner la tête avec la techniques de base *HeadPointing* semble en moyenne 1,5 seconde plus long que *TeleHead* ($p = 2,6e-05$), *TelePad* ($p = 0,0015$) et *Grab* ($p = 0,0007$). Au contraire, nous n'avons trouvé aucune preuve de différence entre les trois techniques *Bring2Me* que nous avons proposées : les temps de sélection totaux étaient proches de 4,3 s pour chacune.

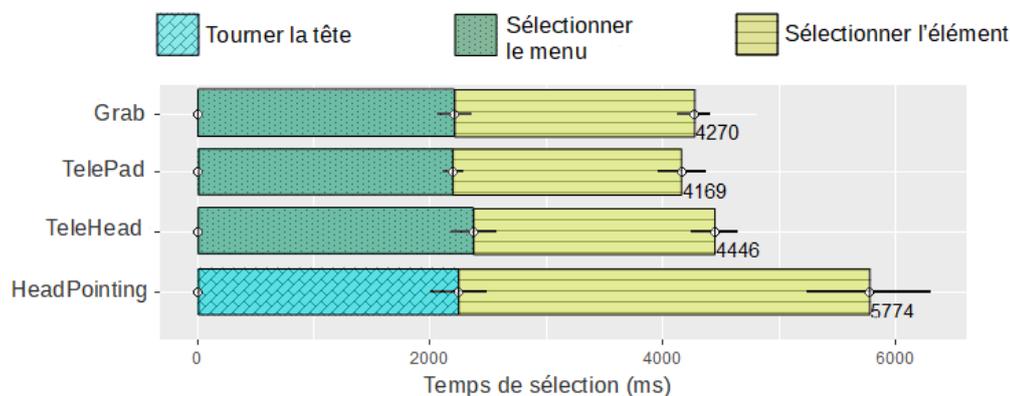


FIGURE 4.7 – Étude 4 : temps de sélection pour chaque technique. Les barres d'erreur représentent les IC à 95 %

Comme illustré à la Figure 4.7, les temps pour atteindre la position du menu en tournant la tête ou sélectionner un menu hors écran étaient également similaires. Cependant, nous avons trouvé de bonnes preuves de différence pour le temps de sélection des éléments entre *HeadPointing* et les autres techniques. Dans l'ensemble, nous n'avons trouvé aucune preuve d'une influence des positions des menus sur les temps de sélection ($p = 0,98$).

Afin de détecter tout effet d'apprentissage dans les données, nous avons comparé les deux blocs de tâches pour chaque technique (les 16 premiers pointages contre les 16 derniers). Nous avons trouvé une bonne preuve de différence uniquement pour *TeleHead* ($p = 0,006$). En nous concentrant sur cette technique particulière, nous avons observé une tendance suggérant que la différence serait principalement due aux temps de sélection des éléments de menu (comparaison des deux blocs d'essais : $p = 0,035$).

Erreurs

Nous avons observé un écart important entre les participants sur le nombre d'erreurs commises au cours de l'étude, comme illustré à la Figure 4.8. Dans l'ensemble, il est difficile de tirer une conclusion claire. Cette observation est soutenue par notre analyse statistique, avec laquelle nous n'avons trouvé aucune preuve claire de différence entre les techniques ($\chi^2 = 5,74$, $p = 0,13$). Cependant, en comparant

uniquement les erreurs liées aux sélections d'éléments de menu, nous avons observé que *HeadPointing* entraînait plus d'erreurs que *TeleHead* ($p = 0,025$), *TelePad* ($p = 0,02$) et *Grab* ($p = 0,012$). En particulier, aucune erreur n'a été commise lors de la phase de sélection des éléments de menu avec *Grab* lors des deux blocs enregistrés. Des erreurs ont bel et bien eu lieu avec cette technique lors des phases de sélection d'éléments, mais uniquement lors du bloc d'entraînement.

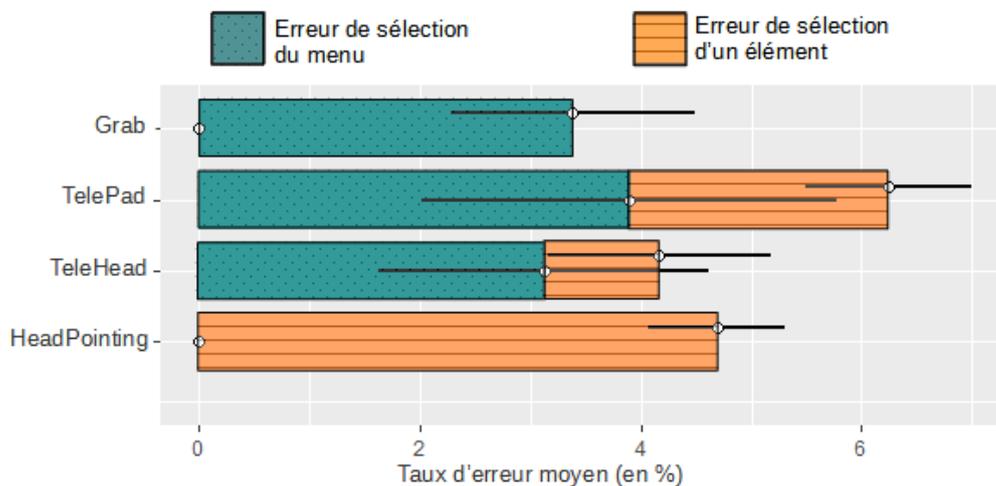


FIGURE 4.8 – Étude 4 : taux d'erreur moyens (en %) pour chaque technique. Les barres d'erreur représentent les IC à 95 %

Sans surprise, le nombre d'erreurs lors de la sélection du menu via raycasting semble être similaire entre les techniques *Bring2Me* (*Grab*, *TelePad* et *TeleHead*). Ce résultat est corroboré par notre analyse statistique : aucune preuve de différence n'a été observée ($p = 1$). Seules deux erreurs d'attention visuelle ont été commises pendant toute l'expérience, ce qui rend ce type d'erreur négligeable.

Trajectoires de la tête et de la main

Avec la technique *Grab*, deux modalités d'interaction étaient disponibles en même temps lors de la phase de sélection des éléments (*i.e.* une fois le menu attrapé). L'analyse des trajectoires de la tête suggère que seulement 2 des 12 participants ont utilisé des mouvements de tête pour sélectionner des éléments avec la technique *Grab*. Pour les 10 autres participants, aucune preuve d'utilisation parallèle des deux modalités disponibles n'a été trouvée.

Par ailleurs, nous avons également étudié le temps qu'ont mis les participants pour commencer la phase de sélection d'un élément une fois que le menu est devenu visible dans le champ de vision. Ce temps était proche de 0.5s pour les trois techniques *Bring2Me*.

Retours qualitatifs

Une tendance claire a émergé à la fois des questionnaires qualitatifs et des entretiens semi-structurés : tourner la tête est l'approche la moins appréciée par les participants, comme le montre la figure 4.9. *HeadPointing* a été jugée plus fatigante physiquement et plus frustrante. Fait intéressant, les participants P10 et P11 ont

déclaré avoir trouvé cette technique plus "*simple et intuitive*" lors des explications initiales données par l'instructeur. Ces deux participants s'attendaient donc à préférer *HeadPointing*. Néanmoins, ils ont ensuite déclaré avoir changé d'avis après quelques essais d'entraînement. Tous les participants ont apprécié pouvoir ramener le menu dans leur champ de vision. Six participants ont qualifié cette approche d'*utile* et les 6 autres de *très utile*.

Lorsque nous leur avons demandé de comparer les techniques *Bring2Me* lors des entretiens semi-structurés, plusieurs tendances ont été rapportées par les participants. Premièrement, le saut de menu instantané implémenté par *TeleHead* et *TelePad* a toujours été jugé meilleur (moins frustrant et moins fatigant) que de tourner la tête. Seul P9 a signalé que "*il n'était pas toujours évident de savoir où le menu apparaîtrait exactement*". Fait intéressant, P2 et P7 ont rapporté que la téléportation "*créait une rupture*" par rapport au mouvement continu avec *Grab*.

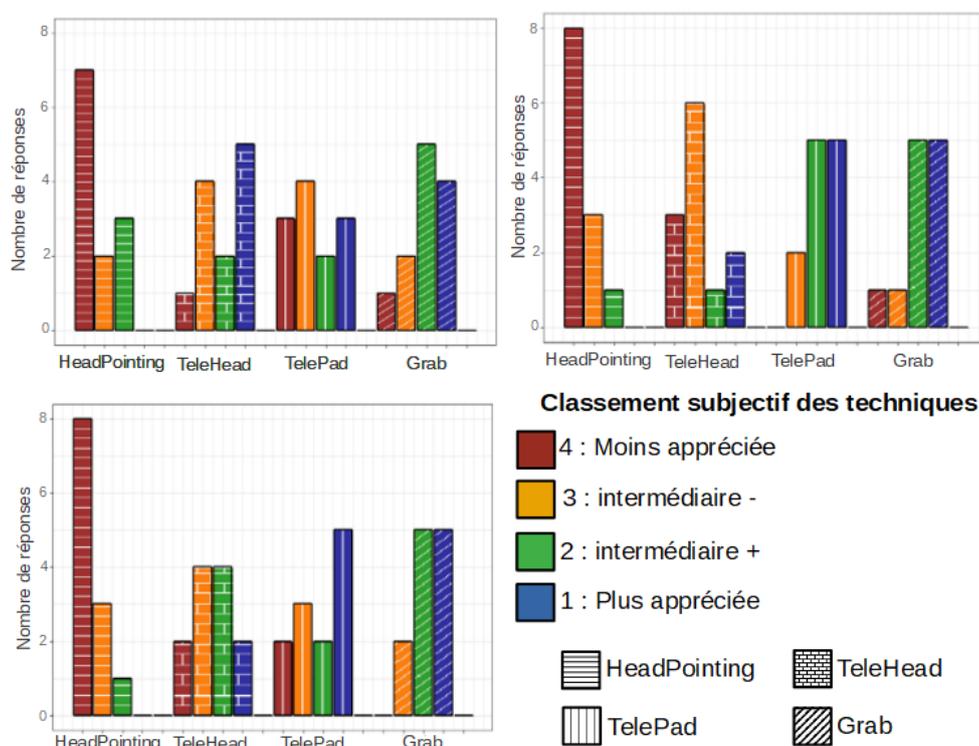


FIGURE 4.9 – Étude 4 : retours qualitatifs pour chaque technique.

5 participants ont choisi *Grab* comme technique préférée, expliquant que cette technique était plus facile, plus intuitive et plus précise. P12 a également signalé qu'il appréciait d'utiliser à la fois la tête et la main pour sélectionner des éléments de menu avec cette technique. Cependant, 3 participants ont également signalé que le geste continu était plus fatigant physiquement que la téléportation du menu, en particulier pour les cibles les plus éloignées. P1, P4 et P6 ont également expliqué que leur main devait restée crispée sur le bouton du pad pour éviter tout erreur, ce qui n'était pas confortable.

4.4 Leçons apprises et discussion

4.4.1 Tourner la tête

Les résultats expérimentaux suggèrent clairement que tourner la tête pour voir le menu et interagir avec lui est la moins bonne approche parmi les techniques d'interaction que nous avons évaluées. En effet, *HeadPointing* était la technique la plus lente et la moins appréciée, même si c'était la plus simple à expliquer aux participants. Cependant, les temps pour rendre le menu visible (en tournant la tête pour le voir ou en le faisant sauter vers le FOV) étaient similaires. Deux raisons peuvent expliquer cette similitude. Premièrement, les participants connaissaient parfaitement les emplacements des quatre menus virtuels grâce aux longues phases d'entraînement. Dans ce cas, ils n'avaient alors pas besoin d'explorer l'environnement pour découvrir des cibles. Deuxièmement, la sélection du menu avec les techniques *Bring2Me* était facilitée en utilisant l'expansion de la cible et le guidage textuel.

La différence de performance entre les techniques repose principalement sur les temps de sélection des éléments. Lors de l'utilisation de *Headpointing*, les participants n'ont aucun contrôle sur l'emplacement du menu virtuel et doivent pointer vers une cible distante tournant la tête vers lui. Au contraire, les techniques *Bring2Me* modifient l'emplacement des menus et peuvent temporairement les rapprocher le plus possible du participant. Les sélections peuvent ainsi être facilitées car les cibles sont plus grandes (voir Section 1.2.2). Il s'agit d'un avantage non négligeable des techniques *Bring2Me* par rapport à la technique de référence. Dans notre configuration expérimentale, la différence de taille visuelle était néanmoins limitée.

Au-delà de la différence de performance, le retour qualitatif négatif sur le fait de tourner la tête doit être souligné. Notre protocole expérimental n'impliquait pas une tâche de distraction nécessitant une attention visuelle de la part des participants (contrairement à l'étude 3, Section 3.3.3). Mais même sans une telle tâche de distraction, tourner la tête et perdre le contexte visuel initial a été jugé fatigant, frustrant et dérangeant. De tels résultats mettent en évidence le potentiel des techniques *Bring2Me* lorsqu'une main est disponible au moment de l'interaction.

4.4.2 Techniques *Bring2Me*

Dans l'ensemble, les trois techniques *Bring2Me* ont des performances similaires, mais *TeleHead* était moins apprécié que *TelePad* et *Grab*. Les raisons évoquées par les participants sont principalement le manque de stabilité et le manque de familiarité du pointage avec la tête par rapport au pointage avec la main. La technique *Grab* est en effet plus proche de l'utilisation habituelle d'une souris. Cette tendance au sujet des préférences des utilisateurs confirme les travaux précédents [118, 124] (voir Section 1.2.2). Kyto et al. ont constaté que les participants préféraient utiliser un dispositif tenu à la main pour affiner précisément la sélection plutôt que le pointage de la tête [118]. De même, Lee et al. ont observé que le raycasting manuel surpassait le raycasting basé sur la tête en termes de temps de complétion des tâches et de taux d'erreur [124].

Par ailleurs, les techniques *Bring2Me* impliquent par conception de diviser l'interaction en trois étapes (sélection du menu, déplacement et sélection de l'élément). Néanmoins, seuls P2 et P7 ont explicitement reporté un "sentiment de rupture"

avec *TeleHead* et *TelePad* lorsqu'ils devaient d'abord téléporter le menu, puis dans un second temps sélectionner l'élément. Décomposer l'interaction avec le menu ne semble donc pas poser problème pour une majorité de participants. Or, cette décomposition comporte également des avantages, notamment pour la gestion des erreurs. Avec les trois techniques *Bring2Me*, la sélection du mauvais menu a des conséquences limitées car il suffit de fermer le menu ou d'effectuer une autre sélection pour échanger le menu indésirable avec celui ciblé. Ce n'est pas le cas avec des techniques de menus où l'utilisateur sélectionne en une seule action continue un élément [125] (*HeadPointing*) et doit donc annuler son action avant de recommencer sa sélection.

L'approche mixte : **Grab**

Il convient de noter que la technique *Grab* a été perçue comme l'une des techniques les plus précises de notre configuration. Les erreurs commises lors de la phase de sélection du menu étaient similaires entre les trois techniques *Bring2Me*. Cependant, aucune erreur de sélection d'élément n'a été commise avec *Grab* lors de la phase de sélection d'éléments (en dehors des phases d'entraînement). Le contrôle continu de la position du menu semble donc être un paramètre de conception important dans des contextes comme la chirurgie augmentée où la précision et la sensation de contrôle sont des facteurs clés [37]. Le contrôle continu permet plus de flexibilité en laissant les utilisateurs contrôler la position du menu pendant la phase de sélection des éléments, ce qui peut par exemple limiter l'occultation visuelle dérangeante par rapport à *TeleHead* et *TelePad*.

De plus, *Grab* est aussi la seule technique permettant d'utiliser deux modalités d'entrée en même temps pour sélectionner des éléments de menu. Il est possible à la fois de déplacer le menu avec la main et d'amener le curseur au centre de l'écran vers l'élément ciblé. Cependant, notre protocole expérimental a empêché les participants de détourner le regard du contexte visuel. Ce choix de conception limitait l'amplitude des mouvements possibles de la tête. L'analyse des trajectoires suggère que seuls 2 participants sur les 12 ont profité de la possibilité de coordonner les mouvements de la tête et de la main. Des recherches supplémentaires sont donc nécessaires pour évaluer l'impact de l'utilisation parallèle des deux modalités d'entrée. En particulier, il serait intéressant d'approfondir le compromis entre :

1. Laisser les utilisateurs tourner librement la tête pour accéder plus rapidement au menu saisi ;
2. La perte partielle du contexte visuel initial, qui impacterait l'intérêt premier de la technique *Grab*.

Sauts instantanés des menus

Contrairement aux techniques *Grab*, *TeleHead* et *TelePad* n'offrent qu'un contrôle discret sur l'emplacement du menu. Lorsqu'il est sélectionné, le menu effectue un saut instantané vers son nouvel emplacement, puis reste fixe à cette position jusqu'à ce qu'un élément soit sélectionné. Ce mécanisme crée une occultation visuelle, mais le comportement du menu est déterministe. La capacité de prédire où le menu apparaîtra peut avoir limité tout effet de désorientation dans notre étude. Aucun effet de désorientation n'a été observé et un seul participant a déclaré ne pas être sûr de l'endroit où le menu apparaîtrait. Cette hypothèse est cohérente avec des travaux

antérieurs comme le *Jumping Menu* d'Ahlstroem et al. [5]. Avec le *Jumping Menu*, le curseur saute au premier élément du sous-menu lorsqu'un clic de souris est détecté à l'intérieur d'un élément parent. Les auteurs ont observé que si certains participants trouvaient ce système de menu quelque peu irritant, la plupart des participants le jugeaient meilleur qu'un menu standard. Ces participants avaient de meilleures performances avec le *Jumping Menu* une fois familiarisés avec le système [5].

4.4.3 Perspectives

Emplacement des cibles

Dans notre installation expérimentale, la phase de sélection de menu pour *Tele-Head* et *TelePad* peut être réalisée sans bouger le bras : il suffit d'utiliser les rotations du poignet pour accéder aux menus. Cependant, cela peut ne pas être le cas dans un environnement plus réaliste avec des cibles tout autour des utilisateurs. La plupart des participants à notre étude expérimentale étaient droitiers et 2 d'entre eux ont indiqué que le menu le plus à droite (+50 degrés horizontalement) était plus difficile à sélectionner. Ces participants ont expliqué qu'ils étaient proches de la limite de l'amplitude de rotation de leur poignet. Au-delà de cette limite, un mouvement du bras peut donc devenir nécessaire pour pointer vers d'autres menus. Des recherches supplémentaires sont donc nécessaires pour étudier les techniques *Bring2Me* dans des environnements plus complexes. Par exemple, il serait intéressant d'avoir une plus grande amplitude de position pour des cibles comme $-180^\circ/+180^\circ$ horizontalement et $-60^\circ/+60^\circ$ verticalement dans [214]. Des amplitudes plus importantes pourraient également être utilisées pour étudier plus en détail l'utilisation parallèle de deux modalités d'entrée (tête + main) pour *Grab* pendant la phase de sélection des éléments.

Généralisation des résultats

Dans ce travail, la chirurgie augmentée est un scénario d'application prometteur et concret, car les chirurgiens ont besoin de techniques RM préservant le contexte visuel (Figure 4.2). Cependant, nous avons choisi de mener une étude en laboratoire sans conditions écologiques afin d'obtenir des résultats généralisables. D'autres domaines avec des tâches visuelles très exigeantes telles que la fabrication [154], l'assemblage/maintenance de machines [216] et les cockpits d'avion [74, 160] pourraient également bénéficier de notre travail. De plus, les résultats suggèrent que même sans tâche visuelle exigeante, tous les participants ont préféré ramener le menu dans leur champ de vision plutôt que de tourner la tête. Par conséquent, les techniques *Bring2Me* pourraient également être utilisées dans des scénarios plus généraux d'interaction MR avec des widgets si une main de l'utilisateur est libre.

Les techniques *Bring2Me* doivent être étudiées dans d'autres scénarios avec des conceptions et des positions de menu plus complexes. De plus, elles doivent être comparées aux techniques existantes impliquant le regard et les mains de l'utilisateur [19, 118]. Ces techniques pourraient être plus rapides et plus faciles à utiliser que les techniques *Bring2Me*, mais au prix de la perte du contexte visuel initial. Nous pensons que vérifier l'une des deux propriétés (préserver le contexte visuel et avoir les mains libres) impacte nécessairement les performances d'interaction. Le choix de conception concernant ces deux propriétés dépendra du domaine d'application.

Dans notre cas, ce choix dépend de la disponibilité des mains du chirurgien comme expliqué à la Figure 4.2. La Figure 4.10 résume les avantages et les limitations des techniques afin d'aider les concepteurs dans leurs choix.

Paramètres	Techniques	Avantages	Limitations
Tourner la tête	<i>HeadPointing</i>	Pas besoin des mains pour interagir Facile à expliquer aux participants	Perte du contexte visuel initial Technique la plus lente et la moins appréciée
Ramener le menu dans le champ de vision	Techniques <i>Bring2Me</i> , <i>TeleHead</i> , <i>TelePad</i> et <i>Grab</i>	Contexte visuel initial préservé Préféré par tous les participants Menu plus proches des utilisateurs : taille visuelle++	1 main requise pour interagir
Contrôle discret de la position du menu	<i>TeleHead</i> , <i>TelePad</i>	Saut instantané : pas d'effet de désorientation observé Limite la fatigue de la main par rapport au contrôle continu	Pointage en deux étapes : d'abord le menu, puis l'élément Impossible de contrôler la position du menu après le saut
Contrôle continu de la position du menu	<i>Grab</i>	Peut placer le menu n'importe où : évite ou limite l'occlusion visuelle	Fatigue physique supplémentaire au niveau de la main
Sélection basée sur la tête	<i>TeleHead</i>	Facile à effectuer une fois le menu dans le champ de vision	Changement de modalité entre la sélection du menu et de l'élément Plus de temps requis pour atteindre une performance optimale
Sélection basée sur la main	<i>TelePad</i>	Même modalité pour la sélection du menu et de l'élément	Fatigue physique supplémentaire au niveau de la main
Sélection mixte (tête + main)	<i>Grab</i>	Possible d'utiliser les deux modalités en parallèle Technique la plus précise pour la sélection de l'élément	Seulement 2 participants ont utilisé les deux modalités en parallèle

 Rendre le menu visible
  Déplacer le menu
  Sélectionner l'élément ciblé

FIGURE 4.10 – Étude 4 : Bénéfices et limitation des techniques étudiées.

Conclusion

Cette étude contribue à résoudre un problème important dans l'interaction en RM avec un casque : la perte du contexte visuel qui se produit lorsque les utilisateurs tournent la tête pour voir et interagir avec des widgets virtuels. Ce phénomène est courant car la taille du champ de vision augmenté est limitée avec les casques actuels. Dans ce chapitre, nous nous concentrons sur les techniques préservant ce contexte visuel dans le cas de l'interaction avec les menus en faisant l'hypothèse qu'une main de l'utilisateur est disponible. Les techniques *Bring2Me* que nous proposons permettent aux utilisateurs de ramener des menus virtuels dans leur champ de vision. Ces techniques tirent parti de la mémoire spatiale, sont basées sur le pointage hors écran par raycasting et prennent en compte l'ensemble de l'interaction du menu, de la sélection du menu à la sélection des éléments dans le menu.

Les résultats suggèrent que le pointage avec un curseur contrôlé par la tête est en moyenne 1,5 s plus long que les techniques *Bring2Me*. Le pointage basé sur les mouvements de la tête (Chapitre 3) ne doit donc être utilisé que lorsque les mains ne sont pas disponibles. Sinon, lorsque : 1) au moins une main peut être utilisée au moment de l'interaction et 2) le contexte visuel des utilisateurs doit être préservé, les techniques *Bring2Me* définissent une approche prometteuse et sont préférées par tous les participants de l'étude expérimentale.

Chapitre 5

Interaction tête + main : techniques de navigation dans une scène 3D en RM

Sommaire

5.1	Éléments de conception	128
5.1.1	Couplage des vues	128
5.1.2	Navigation : changement de point de vue	130
5.2	Conception de nouvelles techniques pour changer le point de vue	132
5.3	Études expérimentales	133
5.3.1	Installation expérimentale	134
5.3.2	Étude pilote : exploration informelle de l'espace de conception	135
5.3.3	Deuxième étude : approfondissement de la navigation virtuelle	137
5.4	Leçons apprises et discussion	143
5.4.1	Stratégies	143
5.4.2	Vue de référence	144
5.4.3	Transitions entre RA et RV	145

Introduction

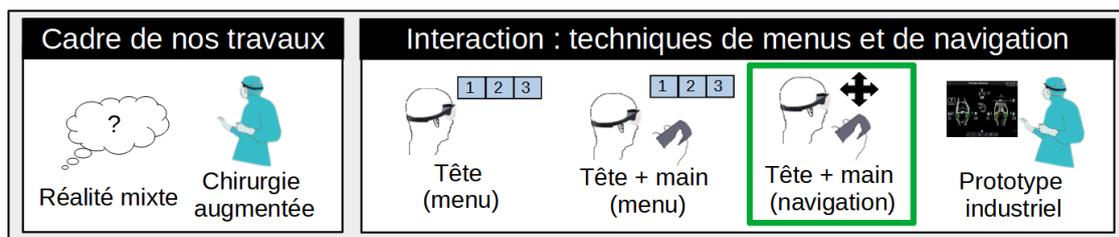


FIGURE 5.1 – Plan du mémoire (Figure B) - Focus sur le cinquième chapitre.

Au sein de ce chapitre, nous nous concentrons sur la tâche d'exploration dans une scène 3D en RM. Comme dans le Chapitre 4, nous considérons les étapes de chirurgie laissant davantage de marge de manœuvre au chirurgien (étapes E3). Nous considérons que le chirurgien peut interagir avec le système en utilisant des techniques basées sur les mouvements de la tête et/ou de la main (Figure 5.2).

Étape	Criticité	Exemple en chirurgie orthopédique	Indisponible	Utilisable pour interagir
E1	Critique	Découpe de l'os pour placer la prothèse	Tout	Rien
E2	Importante	Placement de guides de coupe sur le corps du patient.	Mains occupées	Mouvements de tête Commandes vocales Commandes au pied
E3	Moindre	Plannification de la future position de la prothèse (étape de planning)		Mouvements de la main Mouvements de tête Commandes vocales Commandes au pied

FIGURE 5.2 – Copie de la Figure 4.2. Techniques d'interaction disponibles en fonction des différents types d'étapes d'une chirurgie orthopédique.

L'originalité de ces travaux est de considérer des techniques de navigation appliquées à une scène en RM qui combine plusieurs vues : une vue mixte et une vue virtuelle. Notre objectif est d'étudier quelles techniques permettent à un utilisateur de changer de point de vue afin d'explorer et de comprendre la scène en RM présentée. Pour ce faire, nous considérons deux facteurs de conception : 1) le type de techniques de navigation pour changer le point de vue (action physique ou action virtuelle) et 2) le couplage des vues (fortement couplées vs faiblement couplées) lors du changement de point de vue. Il est à noter que le couplage des vues est bien compris dans les systèmes graphiques classiques à vues multiples [18, 174], mais reste à étudier en RM avec un casque quand il y a à la fois une vue mixte et une vue virtuelle. Par exemple, changer le point de vue sur la vue mixte pose le problème des incohérences visuelles. D'une part, les vues virtuelles ne sont pas directement liées aux caractéristiques du monde physique. Le point de vue de la scène peut ainsi être librement ajusté sans conséquences. D'autre part, la vue mixte a un lien spatial fort avec l'environnement réel puisque le contenu virtuel est directement superposé à l'objet physique. Changer le point de vue à partir duquel cette augmentation virtuelle est vue créerait une incohérence visuelle avec l'objet physique et briserait la métaphore de l'objet augmenté.

Pour illustrer une scène en RM combinant une vue mixte et une vue virtuelle, nous présentons un scénario concret dans notre domaine d'application. Pour cela, nous considérons l'étape de planning chirurgical, qui est une étape de type E3 (Figure 5.2). Au début d'une opération du genou, le chirurgien va planifier la future position de la prothèse sur la jambe du patient. Pour cela, il doit s'assurer que la prothèse remplira correctement son rôle articulaire, que la jambe soit en flexion ou en extension. L'étape de planning chirurgical sert donc à définir avec précision la position et l'orientation de la prothèse sur la jambe en flexion et en extension. La Figure 5.3a montre l'interface graphique utilisateur affichée par le système OrthoPilot®[3]. Lors de cette étape de planification, le chirurgien doit ajuster la position future de la prothèse pour les deux états de la jambe : extension (partie gauche de l'écran) et flexion (partie droite de l'écran). Toute mise à jour de valeur affecte la

position de la prothèse sur les deux états de jambe. Par exemple, la modification de l'orientation sur la jambe en extension a un impact sur la distance entre le fémur et le tibia sur la jambe en flexion (mise en évidence par des flèches rouges à la Figure 5.3a).

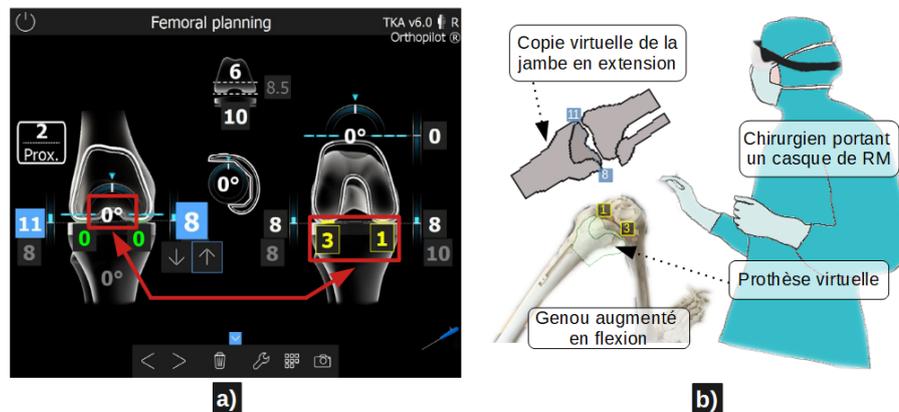


FIGURE 5.3 – Étape de planning chirurgical. a) Sur un écran distant : capture d'écran de l'OrthoPilot®[3], image courtoisie d'Aesculap. La jambe en extension est représentée sur la gauche de l'écran et la jambe en flexion sur la droite. b) Planning chirurgical envisagé en RM.

Cette étape est particulièrement intéressante à étudier car elle est un parfait exemple d'une tâche d'analyse de scène en RM. En portant un casque de RM, le chirurgien peut visualiser les coupes prévues et la prothèse directement sur le genou du patient (vue mixte du genou en flexion à la Figure 5.3b). En parallèle, le chirurgien peut également voir une représentation virtuelle du genou avec la prothèse (vue virtuelle du genou en extension sur la figure 5.3b). Au final, le champ opératoire devient une scène 3D complexe, avec plusieurs vues en RM que le chirurgien doit inspecter, éditer puis valider. La présence de vues multiples (*spatial multiplexing*) est ici nécessaire. En effet, il faut éviter de bouger la jambe du patient, de sorte que passer physiquement de la flexion et l'extension n'est pas approprié. De plus, la vue de la jambe en flexion et celle de la jambe en extension ne sont pas indépendantes l'une de l'autre. La cohabitation spatiale est donc requise afin de pouvoir comparer visuellement les deux vues. Cet exemple de l'étape de planning chirurgical correspond à notre domaine d'application, mais il est important de noter que des scénarios dans d'autres domaines existent également. Ainsi, la Figure 5.4 présente un exemple de combinaison d'une vue mixte et d'une vue virtuelle en RM pour une tâche d'assemblage. Il ne s'agit donc pas d'un sujet niche à portée limitée.

Pour étudier les techniques de navigation dans une scène combinant une vue mixte et une vue virtuelle, nous détaillons les deux éléments de conception considérés (couplage des vues et technique de navigation) dans la section suivante. Sur la base de cet espace de conception à deux axes, nous détaillons ensuite la conception de 4 techniques d'interaction que nous avons évaluées expérimentalement. Après avoir discuté des résultats obtenus, nous concluons en proposant des directives de conception pour l'exploration d'objets 3D avec un système multi-vues en RM avec un casque.

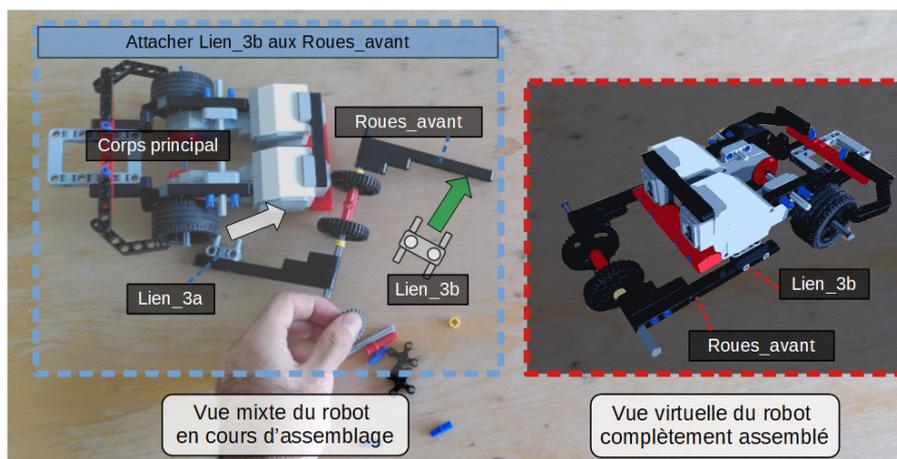


FIGURE 5.4 – Exemple de combinaison d’une vue mixte et d’une vue virtuelle pour une tâche d’assemblage d’un robot.

5.1 Éléments de conception

5.1.1 Couplage des vues

Baldonado et al. ont défini un système à vues multiples comme un système utilisant "*deux ou plusieurs vues distinctes pour soutenir l'enquête sur une seule entité conceptuelle*" [18]. Lorsque les vues affichent le même objet, par exemple la jambe du patient avec la prothèse de la Figure 5.3a, le terme *multi-forme* peut également être utilisé [174, 179]. Les systèmes à vues multiples permettent aux utilisateurs de visualiser des objets d'intérêt sous différentes perspectives. Cela peut améliorer leur performance et réduire leur charge cognitive [18] ou faciliter la collaboration entre plusieurs utilisateurs [117]. De nombreuses approches ont été explorées dans la littérature [179], y compris vue d'ensemble + détail, focus + contexte et interfaces de zoom [44].

Une notion clé des systèmes à vues multiples est la coordination des vues. Cette notion de coordination reflète comment les changements sur une vue donnée affectent les autres. D'une part, la création de liens et de correspondances entre différentes vues peut aider à découvrir des relations imprévues dans les données [18]. Par exemple, Weaver a proposé un système dans lequel les utilisateurs ont un contrôle précis sur la façon dont les actions de navigation et de sélection dans une vue affectent les autres vues [204]. D'autre part, deux vues peuvent être complètement indépendantes (vues faiblement couplées), par exemple une fenêtre personnelle séparée d'un espace de travail partagé [117]. De nombreux travaux antérieurs dans le domaine de la visualisation ont étudié ce concept de coordination entre plusieurs vues [179]. Basé sur ces travaux en visualisation, le domaine de l'analyse immersive vise à incorporer des outils d'analyse pour faciliter la compréhension des données et la prise de décision [137]. Pour ce faire, des technologies comme la réalité virtuelle (VR) sont prometteuses pour permettre aux utilisateurs d'explorer des données multidimensionnelles. Les environnements virtuels sont appropriés pour les systèmes à vues multiples car l'espace de travail visuel n'est pas limité à un périphérique physique. L'ensemble de l'environnement entourant l'utilisateur peut être utilisé pour afficher plusieurs objets et visualisations [21].

En RA, peu de travaux ont étudié des vues multiples au sein d'une scène 3D augmentée. Cordeil et al. proposent un scénario avec deux vues coordonnées (fortement couplées) : une visualisation 3D virtuelle flottant au-dessus du bureau couplée à une vue 2D sur un écran physique [48]. De même, Reipschlager et al. ont présenté une station de travail en RA pour la création de modèles 3D [173]. Les objets créés sont néanmoins des copies virtuelles, de sorte que les utilisateurs ne peuvent travailler qu'avec plusieurs vues virtuelles de l'objet modélisé. Une autre approche étudiée par Au et al. est l'intégration de plusieurs vues 2D sous forme de miroirs virtuels pour mieux comprendre une scène 3D [11]. L'objet physique 3D de la scène n'est pas augmenté en soi. En outre, les techniques d'interaction pour explorer dynamiquement la scène ne sont pas abordées dans ces études. Les études en RA ne considèrent donc pas la combinaison de vues mixtes et virtuelles en RM.

Notre étude se concentre sur l'exploration d'objets en RM basée sur la combinaison d'une vue mixte (l'objet augmenté) et d'une vue virtuelle (une copie virtuelle de l'objet), comme illustré en Figures 5.3b et 5.4. Cette notion de copie virtuelle est similaire au concept de Voodoo Doll proposé par Pierce et al. : une copie virtuelle portative d'un objet physique présent dans l'environnement [165]. Cette copie virtuelle peut être déplacée, orientée et même mise à l'échelle à l'aide de la manipulation manuelle afin d'explorer l'objet initial. De telles manipulations sont prometteuses pour l'exploration d'objets 3D complexes en RM. Cependant, cette approche encourage les utilisateurs à se concentrer uniquement sur la copie virtuelle pour explorer l'objet. Le monde physique est principalement utilisé pour créer les poupées virtuelles [165]. L'objet physique n'étant pas modifiable par le système, il n'y a par conséquent pas de couplage des vues.

Plus étroitement lié à nos travaux, Tatzgern et al. ont exploré des techniques d'interaction pour des points d'intérêt du monde réel en RM [193]. Ils se concentrent sur la manipulation d'une copie virtuelle d'un bâtiment physique. Selon la technique, la copie peut être soit superposée au bâtiment physique (état sur place), soit spatialement séparée de celui-ci. Dans ce dernier cas, les liens visuels entre l'objet physique et sa copie virtuelle sont affichés et les utilisateurs peuvent basculer entre les points de vue discrets de la même manière que [190]. Ce système combine ainsi une vue mixte (c'est-à-dire le bâtiment physique et ses liens vers la copie virtuelle) avec une vue virtuelle en RA. Les auteurs ont souligné l'importance d'avoir une vue d'ensemble rapide de l'objet et ont également observé que les participants préféreraient que la copie virtuelle soit superposée à l'objet réel au lieu d'en être spatialement séparée [193], ce qui met en évidence le potentiel des objets mixtes. Ces travaux soulignent l'apport de vues multiples mais ont été conduits en RM sur support mobile. Contrairement aux études basées sur la RM mobile [193] et des supports multiples [173], nous nous concentrons sur la RM avec un casque.

Nous retenons un élément de conception identifié en visualisation qu'est le couplage des deux vues. Nous constatons que les systèmes de RM qui combinent plusieurs vues :

- gèrent des vues complètement découplées notamment lorsque la vue est physique.
- gèrent des vues fortement couplées lorsque ces vues sont virtuelles.

Dans nos travaux, nous considérons deux types de couplage : les vues fortement couplées et les vues faiblement couplées. Les vues fortement couplées impliquent

que tout changement dans une vue est appliqué aux autres vues. Au contraire, avec des vues faiblement couplées, les modifications d'une vue n'affectent pas les autres vues. Nous ne considérons que la coordination des points de vue entre les vues. Ce type de coordination est appelé *navigation slaving* [18]. En d'autres termes, seul le point de vue à partir duquel l'objet est vu peut différer entre les vues faiblement couplées. Les autres caractéristiques visuelles des différentes formes de l'objet (par exemple les couleurs et l'échelle) sont toujours les mêmes dans les différentes vues.

5.1.2 Navigation : changement de point de vue

Notre deuxième facteur de conception est la technique utilisée pour changer le point de vue. Le fait d'avoir plusieurs vues permet aux utilisateurs de visualiser l'objet cible sous différentes perspectives. Pour avoir encore plus de flexibilité lors de l'exploration de cet objet, il est également crucial de laisser les utilisateurs contrôler le point de vue à partir duquel l'objet est vu. Cette navigation peut être continue ou limitée à des positions discrètes. Un exemple de navigation continue peut être trouvé dans les travaux de Berard et Louis [26]. Les auteurs ont proposé un appareil à perspective corrigée déplaçable (*Handheld Perspective-Coupled Display*, abrégé en *HPCD*) qui donne la sensation d'avoir un objet 3D virtuel à l'intérieur d'une sphère physique. L'utilisateur peut manipuler cette sphère physique portable et la faire pivoter librement pour explorer l'objet 3D affiché à l'intérieur de la sphère [26]. Contrairement à cette approche, la navigation discrète repose sur un ensemble fini de positions existantes. Sukan et al. ont proposé un système de RM avec une tablette portable dans laquelle les utilisateurs peuvent enregistrer des positions spécifiques appelées instantanés dans la scène 3D [190]. Il est alors possible de sélectionner une position précédemment enregistrée pour visualiser la scène à partir de cette position sans se déplacer physiquement.

La comparaison de la navigation discrète et continue sort du cadre de cette étude. Nous nous concentrons uniquement sur la navigation discrète. Ce choix a été motivé par notre domaine d'application, la chirurgie augmentée. Les chirurgiens sont habitués à visualiser des données médicales en trois dimensions (comme des radios par exemple) à partir d'un ensemble de plans anatomiques [210]. Utiliser des points de vue discrets de référence est donc approprié dans un tel contexte. Cela nous permet également de mieux contrôler les stratégies d'exploration et les modèles de navigation possibles afin de pouvoir les comparer.

Sukan et al. ont comparé une technique de navigation physique à une technique de navigation virtuelle en RM [190]. Les auteurs ont observé que pour une tâche d'alignement d'objets, la technique de navigation virtuelle (sélection de positions dans la scène 3D) était plus rapide, plus précise et était préférée par les participants plutôt que la technique de navigation physique (déplacement physique dans la scène 3D) [190]. Cependant, la navigation physique peut également être bénéfique pour les utilisateurs [181]. Simpson et al. ont comparé deux techniques de navigation pour explorer une visualisation de données en RV. Les auteurs ont observé que se déplacer physiquement autour de la visualisation pourrait être plus approprié que déplacer virtuellement la visualisation avec un contrôleur pour les utilisateurs ayant de faibles capacités spatiales [181]. Les capacités spatiales et la charge cognitive peuvent ainsi influencer le choix des techniques de navigation. Cependant, ces résultats sont à vérifier dans le cas de l'exploration d'un objet 3D en RM basée sur un casque. L'étude

expérimentale proposée par les auteurs était basée sur une tâche d’alignement avec une tablette [190], qui diffère significativement d’une tâche d’exploration avec un casque.

Une spécificité de l’exploration d’objets avec une combinaison de vues mixtes et virtuelles dans MR est que changer le point de vue sur la vue mixte peut créer une incohérence visuelle, comme illustré à la Figure 5.5a.

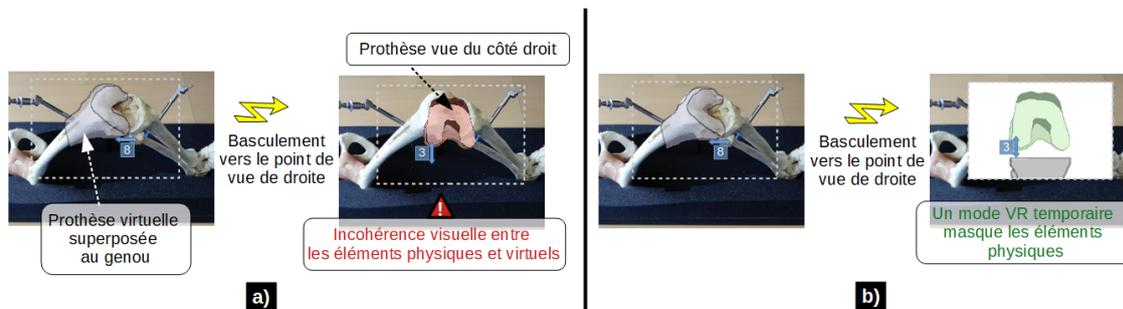


FIGURE 5.5 – Changer virtuellement de point de vue sur la vue mixte. a) L’objet virtuel ne correspond plus à son ancre physique, ce qui crée une incohérence visuelle. b) Ajout d’une transition vers un mode RV temporaire pour éviter toute incohérence visuelle.

Si l’objet physique et son augmentation virtuelle n’ont pas la même orientation parce que les points de vue sont différents, alors la métaphore de l’objet augmenté est remise en cause. Pour éviter toute incohérence visuelle entre la partie physique et la partie virtuelle de l’objet mixte, nous avons choisi d’utiliser un mode RV temporaire comme dans [15, 193]. Dès qu’une incohérence visuelle apparaît, nous masquons temporairement le monde réel à l’aide d’un arrière-plan opaque. Grâce à cette transition du mode RA au mode RV, seuls les objets virtuels restent visibles jusqu’à ce que les utilisateurs reviennent à un point de vue où l’objet mixte est cohérent. Ce principe est illustré à la Figure 5.5b. La principale différence avec les interfaces comportant des transition entre RA et RV (*transitional interfaces*) [30, 78] est que les utilisateurs ne sont pas totalement immergés dans un environnement virtuel. Dans notre cas, seul le champ de vision augmenté est affecté par le changement d’opacité d’arrière-plan du mode RV. L’environnement physique reste ainsi visible dans la vision périphérique.

Nous considérons deux types de techniques de navigation : la navigation physique et la navigation virtuelle. La navigation physique pour explorer un objet en RM peut être réalisée en utilisant deux approches : 1) se déplacer physiquement dans l’environnement et 2) déplacer l’objet physique lui-même pour le voir sous différentes perspectives. Les deux approches sont regroupées sous les termes *Action physique* dans notre espace de conception. Cependant, nous n’explorons pas ci-après la seconde approche qui correspond à la RA tangible [31]. Ce choix est motivé par notre domaine d’application principal, où le déplacement de l’objet augmenté (c’est-à-dire la jambe du patient) n’est pas possible. Ceci est également vrai dans d’autres domaines, en particulier l’industrie où il n’est pas possible de déplacer/faire tourner une machine de production.

Ainsi, nous considérons deux éléments de conception (couplage des vues et technique de navigation) qui définissent un espace de conception présenté à la Figure

5.6.

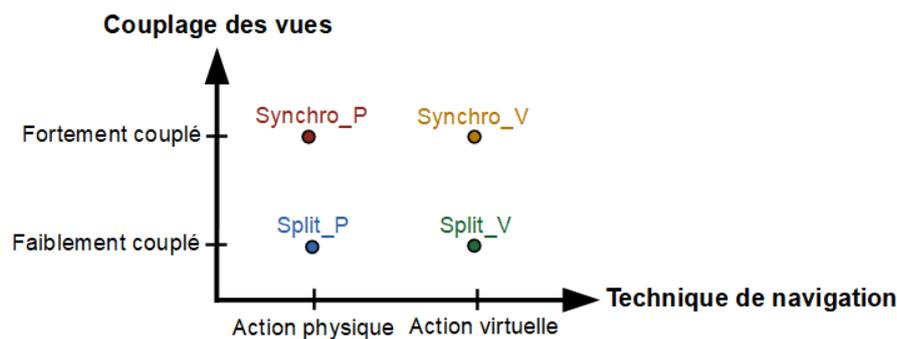


FIGURE 5.6 – Espace de conception considérant deux éléments de conception : le couplage des vues et la technique de navigation. 4 techniques sont ainsi conçues.

5.2 Conception de nouvelles techniques pour changer le point de vue

Nous avons défini 4 techniques d'interaction à partir des deux facteurs de conception étudiés : *Synchro_P*, *Synchro_V*, *Split_P* et *Split_V*. Les techniques *Synchro* sont basées sur des vues fortement couplées (ou coordonnées) tandis que les techniques *Split* utilisent des vues faiblement couplées. La dernière lettre du nom de la technique fait référence à la technique de navigation, soit *Physique* (P) soit *Virtuelle* (V). Ci-après, les techniques sont expliquées en considérant les formes abstraites de la Figure 5.7. La Boîte représente la vue mixte (cube physique augmenté) tandis que le Patron représente la vue virtuelle. La Figure 5.8 illustre le fonctionnement des 4 techniques en utilisant ces deux formes abstraites.

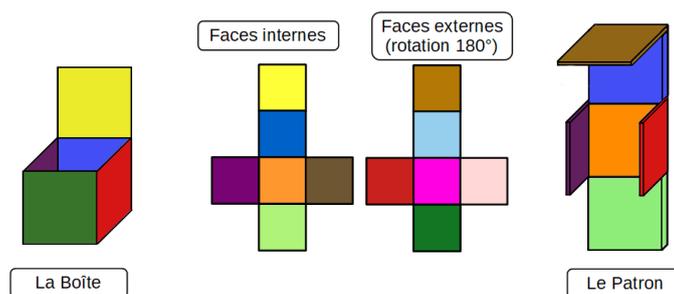


FIGURE 5.7 – Formes abstraites utilisées pour illustrer le fonctionnement des 4 techniques : une forme cubique (la Boîte) et son patron (nommé le Patron).

Couplage des vues

Avec des vues fortement couplées (techniques *Synchro*), les deux vues sont affectées par tout changement de point de vue. Avec des vues faiblement couplées (techniques *Split*), les utilisateurs modifient le point de vue dans une seule vue, laissant l'autre vue inchangée. Pour ce faire, nous ajoutons un mécanisme de sélection de vue. Un retour visuel en haut de l'écran du casque indique quelle vue est

actuellement active, comme schématisé à la Figure 5.8. Lors d'un changement de point de vue, soit à cause d'un mouvement physique (*Split_P*) soit d'une action de navigation virtuelle (*Split_V*), seule la vue active est affectée par le changement. L'autre vue reste la même, ce qui entraîne des points de vue distincts sur l'objet. Avec *Split_P*, la vue sélectionnée de l'objet "suit" les mouvements physiques, ce qui signifie qu'elle se tourne toujours pour faire face à l'utilisateur.

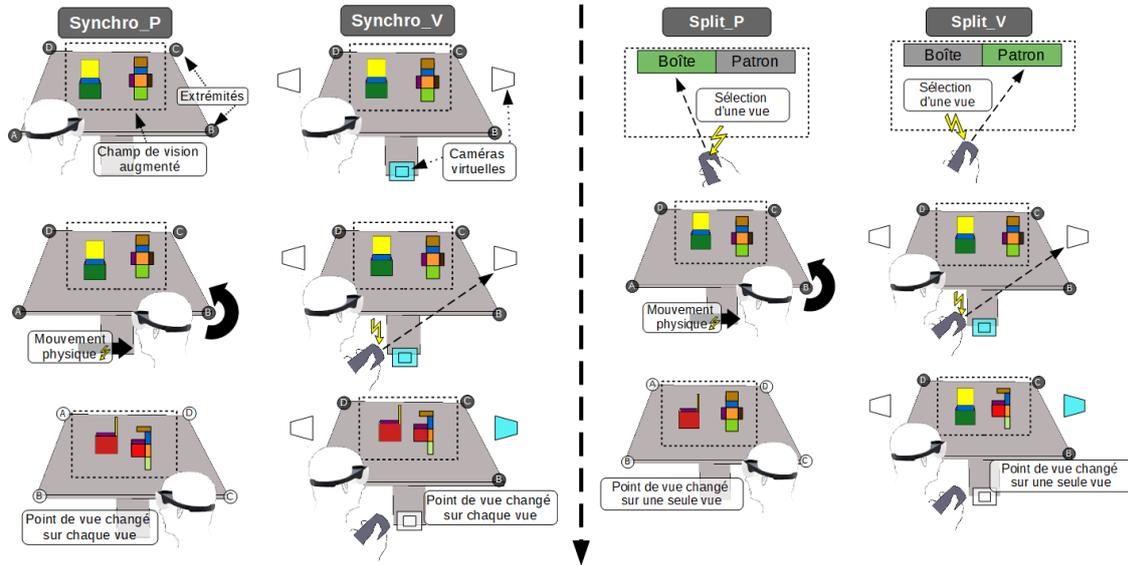


FIGURE 5.8 – Fonctionnement des 4 techniques de changement de point de vue illustré avec la Boîte et le Patron (voir Figure 5.7).

Technique de navigation

La technique de navigation physique (*Synchro_P* et *Split_P*) est l'approche la plus simple : les utilisateurs doivent se déplacer physiquement autour de l'objet pour le voir sous différents angles. Inspiré des travaux de Sukan et al. [190], la technique de navigation virtuelle (*Synchro_V* et *Split_V*) se base sur un mécanisme de raycasting qui permet aux utilisateurs de sélectionner un point de vue discret distant. Le raycasting est couramment utilisé pour sélectionner des objets distants en RM [7, 66, 155]. Avec *Synchro_V* et *Split_V*, les utilisateurs peuvent visualiser l'objet cible d'un point de vue différent sans se déplacer physiquement.

Au final, se déplacer physiquement autour de vues fortement couplées (*Synchro_P*) est la seule technique où le mode RV temporaire n'est pas nécessaire car il ne peut y avoir aucune incohérence visuelle. Toutes les techniques ont été implémentées avec le casque Epson Moverio BT-300 et le logiciel *TrackingTools*, comme pour les expérimentations des Chapitres 3 et 4.

5.3 Études expérimentales

Nous avons mené deux études expérimentales pour évaluer les quatre techniques d'interaction conçues à partir de notre espace de conception (Figure 5.6). Les deux études reposent sur la même installation expérimentale.

5.3.1 Installation expérimentale

La configuration expérimentale (Figure 5.9) est basée sur une tâche d'exploration d'un cube inspirée par l'enseignement de la géométrie 3D aux enfants [202]. Néanmoins, il présentait également quelques similitudes avec la planification chirurgicale [3]. Une table circulaire (35 cm de diamètre, 120 cm de hauteur) constitue l'espace de travail de l'expérience. Les participants se tenaient derrière cette table et se concentraient sur un objet abstrait. La vue mixte de l'objet était représentée par la Boîte et la vue virtuelle de l'objet par un Patron virtuel. La Boîte et le Patron étaient tous deux à la même distance initiale de l'utilisateur et proches l'un de l'autre. Notre objectif est de limiter les problèmes de perception dus au casque, y compris les conflits de vergence-accommodation [152] détaillés à la Figure 2.11.

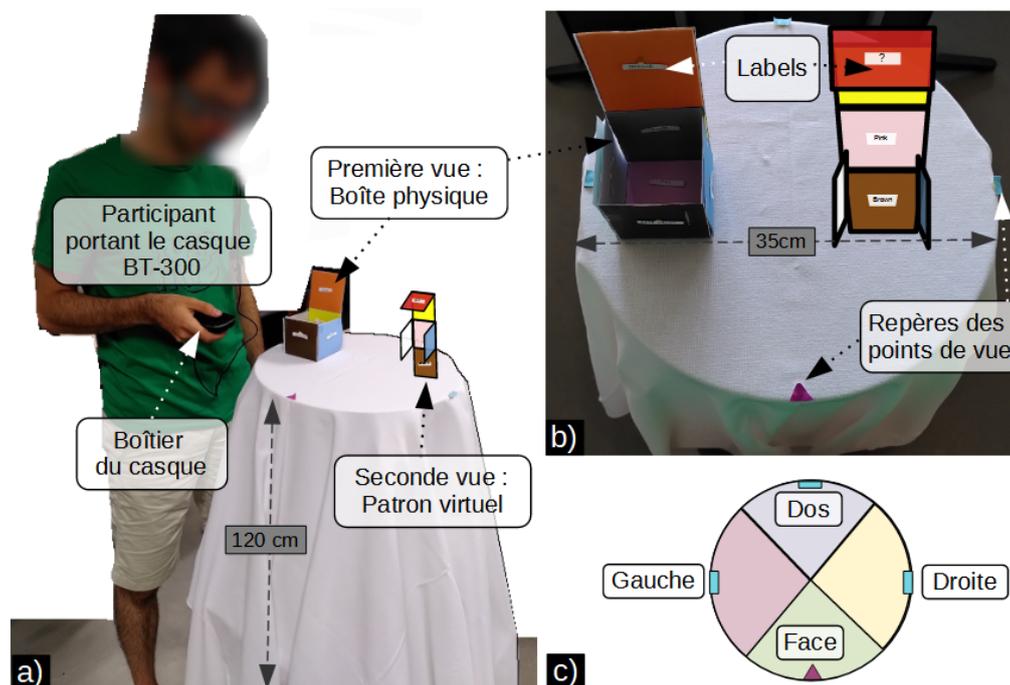


FIGURE 5.9 – Pilote étude 5 : installation expérimentale. a) Participant debout devant la table représentant l'espace de travail. b) Aperçu des deux vues de l'objet : la Boîte physique (gauche) et le Patron virtuel (droite). c) Les 4 secteurs de l'espace de travail représentant les zones de sélection de chaque point de vue existant.

La Boîte était un objet physique en forme de boîte ouverte (avec des faces carrées de 7 cm de long) située sur le côté gauche de la table. Chacune de ses faces avait une couleur donnée (toutes différentes, y compris pour les faces internes et externes). Notre plan initial était d'augmenter chaque face de la Boîte avec une étiquette textuelle virtuelle. Cependant, pour éviter tout biais technologique dû au calibrage, la Boîte était un objet entièrement physique. Les étiquettes des faces n'étaient donc pas virtuelles mais physiques : des étiquettes en papier étaient fixées sur chaque face. Néanmoins, nous avons traité la Boîte comme un objet mixte. Lorsque le point de vue sélectionné par l'utilisateur ne correspondait plus au cube physique, le système déclenchait la transition vers le mode RV pour éviter toute incohérence visuelle, comme illustré à la Figure 5.5. Une version entièrement virtuelle de la Boîte (même position, couleurs et étiquettes) a été utilisée en mode RV au lieu de la version

physique.

Le Patron virtuel correspondant à la vue virtuelle a été placé à droite de la table. Comme pour la Boîte, chaque face avait une couleur spécifique et une étiquette écrite dessus, mais cette copie était entièrement virtuelle. Contrairement à un patron 2D classique d'un cube, nous avons fait pivoter certaines des faces du patron pour obtenir un objet 3D avec des éléments faisant face aux 4 directions possibles (avant, droite, arrière et gauche).

Quatre points de vue discrets étaient présents dans la configuration, un par direction cardinale. Ces points de vue ont été matérialisés sur la table circulaire par des repères (un triangle violet pour le point de vue initial, des rectangles bleus pour les autres) comme montré à la Figure 5.9b et c. Avec les techniques *Synchro_P* et *Split_P*, les participants se déplaçaient physiquement uniquement entre ces positions. Les vues se mettaient à jour uniquement lorsque les participants atteignaient un point de vue existant. Lors de l'utilisation de *Synchro_V* et *Split_V*, les points de vue étaient également matérialisés par des triangles virtuels. Pour faciliter la sélection de ces points de vue virtuels, l'espace de travail circulaire a été divisé en 4 secteurs de 90 ° chacun (Figure 5.9c). Les participants devaient orienter le boîtier du casque BT-300 dans la direction d'un secteur. Le bouton central du boîtier était utilisé pour confirmer la sélection et déclencher la mise à jour du point de vue correspondant. De cette façon, les participants pouvaient virtuellement changer le point de vue tout en se concentrant visuellement sur l'objet.

5.3.2 Étude pilote : exploration informelle de l'espace de conception

Nous avons d'abord mené une étude pilote exploratoire pour évaluer les quatre techniques d'interaction que nous avons conçues pour un système de RM à vues multiples. Notre objectif n'était pas seulement d'étudier chaque technique indépendamment, mais aussi d'explorer la combinaison de ces techniques. Une seconde étude menée dans des conditions similaires est décrite à la Section 5.3.3.

Tâche

Les participants devaient réaliser une tâche d'exploration : compter le nombre de faces invalides sur l'objet. Une face était considérée invalide lorsque sa couleur ne correspondait pas à son étiquette (ainsi, une face verte avec une étiquette "Violet" n'était pas valide). Les participants exploraient la scène à travers au moins 3 points de vue différents pour déterminer le nombre de faces invalides. Une fois sûrs de leur réponse, les participants devaient appuyer sur le bouton de validation pour arrêter l'essai. Ils étaient ensuite invités à saisir leur réponse à l'aide des flèches du boîtier du casque.

Quand une face était étiquetée avec un symbole "?" au lieu d'un nom de couleur (ci-après appelée *face spéciale*), les participants devaient rechercher la face correspondante sur l'autre vue pour déterminer la validité de cette face courante. Par exemple, un "?" sur une face rouge de la Boîte obligeait les participants à vérifier si la face correspondante sur le patron était valide. Les faces spéciales étaient équitablement présentes sur la Boîte et le Patron. Le but de cette configuration expérimentale était de s'assurer que les participants utilisent les deux vues pour accomplir la tâche.

Conception de l'étude

4 configurations différentes d'objets ont été conçues pour cette étude pilote. Chaque configuration avait une disposition spécifique d'étiquettes pour créer des essais avec des difficultés différentes. La configuration la plus simple n'avait qu'une seule face invalide immédiatement visible et seulement 2 étiquettes "?". La configuration la plus difficile avait 4 étiquettes "?" et 3 faces non valides nécessitant des points de vue différents (dont 1 face non valide sur un "?"). L'ordre des configurations était aléatoire. Deux configurations supplémentaires ont été utilisées exclusivement pour les phases d'entraînement.

L'expérience était divisée en deux phases : la phase imposée et la phase libre. Au cours de la phase imposée, les participants devaient effectuer 4 blocs d'essais, un par technique d'interaction. L'ordre était imposé et contrebalancé entre les participants en utilisant un carré latin. Chaque bloc d'essai était constitué de 2 étapes d'entraînement suivies de 4 étapes enregistrées. Les participants pouvaient faire une courte pause entre chaque bloc s'ils le voulaient. Lors de l'utilisation des techniques *Split_P* et *Split_V*, les participants pouvaient basculer entre les modes "Boîte active" et "Patron actif" en utilisant les flèches droite et gauche du pavé du boîtier du casque. Un retour visuel en haut de l'écran du casque indiquait quel mode était actuellement sélectionné.

Au cours de la phase libre, les participants étaient informés qu'ils pouvaient utiliser n'importe quelle technique pour accomplir la tâche. Ils pouvaient se déplacer physiquement autour de la table, changer virtuellement de point de vue et basculer entre des vues fortement et faiblement couplées. Pour ce faire, les participants devaient simplement sélectionner le mode correspondant (Boîte active, Patron actif ou Synchro) en utilisant le boîtier du casque comme auparavant. Le but de cette phase était de permettre aux participants de choisir librement la combinaison de techniques qui convenait le mieux à leur stratégie pour accomplir la tâche d'exploration. Les participants ont effectué 2 essais d'entraînement au début, puis 4 essais enregistrés. L'expérience incluant la phase imposée et la phase libre a duré 90 minutes en moyenne.

Leçons apprises

Quatre participants (3 hommes, 1 femme) âgés de 23 à 30 ans ($sd = 4,6$) ont été recrutés pour cette étude. Ils venaient tous de notre laboratoire et n'étaient pas habitués à travailler en RM. Les participants pouvaient porter des lentilles de contact ou des lunettes avec le casque s'ils le souhaitaient. Bien qu'il s'agisse d'une étude pilote, cette première exploration de nos 4 techniques d'interaction a suggéré quelques tendances. Dans l'ensemble, 3 participants sur 4 ont préféré utiliser des vues fortement couplées pour accomplir la tâche (techniques *Synchro_V* et *Synchro_P*), mais seulement 1 participant sur 4 (P2) n'a jamais utilisé la technique *Split_V* pendant la section libre. P4 a mentionné que la navigation virtuelle était "*utile mais exigeait plus de réflexion, comme une tâche supplémentaire*". En particulier, 2 participants semblaient avoir une représentation mentale différente du changement de point de vue. Ces participants s'attendaient à faire pivoter l'objet lors de la sélection d'un autre point de vue au lieu de le voir à partir d'une autre position virtuelle. Cette attente initiale a surtout affecté les points de vue latéraux. Par

exemple, sélectionner le point de vue gauche impliquait de se déplacer virtuellement vers le côté gauche de la table (*i.e.* pour voir le côté *droit* de l'objet), sans voir le côté *gauche* de l'objet.

Changer physiquement le point de vue a également été jugé moins exigeant mentalement. En particulier, P3 a expliqué que se déplacer autour de la table était plus facile pour s'assurer qu'il vérifiait chaque face sur les deux vues et qu'il basait sa stratégie sur la visite de chaque point de vue dans le sens anti-horaire. Cependant, cette technique de navigation physique a également été jugée plus fatigante physiquement, ce qui confirme les conclusions de Sukan et al. [190]. P4 a expliqué que "*se déplacer autour de la table n'est pas vraiment nécessaire car il y a des points de vue [virtuels]*".

Dans un questionnaire final, les 4 participants ont classé la technique *Split_P* comme leur moins préférée. Cette approche était perçue comme déroutante, plus complexe et plus fatigante. Un tel résultat suggère que lorsque les participants se déplacent autour de la table, ils s'attendent à ce que les deux formes se comportent comme des objets physiques. Ce n'était pas le cas avec cette technique car la forme sélectionnée se retournait pour toujours faire face à l'utilisateur. L'utilisation de vues faiblement couplées peut être plus appropriée avec une technique de navigation virtuelle (*Split_V*).

La fatigue supplémentaire des techniques de navigation physique et les mauvais résultats concernant *Split_P* nous ont encouragés à nous concentrer sur les techniques de navigation virtuelle. Un tel choix est également cohérent avec certains des contextes d'application tels que la chirurgie augmentée et l'assemblage de machines en MR, où il est souvent impossible de se déplacer physiquement dans l'espace de travail. Par conséquent, nous avons mené une deuxième étude formelle des deux techniques *Syn_V* et *Split_V*.

5.3.3 Deuxième étude : approfondissement de la navigation virtuelle

Le but de cette seconde étude était d'explorer la complémentarité de nos techniques d'interaction lors de la navigation virtuelle. De plus, nous voulions également étudier si l'une des deux vues était utilisée comme référence par les participants. Nous avons émis l'hypothèse que, puisque la Boîte est physique, les participants seraient plus enclins à la percevoir comme une vue principale. 12 nouveaux participants (10 hommes, 2 femmes) âgés de 23 à 41 ans ($sd = 6,4$) ont été recrutés pour cette étude. La configuration expérimentale et les mesures étaient similaires à celles de l'étude pilote, mais nous avons légèrement adapté la tâche d'exploration.

Comme dans l'étude pilote, le but de cette tâche d'exploration était de compter le nombre de faces invalides dans les deux vues. Le processus de saisie des réponses était le même qu'auparavant. Cependant, cette fois, les participants devaient compter le nombre de faces spéciales (marquées d'un caractère "?") qui n'étaient pas valides. Les faces normales devaient être ignorées. Pour augmenter la complémentarité des deux vues, nous avons adapté la conception des faces spéciales. Pour chaque face spéciale marquée d'un "?", la couleur de la face correspondante sur l'autre vue était changée en gris. De cette façon, les participants devaient utiliser les informations des deux vues (la couleur d'une vue et le nom de la couleur de l'autre vue). Ceci

est différent de l'étude précédente où toutes les informations (couleur de la face + étiquette) étaient présentes dans la deuxième vue, comme illustré à la Figure 5.10. Ce changement reflète également le scénario du planning chirurgical de la Figure 5.3 où les deux états de la jambe sont complémentaires et également importants.

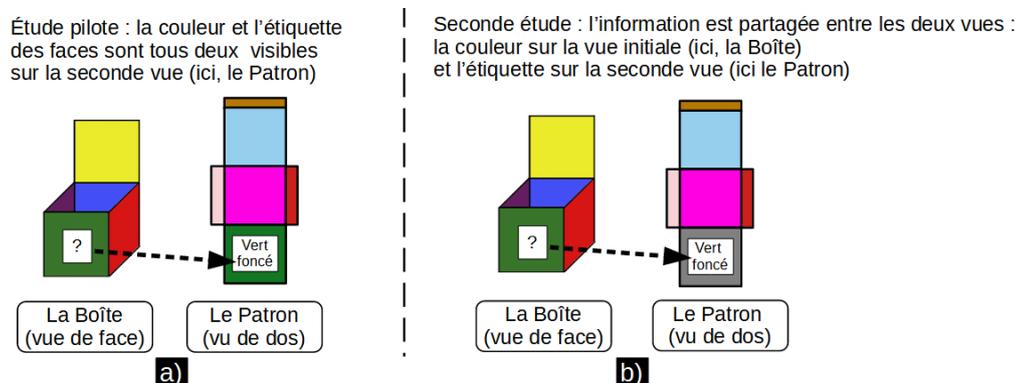


FIGURE 5.10 – a) Étude pilote : la couleur de la face et l'étiquette sont tous deux visibles sur la même forme. b) Seconde étude : l'information est partagée entre les deux vues. La couleur est visible sur la vue initiale (ici, la Boîte) et l'étiquette sur la seconde vue (ici, le Patron).

Conception de l'étude

4 nouvelles configurations (+2 pour l'entraînement uniquement) de difficulté croissante ont été définies sur la base des considérations de conception suivantes. La Boîte et le Patron avaient chacun le même nombre de faces spéciales dans une configuration donnée. De plus, nous avons également considéré les positions des faces spéciales par rapport aux techniques d'interaction. D'une part, certaines faces spéciales pouvaient être vérifiées en utilisant le même point de vue pour la Boîte et le Patron. Dans ce cas, la tâche pouvait être effectuée en utilisant *Synchro_V*. D'un autre côté, certaines faces spéciales obligeaient les participants à avoir des points de vue distincts pour la Boîte et le Patron, encourageant l'utilisation de *Split_V*.

Les participants ont commencé par une phase d'entraînement pour se familiariser avec les deux techniques d'interaction. Ils ont dû faire 2 configurations d'entraînement avec *Synchro_V* et 3 avec *Split_V* qui selon l'étude pilote était un peu plus difficile à prendre en main. Après avoir recueilli des retours qualitatifs, nous avons demandé aux participants de réaliser une phase libre avec 2 configurations d'entraînement suivies de 2 blocs de 4 configurations enregistrées chacun. Les participants ont donc effectué chaque configuration deux fois. Chaque configuration était identifiée avec 2 numéros spécifiques (un numéro par bloc) pour empêcher les participants de comprendre qu'ils ont déjà vu une configuration donnée. Un questionnaire final était remis pour recueillir les retours des participants. L'expérience a duré en moyenne 50 minutes.

Mesures

Les temps de complétion des tâches et les réponses erronées ont été enregistrés. Par ailleurs, nous voulions également garder une trace des stratégies d'exploration

utilisées par les participants. Le nombre de fois où les participants ont changé de point de vue est un indicateur révélateur de ces stratégies.

Nous avons enregistré chacun de ces changements de points de vue avec le mode faiblement couplé (changement sur la Boîte ou sur le Patron) ainsi qu'avec le mode mode fortement couplé. Un changement de point de vue en mode Synchro (fortement couplé), qui implique un changement de point de vue sur les deux vues, a été enregistré comme un seul changement de point de vue (*i.e.* nous avons considéré le nombre d'actions de changement de point de vue effectuées par les participants).

Après la phase d'entraînement, un questionnaire Raw-TLX était remis aux participants (Annexe A). Notre objectif était de recueillir des informations sur la technique et la stratégie adoptées pour mener à bien la tâche. À la fin de la phase libre, un deuxième questionnaire Raw-TLX suivi d'un entretien semi-structuré (Annexe E) était réalisé pour collecter des données qualitatives sur les stratégies utilisées par les participants et leurs préférences quant à la combinaison de techniques.

Résultats

Pour effectuer une analyse statistique des mesures de temps, nous avons utilisé une ANOVA à mesures répétées (avec $\alpha = 0,05$). Des tests post-hoc ont été réalisés en utilisant des t-tests pairés et des corrections de Bonferroni. Lorsque l'ANOVA n'était pas appropriée, nous avons utilisé un test de Friedman à la place, suivi d'un test post-hoc de Wilcoxon avec des corrections de Bonferroni. Les résultats de ces tests statistiques sont rapportés ci-dessous. Cependant, comme recommandé par Dragicevic [61], nous mettons davantage l'accent sur les intervalles de confiance et la taille des effets. Dans toutes les figures, les barres d'erreur sont des intervalles de confiance (IC) à 95%.

Temps de complétion des tâches

Nous avons d'abord étudié les temps de complétion des tâches pour chaque configuration possible. Comme l'illustre la figure 5.11a, les participants ont mis en moyenne plus de temps pour réaliser la configuration C4 la plus difficile. Cette tendance est corroborée par l'analyse statistique : nous avons observé de légères preuves de différence entre C4 et la configuration la plus simple C1 ainsi qu'entre C4 et la difficulté moyenne C2 ($p = 0,079$ et $p = 0,061$ respectivement). Nous avons observé de bonnes preuves de différence entre C4 et la difficulté moyenne C3 ($p = 0,035$). Cependant, nous n'avons pas trouvé de preuve de différence entre la configuration C1 la plus simple et les deux configurations de difficulté moyenne C2 et C3. Nous avons également vérifié le nombre de changements de point de vue effectués lors des essais pour chaque configuration, mais aucune preuve notable de différence n'a été observée ($p = 0,2$).

Les deux configurations de difficulté moyenne C2 et C3 avaient 2 faces spéciales. La première face spéciale était valide mais demandait des orientations différentes de la Boîte et du patron pour voir à la fois la couleur de la face et son étiquette. La deuxième face spéciale était invalide sur le Patron (C2) ou sur la Boîte (C3), mais ne nécessitait pas des orientations différentes. Le fait que nous n'ayons observé aucune preuve de différence entre C2 et C3 montre que la conception de la Boîte et du Patron est équilibrée et comparable, bien qu'un objet soit mixte (Boîte) et l'autre (Patron)

soit virtuel. La configuration la plus simple C1 avait deux faces spéciales valides, ce qui suggère qu'avoir une face invalide comme dans C2 et C3 n'a pas d'impact sur les performances. Cependant, la preuve de la différence avec la configuration la plus difficile C4 est bien plus claire. Or, nous n'avons observé aucune preuve de différence pour le nombre de changements de point de vue (environ 10 en moyenne, $\chi^2 = 4,6$, $p = 0,2$). Le nombre de changements de point de vue (*i.e.* la navigation) était en moyenne le même pour chaque configuration. Par conséquent, nous expliquons ces moins bonnes performances par la charge cognitive des utilisateurs liée au nombre plus élevé de faces spéciales.

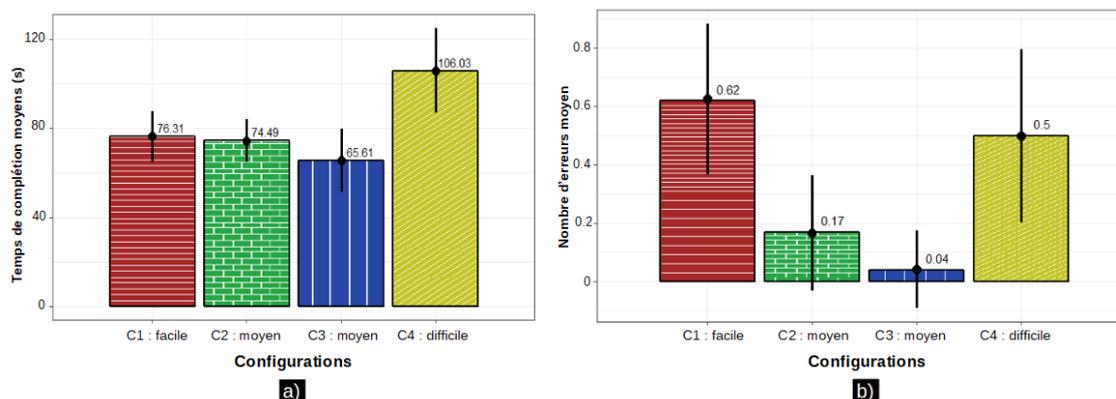


FIGURE 5.11 – Étude 5 : a) Temps de complétion moyens pour chaque configuration. b) Nombre moyen d'erreurs (réponses incorrectes) pour chaque configuration. Les barres d'erreur sont les intervalles de confiance à 95%.

Erreurs

Les réponses non valides des participants ont été enregistrées comme des erreurs. Étonnamment, nous avons trouvé de légères preuves de différence entre la configuration facile et les deux configurations de difficulté moyenne. C1 semble conduire à un peu plus d'erreurs que C2 ($p = 0,088$) et C3 ($p = 0,062$), comme le montre la figure 5.11b. De même, nous avons observé une tendance plus nuancée avec la configuration la plus difficile C4 qu'avec les configurations de difficulté moyenne ($p = 0,49$ pour C2 et $p = 0,16$ pour C3).

C1 n'avait pas de face spéciale invalide, mais les participants étaient informés au début de la phase d'entraînement qu'une configuration pouvait avoir zéro ou plusieurs faces invalides. Une explication possible est que certains participants ont été déconcertés par la couleur blanche d'une face spéciale du Patron dans la configuration C1. Pour rendre cette face visible malgré le fond blanc opaque du mode RV, sa couleur blanche a été changée en un gris très clair. 9 participants ont fait au moins une erreur sur cette configuration, probablement parce qu'ils ont considéré cette face spéciale comme invalide (couleur gris clair sur le Patron, étiquette "Blanc" sur la Boîte). Heureusement, C1 était la seule configuration impliquant la couleur blanche dans une face spéciale.

Commentaires qualitatifs

Les commentaires qualitatifs de la phase imposée suggèrent que la tâche n'était pas physiquement fatigante ni frustrante. Seuls 3 participants sur 12 ont estimé que la tâche était mentalement exigeante ou très exigeante. La majorité des participants (8/12) a préféré utiliser Split, même si P5 a souligné que cette technique semblait plus complexe au départ. Lorsqu'il leur été demandé d'imaginer leur stratégie s'ils pouvaient librement passer de Split à Synchro, 4 participants ont expliqué qu'ils utiliseraient Synchro pour avoir un aperçu des deux vues, puis Split pour traiter les faces spéciales. 3 participants ont déclaré qu'ils utiliseraient essentiellement Split et les 5 participants restants étaient indécis.

Les commentaires qualitatifs du questionnaire Raw-TLX suggèrent que la phase libre n'était pas physiquement fatigante ni frustrante. 7 participants sur 12 ont estimé qu'ils avaient assez bien réussi la tâche d'exploration. Cependant, les retours étaient plus mitigés sur la demande mentale et l'effort requis pour atteindre un bon niveau de performance. Un aperçu de ces résultats est disponible à la Figure 5.12.

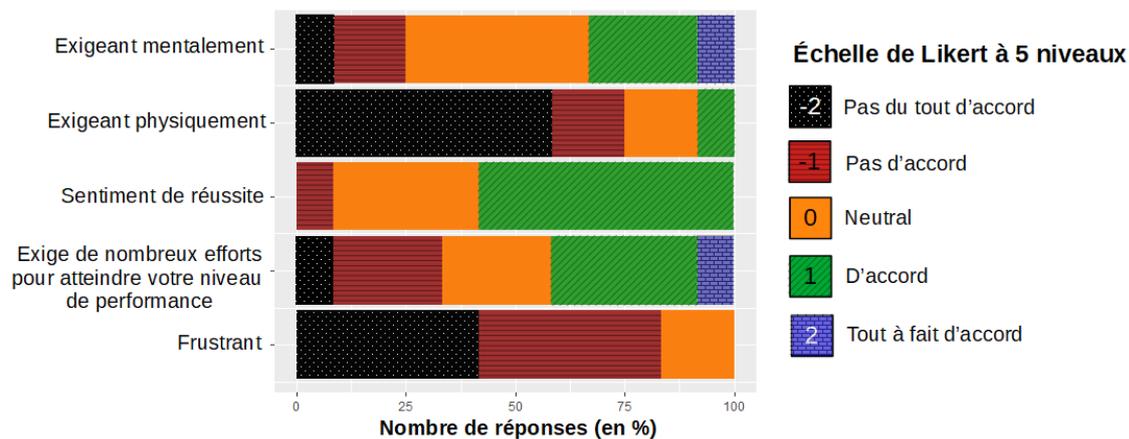


FIGURE 5.12 – Étude 5 : retours qualitatifs liés au questionnaire Raw-TLX à la fin de la phase libre.

Stratégies : techniques Synchro et Split

Nous avons observé que les participants utilisaient différentes stratégies pour accomplir la tâche. Nous avons identifié 6 stratégies regroupées en 3 catégories comme illustré à la Figure 5.13 : 2 stratégies basées principalement sur Split, 2 stratégies principalement basées sur Synchro et 2 stratégies hybrides combinant de façon égale les modes Split et Synchro. Plutôt que de déterminer quelle stratégie est la meilleure pour inspecter un objet 3D dans un système de RM à vues multiples (en plus de ne pas avoir assez de participants par stratégie pour conclure sur les performances), nous nous concentrons sur la compréhension des stratégies observées. Pour ce faire, nous utilisons 1) les retours qualitatifs des questionnaires et des entretiens Raw-TLX et 2) des données quantitatives enregistrées lors des changements de point de vue.

6 participants sur 12 ont expliqué qu'ils utilisaient une combinaison des techniques *Synchro_V* et *Split_V* (principalement les stratégies S3 et S4, plus S5 de

manière plus nuancée). Dans l'ensemble, 5 participants ont déclaré avoir principalement utilisé *Split_V* (S1 et S2) et 4 participants principalement *Synchro_V* (S5 et S6). Nous avons vérifié que les stratégies déclarées par les participants (retour qualitatif des entretiens) étaient cohérentes avec les données quantitatives enregistrées sur l'utilisation des modes Synchro et Split sur chaque vue. Fait intéressant, nous avons observé que le mode Synchro était proportionnellement plus utilisé. Les participants qui ont principalement utilisé Synchro (S5 et S6) ont à peine utilisé le mode Split (environ 1,5 changement de point de vue en moyenne). Cependant, l'inverse n'est pas vrai : les 5 participants qui ont le plus utilisé Split (S1 et S2) ont utilisé le mode Synchro pour effectuer entre 3 et 4 changements de point de vue en moyenne.

Stratégie	Description	Participants	
S1	Split sur une vue, puis sur l'autre vue	2, 9, 10, 11	
S2	Comme S1, mais commence toujours par vérifier la Boîte physique en premier	1	
S3	Synchro pour avoir un aperçu Split pour traiter toutes les faces spéciales (indépendamment de l'orientation requise pour la Boîte et le Patron)	5, 8	
S4	Comme S3. Après avoir traité une face spéciale, Split pour avoir la même orientation Boîte & Patron	4	
S5	Synchro pour explorer et traiter les faces spéciales faciles en premier (même orientation pour la Boîte et le Patron). Ensuite, Split lorsque des orientations différentes sont nécessaires.	3, 6, 12	
S6	Synchro uniquement, navigation sens anti-horaire	7	

FIGURE 5.13 – Résumé des stratégies adoptées par les participants pour mener à bien la tâche d'exploration.

6 participants sur 12 ont expliqué que le mode Synchro était surtout utile pour explorer les vues globalement et pour trouver rapidement les faces spéciales. P7, qui a exclusivement utilisé ce mode, a déclaré que "*c'était plus naturel, plus proche de ce qui peut être fait physiquement*". P6 a en outre expliqué qu'il était possible de mémoriser la couleur de la face spéciale avant de vérifier l'étiquette sur l'autre vue et a donc considéré que le mode Synchro était suffisant pour accomplir la tâche. P9 n'a pas beaucoup utilisé le mode Synchro mais a mentionné que c'était un bon moyen de réinitialiser les deux vues (*i.e.* pour revenir au contexte initial comme au début de l'essai).

Cependant, les 6 participants restants n'ont pas considéré que le mode Synchro était vraiment utile. P2 a expliqué que l'utilisation de Synchro "*pouvait casser le contexte que j'ai créé*" tandis que P1 a déclaré qu'il avait besoin de "*garder le '?' visible à l'écran*", ce qui n'était possible qu'avec le mode Split. Les participants qui ont principalement utilisé le mode Split ont souligné cette possibilité de maintenir le contexte visuel de la vue où ils avaient trouvé la face spéciale. P10 ajoute que "*cela donne une plus grande liberté de mouvement et que cela donne plus l'impression de prendre un objet et de le tourner au lieu de se déplacer dans les deux*".

Le nombre limité de participants par type de stratégie nous empêche de tirer une conclusion ferme sur la performance de chaque stratégie, comme le montrent les

très grands intervalles de confiance de la figure 5.14. Néanmoins, nous avons noté une première tendance dans notre étude. Les participants qui utilisaient principalement le mode Synchro semblaient terminer la tâche plus rapidement et avec moins d'actions de changement de point de vue.

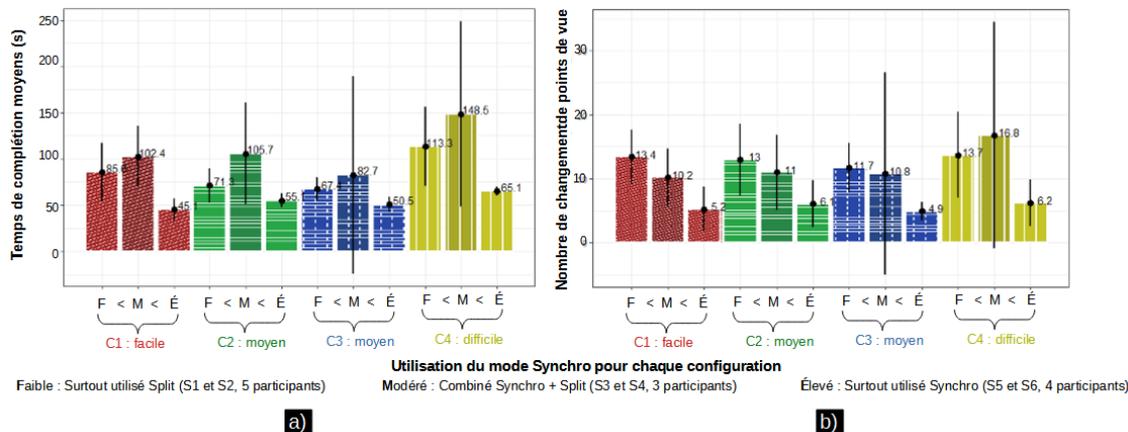


FIGURE 5.14 – Étude 5 : Performance des participants en fonction du type de stratégie et de la configuration. a) Temps de complétion moyens. b) Nombre moyen de changements de point de vue. Les barres d'erreurs représentent les intervalles de confiance à 95%.

5.4 Leçons apprises et discussion

À partir des résultats quantitatifs et qualitatifs que nous avons recueillis pour l'exploration d'un cube et les couleurs de ses faces, nous proposons un ensemble de lignes directrices pour les concepteurs.

5.4.1 Stratégies

Les résultats expérimentaux suggèrent que des techniques fortement couplées et faiblement couplées sont utilisées à des fins différentes. Le couplage fort des vues est plus adapté pour explorer rapidement les deux vues et trouver des informations spécifiques (dans notre cas, des faces spéciales). En plus des commentaires qualitatifs, cela se reflète également dans le nombre de changements de point de vue pendant l'exploration. Nous rappelons que nous considérons un changement de point de vue en mode Synchro, donc un changement de point de vue sur les deux vues, comme un changement de point de vue (*i.e.* nous considérons le nombre d'actions de changement de point de vue effectuées par les participants). Dans ce cas, moins d'actions de changement de point de vue sont requises de la part des participants pour voir toutes les faces sur les deux vues.

Néanmoins, lorsque des points de vue distincts sont requis sur les différentes vues, des vues fortement couplées impliquent de perdre le contexte visuel de la vue initiale. Il est toujours possible de mémoriser les informations avant de se concentrer sur l'autre vue. Cependant, cela ne convient pas nécessairement dans des contextes d'application où les utilisateurs ont déjà une charge cognitive importante telle que

la chirurgie augmentée. Dans notre cas, seul P6 a explicitement mentionné cette approche. Ce participant a surtout utilisé la stratégie S5 (principalement Synchron, Split quand 2 points de vue différents sont nécessaires) lors de l'étude, mais a commencé à mémoriser la couleur de la face lors des derniers essais. Le mode Split n'était donc plus nécessaire dans ce cas. Les 11 autres participants (y compris P7 qui n'utilisait que Synchron) ont exécuté la tâche en comparant visuellement les faces spéciales. Nous avons observé que les participants utilisaient principalement des vues faiblement couplées pour comparer des formes avec des points de vue distincts. La technique *Split_V* a permis aux participants de traiter des faces spéciales difficiles sans avoir à mémoriser des informations supplémentaires ou à changer de point de vue sur la vue initiale par la suite. Ces résultats confortent la règle de complémentarité décrite par Baldonado et al. [18]. Les auteurs mentionnent que la comparaison visuelle est plus facile à réaliser que la comparaison de mémoire. Cela met en évidence les avantages de l'utilisation de plusieurs vues faiblement couplées par rapport aux vues fortement couplées.

D'une part, avoir des vues faiblement couplées peut être initialement plus complexe car les utilisateurs doivent avoir une bonne représentation mentale des deux formes. Les commentaires qualitatifs des participants de l'étude pilote et de la deuxième étude appuient cette observation. D'autre part, les vues faiblement couplées permettent de faire des comparaisons directes visuellement au lieu de mémoriser une couleur pour faire la comparaison. Le compromis peut donc dépendre de la complexité de l'objet 3D et de la nature de la tâche (*i.e.* le nombre de comparaisons entre vues). Dans cette deuxième étude, la plupart des participants ont utilisé une combinaison des deux types de couplage de vues. Un tel résultat met en évidence la complémentarité des deux approches et le potentiel des stratégies hybrides. Nous encourageons donc les concepteurs à considérer à la fois les vues fortement et faiblement couplées afin de permettre aux utilisateurs d'adopter la stratégie qui leur convient le mieux (Recommandation R1, voir Tableau 5.1).

5.4.2 Vue de référence

Nous avons émis l'hypothèse que la vue mixte (la Boîte) serait considérée comme la vue principale car elle a des liens plus forts avec le monde réel et que la vue virtuelle (le Patron) comme une copie virtuelle de l'objet avec une géométrie différente. Cependant, même dans notre cas où la Boîte était purement physique, les participants ne la percevaient pas comme une vue de référence. Nous n'avons observé aucune prédominance de la vue mixte sur la vue virtuelle. 6 sur 12 participants se sont concentrés sur la vue où ils ont trouvé la première face spéciale. Selon P7, les deux points de vue étaient équivalents. P12 a expliqué qu'il préférerait déplacer le Patron virtuel parce que la Boîte était physique. Cependant, P1 a toujours commencé par inspecter la boîte physique.

Dans l'ensemble, les participants étaient donc principalement influencés par la tâche lors du choix de leur vue de référence. D'autres participants ont simplement été influencés par les différences dans les formes 3D, comme P9 expliquant qu'il était plus familier avec la forme cubique de la Boîte. Un tel résultat suggère que la vue mixte et la vue virtuelle peuvent être considérées de même importance si elles sont toutes deux complémentaires pour accomplir la tâche (Recommandation R2, Table 5.1).

5.4.3 Transitions entre RA et RV

Les transitions entre les modes RA et RV peuvent avoir impacté cette absence de vue de référence puisque la Boîte a été remplacée par une copie virtuelle dès l'apparition d'une incohérence virtuelle. Dans l'ensemble, les participants ont estimé que les transitions étaient utiles et n'en ont pas été dérangées, ce qui confirme les résultats des travaux antérieurs [15]. Néanmoins, le passage du mode RV au mode RA a été jugé moins fluide que le passage du mode RA au mode RV par 3 participants. Nous avons également noté que pendant la session de formation, certains participants étaient désorientés en quittant le mode RV. Ils ont déclaré que la Boîte virtuelle avait disparu et n'étaient pas sûrs de comprendre si c'était voulu ou non. Le superviseur a dû expliquer à nouveau qu'il pouvait utiliser la Boîte physique pour accomplir la tâche. L'asymétrie des transitions entre les modes RA et RV mérite d'être explorée plus avant dans des travaux futurs. Le fait que le passage de la RV à la RA semble plus difficile que celui de la RA à la RV est à considérer par les concepteurs travaillant sur des interfaces de transition [30, 78] en RM.

Les transitions automatiques entre les modes RA et RV constituent une approche prometteuse pour explorer la vue physique ou mixte d'un objet en RM. Il permet aux utilisateurs d'inspecter cette vue de n'importe quel point de vue sans être dérangés par des incohérences visuelles. De plus, la taille limitée du champ de vision du casque n'est pas nécessairement un inconvénient. Dans notre cas, une partie importante de l'environnement physique était encore visible dans la vision périphérique lors de l'utilisation du mode RV. Le maintien d'un lien avec l'environnement physique peut être important dans des scénarios d'application tels que la chirurgie augmentée. Le maintien de la vision périphérique sur l'environnement physique environnant et les transitions entre les modes RA et RV semblent prometteurs pour minimiser la charge cognitive d'utilisateurs occupés comme les chirurgiens.

Recommandation	Description
R1	Inclure les deux types de couplage : des vues faiblement couplées et des vues fortement couplées afin de laisser les utilisateurs adopter des stratégies hybrides.
R2	Si la vue mixte et la vue virtuelle sont complémentaires, les concevoir avec le même niveau d'importance.
R3	Inclure une transition entre les modes RA et RV pour éviter les incohérences visuelles. Cependant, quitter ce mode RV peut être plus difficile que de quitter le mode RA. Dédier une importance particulière à la manière d'aider les participants à reprendre leur tâche après avoir quitté la RV.

TABLE 5.1 – Étude 5 : recommandations proposées pour l'exploration d'objets 3D en RM combinant une vue mixte et une vue virtuelle.

Conclusion

Ce chapitre contribue à explorer la combinaison d'une vue mixte et d'une vue virtuelle pour l'exploration d'objets 3D en réalité mixte (RM). Nous avons d'abord proposé un espace de conception basé sur différents types de couplage de vues et de techniques de navigation. Nous avons défini 4 techniques d'interaction à partir de cet espace de conception. Une étude pilote exploratoire a suggéré que l'utilisation de vues faiblement couplées peut ne pas convenir lors de déplacements physiques (navigation physique) dans l'espace de travail. Au contraire, pour la navigation virtuelle, les résultats expérimentaux d'une deuxième étude formelle mettent en évidence la complémentarité des techniques fortement couplées et faiblement couplées pour explorer des objets 3D en RM. En outre, nous avons constaté que les deux vues (mixte et virtuelle) étaient considérées de manière égale par les participants. À partir des stratégies observées et des commentaires des participants, nous proposons des recommandations de conception, résumées dans le tableau 5.1, pour aider à la conception de systèmes de RM combinant des vues mixtes et des vues virtuelles.

Tandis que les vues multiples virtuelles constituent un sujet largement étudié en visualisation d'information, nous soulignons l'originalité de nos travaux : la combinaison de vues multiples de natures différentes (mixte et virtuelle). De plus, ces travaux peuvent bénéficier à d'autres domaines applicatifs que la chirurgie augmentée. C'est par exemple le cas de l'enseignement de la géométrie 3D, où les enfants doivent explorer des formes 3D dans différents états (c'est-à-dire des polyèdres 3D et les patrons 2D correspondants) [202]. Dans ce cas, la RM pourrait être un outil puissant pour permettre aux élèves d'explorer à la fois des polyèdres mixtes comme des cubes ou des pyramides augmentés et leurs patrons virtuels. La fabrication et la maintenance des machines en RM [154, 216] partagent également des caractéristiques similaires (par exemple, avec des vues externes et internes d'une machine complexe). De même, les tâches d'assemblage en RM comme dans [172, 219] pourraient bénéficier de l'exploration d'objets à vues multiples. Le système pourrait augmenter les différents composants de l'objet tout en affichant également la forme finale de l'objet assemblé, comme illustré à la Figure 5.4.

Le chapitre suivant aborde l'implémentation de nos travaux de recherche au sein d'un prototype industriel couplé avec l'OrthoPilot®.

Chapitre 6

Prototype d'OrthoPilot® en RM

Sommaire

6.1	Fonctionnement général du prototype	148
6.1.1	OrthoPilot® sans Réalité Mixte	148
6.1.2	Ajout des techniques de Réalité Mixte	149
6.1.3	Visualisation d'informations médicales	149
6.2	Techniques de menus basées sur les mouvements de la tête	151
6.2.1	Curseur virtuel contrôlé avec la tête	151
6.2.2	Menu virtuel contrôlé avec la tête	152
6.3	Techniques d'interaction tête + main : <i>Bring2Me</i>	152
6.4	Évaluation du prototype	154

Introduction

Nos travaux se concentrent sur les techniques d'interaction utilisables en bloc opératoire pour la chirurgie augmentée avec l'entreprise Aesculap, partenaire de cette thèse CIFRE. Pour cette entreprise, l'objectif de cette thèse était d'explorer les apports de la Réalité Mixte (RM) pour la chirurgie du genou. Dans ce chapitre, nous présentons l'implémentation de nos techniques issues des Chapitres 3 et 4 au sein d'un prototype de système de RM basé sur l'OrthoPilot®[3], le logiciel de navigation chirurgicale de l'entreprise.

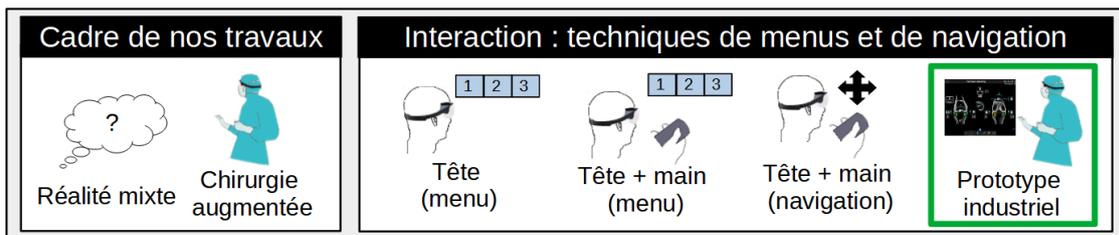


FIGURE 6.1 – Plan du mémoire (Figure B) - Focus sur le 6ème chapitre.

Les techniques de changement de point de vue détaillées au Chapitre 5 ne sont pas encore implémentées. Ces techniques ont pour but d'explorer une scène 3D en RM combinant deux types de vues (Figure 5.3) : une vue mixte (jambe du patient augmentée en extension) et une vue virtuelle (copie virtuelle de la jambe en flexion). Pour mener l'étude expérimentale 5 sur le sujet, nous avons utilisé des formes géométriques basées sur un cube : la Boîte et le Patron. Dans le cas de la chirurgie du genou, la Boîte (vue mixte) correspond à la jambe du patient en extension qui est augmentée avec la future prothèse et le Patron (vue virtuelle) un modèle virtuel de la jambe du patient en flexion avec la future prothèse. Néanmoins, implémenter l'ensemble de l'étape de planning chirurgicale au sein du prototype aurait demandé un temps de développement très important. Les techniques liées à cette étude ne sont donc pas encore incluses dans le prototype actuel par faute de temps.

6.1 Fonctionnement général du prototype

6.1.1 OrthoPilot®sans Réalité Mixte

La solution OrthoPilot®est introduite à la Section 2.1.1. Rappelons simplement ici qu'il s'agit d'un dispositif qui aide le chirurgien à positionner une prothèse de genou. Tout au long de l'opération, le chirurgien a accès aux informations médicales via un écran distant et peut interagir avec le système via une technique de raycasting. L'OrthoPilot®actuel ne comporte aucune technique de RM, comme illustré à la Figure 6.2.

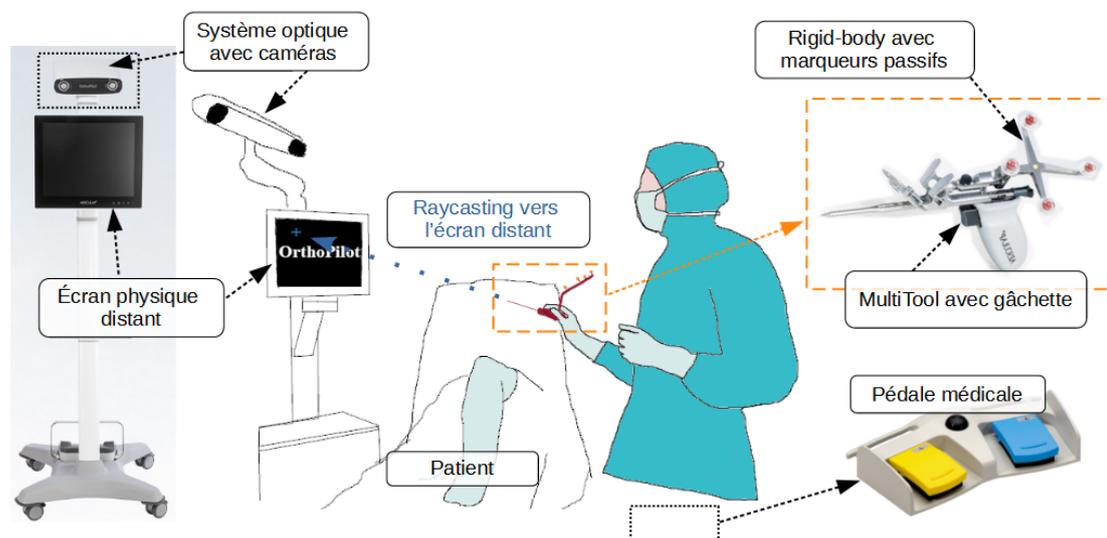


FIGURE 6.2 – L'OrthoPilot®sans Réalité Mixte : le chirurgien pointe vers l'écran distant à l'aide d'un outil chirurgical : le *MultiTool*. Cet outil est détecté par les caméras grâce à un rigid-body comportant des marqueurs réfléchissants. Une pédale médicale peut aussi être (optionnellement) présente au sein du bloc opératoire. Schéma basé sur la Figure 2.3.

Sans RM, l'ensemble de l'interface graphique de l'OrthoPilot®est concentrée sur l'écran distant, ce qui limite la quantité d'information qu'il est possible d'afficher

en même temps. La distance entre le chirurgien et cet écran, comprise entre 1 et 3 mètres, aggrave encore cette limitation. Les éléments à l'écran doivent être de taille suffisante pour être lisibles de loin.

6.1.2 Ajout des techniques de Réalité Mixte

Pour permettre au chirurgien de percevoir des objets virtuels, nous avons utilisé le même casque de RA que lors de nos études expérimentales : le Epson Moverio BT-300 [68].

Dans un premier temps, nous avons exploré une première approche dans laquelle le casque utilise ses propres capteurs et sa propre caméra pour se situer dans l'espace et repérer les autres objets liés à l'interaction. En anglais, ce principe est qualifié de *visual Simultaneous Localization And Mapping* (SLAM). Cette approche est utilisée par différents casques de RA actuels comme l'Hololens [145]. Néanmoins, la précision des capteurs du casque au cours du temps pose problème. De plus, la puissance de calcul des lunettes BT-300 est limitée. Nous avons donc abandonné l'idée de laisser les lunettes BT-300 gérer la détection de sa propre position dans l'espace.

Par conséquent, nous avons choisi de dédier le casque de RA uniquement à l'affichage des informations virtuelles via la librairie OpenGL ES 2.0. Pour la détection de la pose de la tête de l'utilisateur, nous avons préféré utiliser les caméras du système optique de l'OrthoPilot®. Ces caméras sont faites pour détecter des groupes de marqueurs réfléchissants spécifiques. Nous avons donc conçu un rigid-body personnalisé grâce une imprimante 3D. Ce rigid-body a été fixé aux lunettes afin que les caméras puissent détecter en temps réel la pose de la tête de l'utilisateur.

Pour simuler un champ opératoire augmenté, nous nous sommes basés sur une maquette en plastique représentant les os de la jambe. Cette maquette est installé sur une table en face des caméras, en position de flexion ou d'extension. Deux rigid-bodies sont fixés sur la jambe : un sur le fémur et le second sur le tibia. Un aperçu du prototype est disponible à la Figure 6.3.

Le fonctionnement du prototype est similaire à celui de nos installations expérimentales (Section 3.2.2). Les caméras de l'OrthoPilot® détectent les marqueurs présents dans le champ opératoire, y compris ceux attachés aux lunettes. Le logiciel OrthoPilot®, qui s'exécute sur un ordinateur proche, va récupérer ces positions et envoyer tout changement au casque de RA via Bluetooth. L'OrthoPilot® reposant sur un programme C++, un protocole de communication basé sur des messages Json a été mis en place afin de pouvoir envoyer des commandes au programme Android qui s'exécute sur le casque. Ce programme analyse les messages reçus et met à jour le système et l'affichage en conséquence.

6.1.3 Visualisation d'informations médicales

Le premier objectif du prototype est de souligner le potentiel de la RM pour la chirurgie en montrant à quoi pourrait ressembler un bloc opératoire mixte. Pour cela, nous avons défini deux zones dans lesquelles ajouter du contenu virtuel : le champ opératoire et la zone distante.

La première zone est le champ opératoire, c'est à dire la zone à proximité immédiate du patient (Figure 6.3b). Au sein de cette zone, nous avons ajouté deux

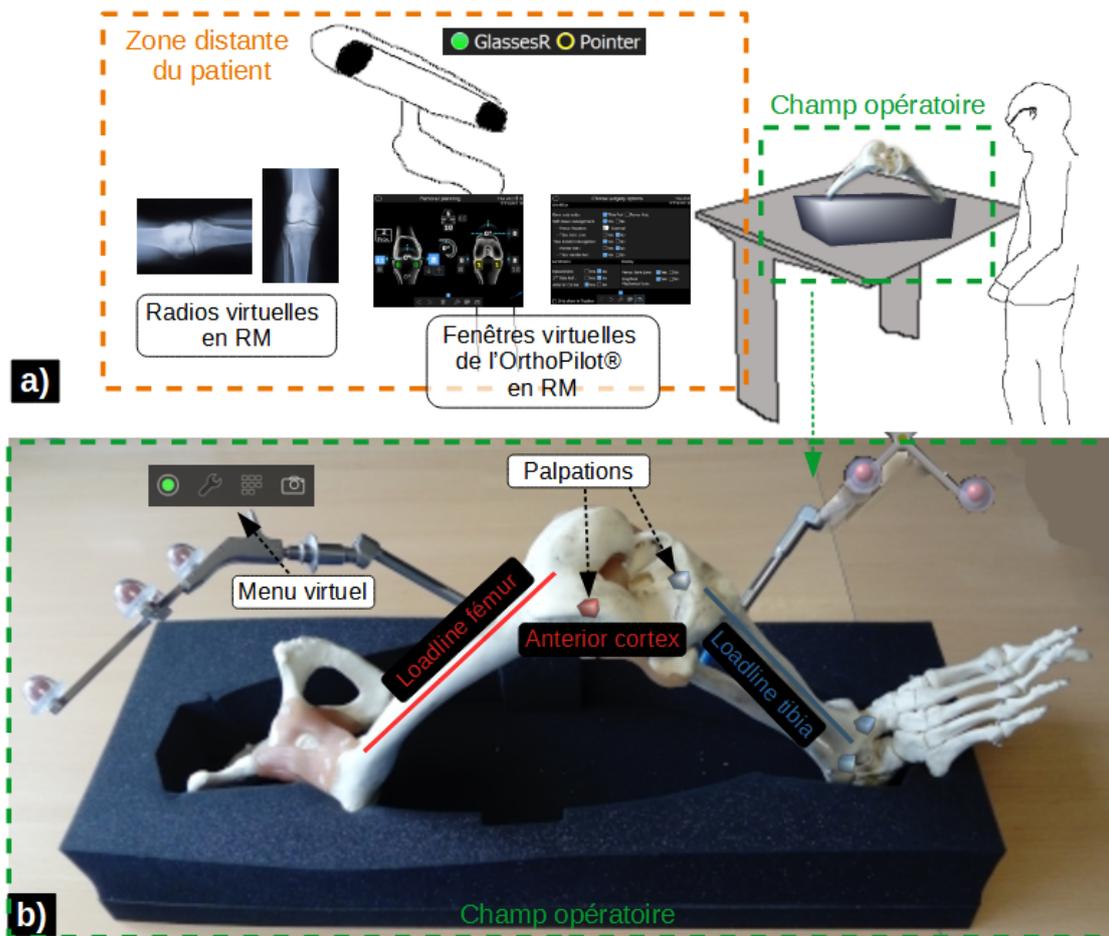


FIGURE 6.3 – Prototype de système de RM combiné à l'OrthoPilot®classique. a) Aperçu global du prototype. b) Schéma du champ opératoire vu au travers du casque.

types d'objets virtuels. Tout d'abord, des informations médicales ont été superposées sur le corps du "patient" en augmentant la maquette en plastique de la jambe. Par exemple, au début d'une opération, le chirurgien va pointer plusieurs repères anatomiques de référence sur le corps du patient et enregistrer leur position au sein du système. Ces points anatomiques permettent à l'OrthoPilot®de modéliser de manière précise la jambe du patient. En nous appuyant sur l'OrthoPilot®, nous avons implémenté cette étape chirurgicale au sein du prototype. Chaque palpation est rendue visible grâce à un cube virtuel suivant le code couleur habituel du système (bleu pour les palpations liées au tibia et rouge pour le fémur). Ainsi, le chirurgien peut voir à tout instant les points qu'il a palpés sur le patient (Figure 6.3b).

Une fois l'ensemble des palpations enregistrées dans le système, l'OrthoPilot®calcule le modèle géométrique de la jambe du patient, dont son axe mécanique. Cette droite imaginaire représente la droite par laquelle l'essentiel du poids du corps va traverser la jambe. Il s'agit d'une information clé pour le travail du chirurgien, qui doit veiller à ce que la pose de la prothèse soit en adéquation avec cet axe mécanique. Dans notre prototype, il est possible de visualiser cet axe directement sur le corps du patient grâce à l'ajout de droites virtuelles (Figure 6.3b).

La seconde zone dans laquelle nous avons ajouté du contenu virtuel est située au delà du champ opératoire, c'est à dire la zone éloignée du patient. Cela corres-

pond par exemple à la position classique de l'écran de l'OrthoPilot®[®], entre 1 et 3 mètres de la table d'opérations. Pour illustrer à quoi pourrait ressembler une salle d'opération mixte, nous avons ajouté dans cette zone plusieurs objets virtuels qui n'ont pas besoin d'être présents en permanence dans le champ opératoire. Ainsi, plusieurs exemples de radios virtuelles sont affichées. L'interface graphique de l'OrthoPilot®[®] est étendue à plusieurs fenêtres virtuelles au lieu d'être limité à une seule fenêtre sur l'écran physique. Des données techniques comme l'état des caméras sont également disponibles via divers panneaux virtuels (Figure 6.3a).

Notre prototype permet donc de visualiser à la fois des informations médicales pré-enregistrées (les palpations, les radios) ou des données calculées (l'axe mécanique de la jambe). A cela s'ajoutent des menus virtuels contextuels à proximité immédiate du patient, comme détaillé en Section 6.2.

6.2 Techniques de menus basées sur les mouvements de la tête

Au delà de la simple visualisation d'information, l'interaction avec du contenu virtuel est également important pour le travail du chirurgien. Nous décrivons dans cette section l'implémentation au sein du prototype de nos techniques d'interaction évaluées en laboratoire.

6.2.1 Curseur virtuel contrôlé avec la tête

Les techniques d'interaction basées sur la tête que nous proposons (Section 3.2) ont été implémentées au sein du prototype. La technique $HCC \rightarrow Virtuel$ permet de sélectionner des éléments du menu virtuel classique de l'OrthoPilot®[®], appelé *ActionBar*. La confirmation de la sélection peut être effectuée avec la pédale ou via le mécanisme de temporisation (*dwell-time*). La méthode de confirmation par temporisation est connue des chirurgiens et est utilisée lorsqu'ils pointent en raycasting sur l'écran. Lors de cette temporisation, le nom de l'élément ciblé est également affiché temporairement. La Figure 6.4 illustre cette technique implémentée au sein du prototype.

La technique $HCC \rightarrow Mixte$ a également été implémentée au sein du prototype. Elle complète un dispositif déjà existant au sein de l'OrthoPilot®[®] : la sélection d'une commande basée sur un rigid-body, ou *toolProximity*. Dans l'OrthoPilot®[®] classique, les chirurgiens peuvent approcher la pointe d'un outil chirurgical vers un des marqueurs des deux rigid-bodies de la jambe. Une fois l'outil suffisamment proche, la commande associée à ce marqueur est déclenchée suite à une temporisation (*dwell-time*). Néanmoins, sans casque de RM, le chirurgien doit apprendre et retenir quelle commande est associée à chaque marqueur, ce qui impacte sa charge cognitive. La technique $HCC \rightarrow Mixte$ permet de voir cette correspondance grâce aux objets virtuels superposés sur chaque marqueur. Par ailleurs, il est possible de déclencher ces commandes rien qu'en amenant le curseur au centre de l'écran vers l'élément désiré. Il n'est donc plus nécessaire d'avoir un outil à la main pour interagir. L'implémentation de cette technique est illustrée à la Figure 6.5.

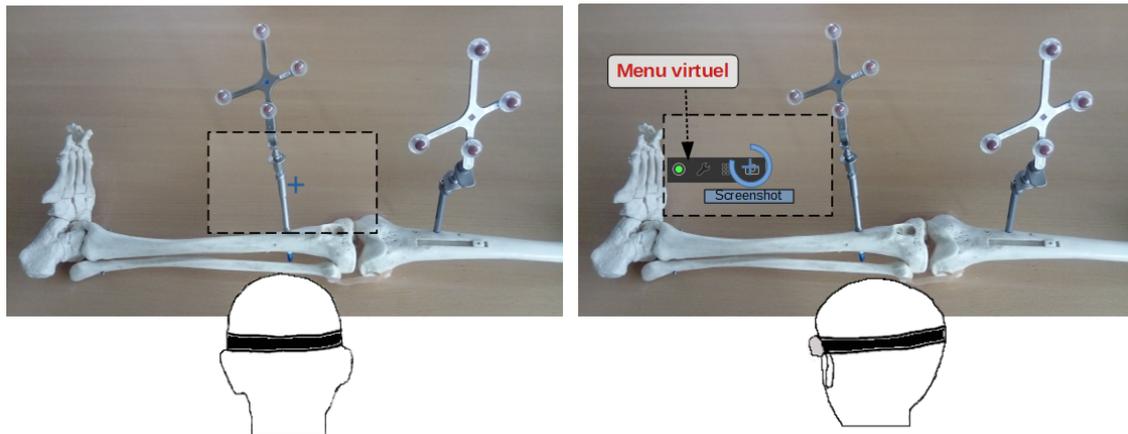


FIGURE 6.4 – Technique HCC→Virtuel implémentée au sein du prototype. L’arc de cercle bleu représente l’avancée de la temporisation (1 seconde de survol nécessaire afin de confirmer la sélection)

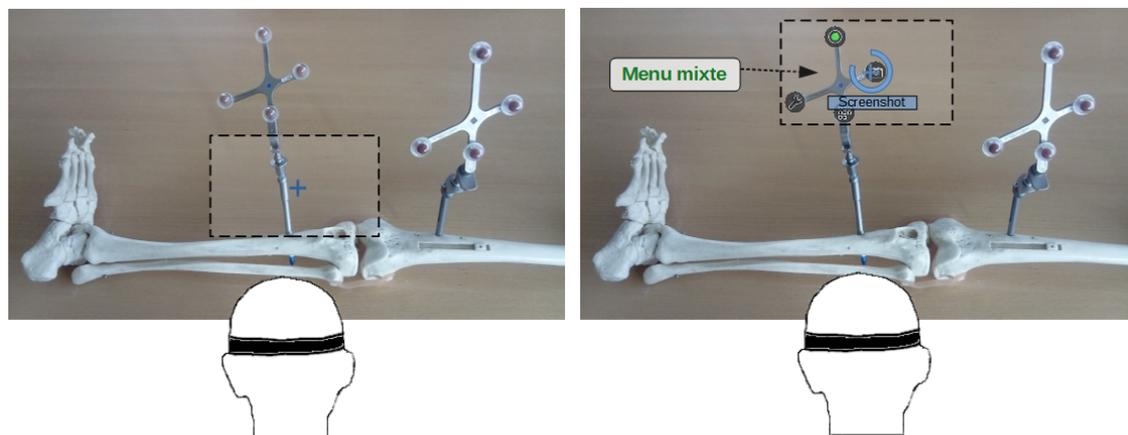


FIGURE 6.5 – Technique HCC→Mixte implémentée au sein du prototype. L’arc de cercle bleu représente l’avancée de la temporisation (1 seconde de survol nécessaire afin de confirmer la sélection).

6.2.2 Menu virtuel contrôlé avec la tête

Enfin, les deux technique HCM ont également été ajoutées au prototype. Le menu *ActionBar* peut ainsi être contrôlé avec la tête et amené jusqu’à un curseur virtuel ou mixte (un marqueur de rigid-body mixte). L’implémentation de ces deux techniques reste néanmoins à affiner. Tout d’abord, le moment précis de l’apparition du menu au centre de l’écran au début de l’interaction pose encore question, et nécessite de plus amples recherches. Ensuite, le prototype est actuellement limité à un seul menu pouvant être contrôlé avec la tête. L’interaction avec plusieurs menus n’est pour l’instant pas implémentée au sein du prototype.

6.3 Techniques d’interaction tête + main : *Bring2Me*

Les techniques *Bring2Me* du Chapitre 4 ont également été implémentées au sein du prototype. A l’aide d’un outil chirurgical appelé le *MultiTool* (voir Figure 6.2),

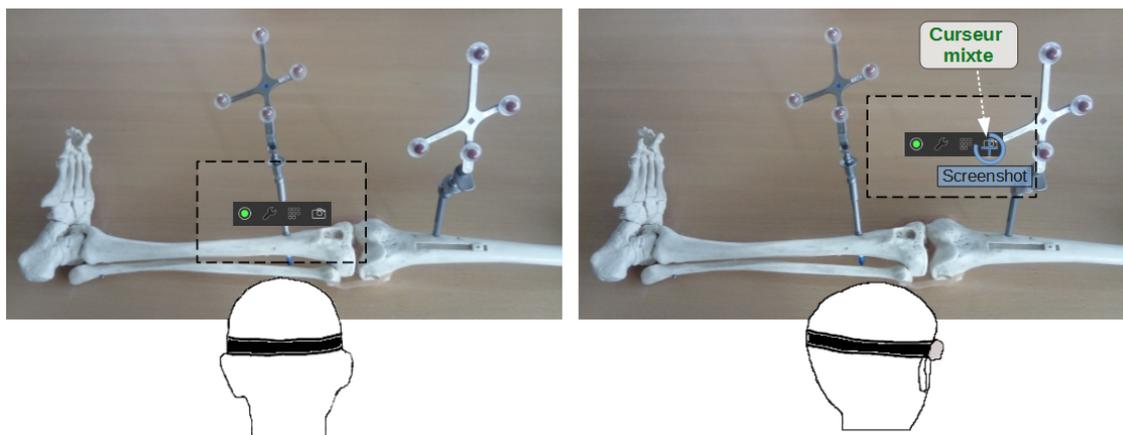


FIGURE 6.6 – Technique HCM→Mixte implémentée au sein du prototype. L’arc de cercle bleu représente l’avancée de la temporisation.

l'utilisateur peut ramener un objet virtuel dans son champ de vision sans quitter des yeux le champ opératoire. Pour cela, nous avons utilisé la gâchette mécanique intégrée au MultiTool comme déclencheur. Un simple clic permet de téléporter l'objet virtuel visé dans le champ de vision. Dans le cas d'un menu, un élément peut ensuite être sélectionné à l'aide du curseur au centre de l'écran (technique *TeleHead*) ou via raycasting avec le MultiTool (technique *TelePad*).

Dans le cas de la technique *Grab*, l'utilisateur doit viser un objet virtuel avec le MultiTool, puis maintenir la gâchette de l'outil enfoncée pour déplacer cet objet. Nous avons cependant constaté une latence significative lors du déplacement continu d'un objet virtuel. Il semble en effet que le casque n'a pas la capacité technique de traiter l'ensemble des messages de mise à jour afin d'obtenir un déplacement fluide. Des optimisations techniques sont sans doute possibles pour limiter la latence visuelle et conserver le geste continu de la technique *Grab*. Néanmoins, nous avons préféré implémenter une variante de cette technique ne souffrant pas de ce problème.

Cette variante, nommée *SilentGrab*, évite ce problème de latence en ne mettant à jour la position de l'élément attrapé qu'au relâchement de la gâchette du MultiTool. L'utilisateur effectue toujours un geste continu du poignet ou du bras, mais ne voit apparaître le menu qu'une fois son geste de glisser/déposer terminé. Pendant l'interaction, un retour textuel est affiché en bas de l'écran du casque pour indiquer quel objet virtuel est actuellement déplacé. Cette approche évite les problèmes de latence mentionnés ci-dessus. Les bénéfices de cette variante restent cependant encore à évaluer expérimentalement.

Par ailleurs, nous nous sommes inspirés du MultiTool tenu à la main par les chirurgiens pour implémenter un système de *marking menu* [221]. L'idée est de faire apparaître un menu circulaire dans le champ de vision augmenté en cliquant sur la gâchette de l'outil. Ensuite, l'utilisateur peut sélectionner un élément (ou sous-élément) en traçant une marque dans les airs dans la direction de la cible avant de relâcher la gâchette. En mode expert, le menu n'apparaît plus dans le champ de vision afin de limiter toute occultation visuelle : l'interaction devient purement gestuelle en mid-air.

6.4 Évaluation du prototype

Le prototype a fait tout d'abord l'objet de deux démonstrations en interne au sein de l'entreprise. L'objectif de ces démonstrations était de mettre en valeur les bénéfices de la RM pour la chirurgie et l'intérêt d'approfondir cette technologie dans les années à venir. Ces démonstrations ont été faites auprès de membres de l'équipe de direction et de collègues d'un autre service de la maison mère Bbraun, en Allemagne.

Deux apprentis chirurgiens du CHU de Grenoble ont également pu tester le prototype lors d'une visite dans les locaux d'Aesculap. Les deux internes n'ont pas pu tester l'ensemble des techniques d'interaction présentes dans le prototype. Néanmoins, ces démonstrations ont tout de même permis de collecter des retours qualitatifs vis à vis du prototype. Tout d'abord, la visualisation d'information médicale directement sur le corps du "patient" (*i.e.* sur les os en plastique) a été appréciée par les chirurgiens. La possibilité de voir des données enregistrées (les palpations) et calculées (l'axe mécanique de la jambe) a été considérée comme prometteuse. Ne plus être limité à un seul écran a aussi été jugé comme un élément important. Cela permet en effet au chirurgien d'accéder à plus d'informations réparties autour du champ opératoire.

Les deux internes ont souligné la nécessité de pouvoir masquer l'ensemble des objets virtuels à n'importe quel moment. La possibilité de couper toute RM pour se concentrer uniquement sur le patient semble être un impératif fort pour le chirurgien. En revanche, ils ont confirmé qu'un mode VR temporaire était parfaitement envisageable lors d'une opération. Tant que le chirurgien reste en situation de contrôle, passer en mode VR le temps de l'interaction leur a semblé pertinent si cela les aide dans leur travail. Un tel retour est encourageant pour les transitions entre les modes RA et RV utilisées dans nos techniques d'interaction.

Par ailleurs, les deux internes se sont montrés enthousiastes vis à vis du principe des techniques *Bring2Me*. Ils ont immédiatement évoqué un scénario où ils pourraient ramener une radio virtuelle tout en restant concentrés sur le patient. Les techniques permettant d'interagir avec du contenu virtuel sans impliquer l'attention visuelle du chirurgien semblent particulièrement prometteuses en bloc opératoire. Les deux internes étaient également intéressés par des questions techniques au niveau des lunettes de RM, notamment par la possibilité de les nettoyer après des projections de sang. La taille réduite du champ de vision augmenté a également été citée comme une limitation importante puisqu'il n'est pas possible de voir l'ensemble de la jambe augmentée d'un seul coup.

Enfin, une discussion plus complète autour du prototype était prévue début 2020 avec un chirurgien travaillant dans un hôpital britannique. Cette rencontre a néanmoins du être annulée suite à la propagation de la COVID-19.

Conclusion

Grâce à la plateforme existante de l'entreprise Aesculap partenaire de cette thèse Cifre, nous avons pu concrétiser les résultats de nos travaux de recherche au sein d'un prototype. Ce prototype se base sur un logiciel professionnel de navigation chirurgicale, l'OrthoPilot®[®], qui ne disposait pas de technique de RM. Nous avons

implémenté les premières étapes d'une chirurgie du genou en RM afin de mettre en valeur le potentiel de cette technologie pour le travail du chirurgien. De plus, plusieurs de nos techniques d'interaction conçues et évaluées expérimentalement en laboratoire ont été implémentées au sein de ce prototype.

Pour l'instant, l'évaluation du prototype est limitée à des démonstrations en interne de l'entreprise ainsi qu'une démonstration à deux apprentis chirurgiens. Les premiers retours sont néanmoins encourageants sur plusieurs aspects du prototype et de nos techniques d'interaction.

Conclusion et perspectives générales

Résumé des contributions

La Réalité Mixte (RM) vise à combiner des éléments physiques et des éléments virtuels pour former un espace de travail mixte. L'ensemble de l'environnement devient alors un espace d'interaction pour l'utilisateur. Ainsi, la RM bouscule les interfaces utilisateurs traditionnelles basées sur un écran physique pour définir des interfaces qualifiées de post-WIMP. Combiner les avantages de l'environnement physique avec la flexibilité des objets virtuels offre de multiples possibilités pour l'entraînement, le guidage ou encore la coopération des utilisateurs. De nombreux secteurs professionnels commencent à mettre en œuvre des solutions de RM, par exemple des cockpits d'avion augmentés et des systèmes de maintenance de machines industrielles.

L'un des domaines d'application majeur de la RM est le domaine médical et en particulier la chirurgie. La RM apporte en effet une aide précieuse au chirurgien. Pouvoir visualiser des informations médicales directement sur le corps du patient, s'entraîner sur un mannequin augmenté ou encore coopérer à distance avec un autre chirurgien expert sont autant d'exemples du potentiel de la chirurgie augmentée. Une opération chirurgicale est une tâche particulièrement critique, complexe et stressante. Aider le chirurgien à mener à bien une opération grâce à un système de RM est donc crucial, aussi bien pour l'équipe médicale que pour le patient.

Néanmoins, de nombreuses contraintes doivent être respectées au sein du bloc opératoire. La stérilité stricte de tout élément à proximité du patient, l'environnement bruyant et le temps limité définissent des exigences à prendre en compte lors de la conception de techniques d'interaction en RM. Ainsi, de nombreuses techniques existantes ne sont pas adaptées à ces contraintes et ne sont donc pas utilisables en bloc opératoire.

Dans ce contexte, nos travaux de recherche ont visé à répondre à ce besoin de nouvelles techniques permettant au chirurgien d'interagir avec du contenu virtuel lors d'une opération. Guidés par les exigences de ce domaine d'application, nos travaux se concentrent sur les techniques de RM basées sur un casque (*Head-Mounted Display*). Nous avons centré nos recherches sur les dispositifs déjà présents au sein du bloc opératoire, notamment certains outils et équipements chirurgicaux existants. Par ailleurs, nos contributions concernent des étapes de criticité différentes de la chirurgie et en particulier de la chirurgie orthopédique.

Dans un premier temps, nous considérons les étapes chirurgicales critiques au cours desquelles le chirurgien a les deux mains occupées. Dans ce cas, l'interaction possible avec le système est limitée et ponctuelle. Nous étudions l'interaction avec

un menu à un seul niveau. Pour cela, nous proposons un espace de conception pour l'interaction avec un menu basée sur les mouvements de la tête, qui respecte les contraintes du bloc opératoire. Nous explorons cet espace de conception en définissant trois nouvelles techniques d'interaction basées sur des menus virtuels ou mixtes. Ces techniques ont été évaluées lors de deux études expérimentales en laboratoire. Nous avons ensuite comparé ces deux types de menus à des menus physiques lors d'une troisième étude. Les résultats expérimentaux suggèrent que contrôler un menu avec la tête peut conduire à des performances similaires à un curseur contrôlé avec la tête, tout en étant plus confortable à utiliser. Par ailleurs, ancrer physiquement un menu virtuel semble permettre aux utilisateurs d'atteindre plus rapidement des performances optimales, un résultat important dans le cadre de la chirurgie augmentée.

Dans un second temps, nous élargissons nos recherches aux étapes chirurgicales moins critiques. Le chirurgien a alors au moins une main libre pour interagir avec le système. Nos contributions concernent deux tâches fréquentes en RM : spécifier une commande à l'aide d'un menu et explorer une scène 3D mixte. Tout d'abord, nous proposons trois nouvelles techniques d'interaction avec un menu nommées Bring2Me. Le but de ces techniques est d'exploiter la main libre pour permettre à l'utilisateur d'interagir avec des menus distants sans avoir à changer de contexte visuel, contrairement aux techniques précédentes de menus basés uniquement sur les mouvements de la tête. Les techniques Bring2Me permettent de ramener par un geste de la main un menu virtuel dans le champ de vision pour préserver le contexte visuel initial. Les résultats expérimentaux d'une étude sont très encourageants à cet égard. Les techniques Bring2Me ont conduit à de meilleures performances que la technique de référence (tourner la tête pour interagir avec le menu) et ont été préférées par les participants.

Enfin, nous allons au-delà de la tâche de spécification de commandes via un menu en étudiant l'exploration interactive d'une scène 3D mixte comme un champ opératoire augmenté. Nous considérons une scène 3D mixte combinant une vue augmentée et une vue virtuelle (par exemple en orthopédie la jambe augmentée du patient en extension et une copie virtuelle de la jambe du patient en flexion). Nous étudions deux facteurs clé de cette tâche d'exploration : le couplage des deux types de vues et la technique de changement de point de vue. Une expérimentation en laboratoire nous a permis d'étudier les stratégies d'exploration possibles et de proposer un ensemble de recommandations pour la conception de vues multiples de nature différente (vue mixte et virtuelle) en RM.

Ces contributions ont fait l'objet de deux publications dans des conférences internationales en Interaction Homme-Machine : Interact'2019 [15] et AVI'2020 [16]. Deux autres articles sont actuellement soumis et en cours d'évaluation. Nous schématisons l'ensemble de nos contributions à la Figure C.

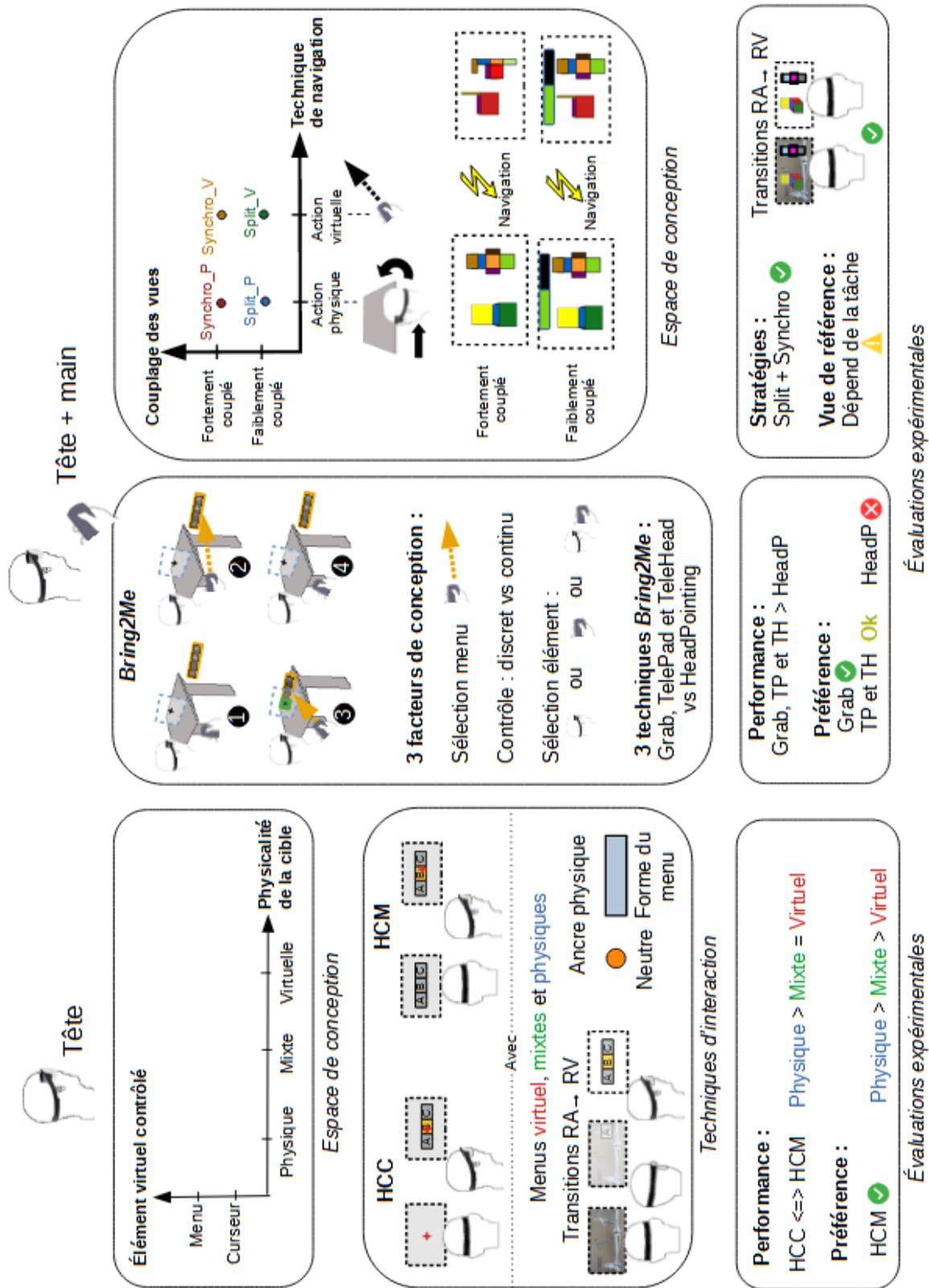


FIGURE C – Résumé des contributions.

Nous identifions plusieurs axes pour la poursuite de ce travail de recherche, dont la portée se veut à la fois générale pour les techniques d'interaction et locale pour le contexte per-opératoire. Nous organisons ces perspectives en deux classes, celles à court terme qui sont des extensions directes de résultats obtenus et celles à plus long terme qui ouvrent de nouvelles voies de recherche.

Perspectives à court terme

Nous identifions une première perspective à court terme qui vise à aller plus loin dans l'étude des techniques en contexte chirurgical (en d'autres termes, qui vise à augmenter la validité écologique des travaux expérimentaux). Ensuite, nous identifions des extensions à nos travaux sur les menus et sur la physicalité des cibles en RM.

Prototype et évaluation avec des chirurgiens

Les premiers travaux envisagés consistent à compléter le prototype conçu pour enrichir l'OrthoPilot® avec des techniques d'interaction en RM. Ces travaux peuvent être divisés en deux phases : une phase de développement et une phase d'évaluation.

La phase de développement consiste principalement à étendre les fonctionnalités existantes du prototype (Chapitre 6). En effet, nos techniques d'interaction ne sont pas toutes développées dans l'OrthoPilot®. Plus précisément, les techniques de changement de point de vue pour l'exploration de scène 3D en RM que nous proposons au Chapitre 5 ne sont pas encore intégrées au prototype. Ces techniques sont importantes à intégrer car elles répondent directement aux besoins de l'étape de planning chirurgical. L'objectif serait ici d'évaluer les techniques dans des conditions plus écologiques. En effet, nous avons mené à dessein une évaluation expérimentale la plus abstraite possible afin de pouvoir généraliser les résultats. Or, nous avons observé que la stratégie des participants pour explorer l'objet au travers des vues multiples était souvent liée à la tâche elle-même (dans notre cas, trouver et compter les erreurs).

Maintenant que nous avons des résultats généraux sur une tâche abstraite, nous envisageons d'étudier une tâche proche de celle effectuée lors de l'étape de planning chirurgical. Nous voudrions notamment remplacer la Boîte et le Patron par une maquette de jambe augmentée et une copie virtuelle de la jambe. L'utilisateur pourrait non seulement interagir avec la jambe du patient augmentée, mais aussi avec une copie virtuelle de cette jambe. Par ailleurs, nous souhaitons inclure l'édition de valeurs (au lieu de la simple inspection) afin de nous rapprocher d'un scénario réaliste en bloc opératoire. Cela permettrait également d'approfondir différents types de couplage des vues multiples en allant au-delà de la navigation coordonnée ou indépendante entre les vues.

Par ailleurs, un autre volet de cette phase de développement serait de reprendre l'implémentation de nos travaux de recherche avec un casque de RM plus performant. Le développement rapide des technologies associées fait qu'après trois ans de travaux doctoraux, des casques récemment disponibles comme le casque Hololens 2 disposent d'un champ de vision sensiblement plus important, d'une puissance de calcul supérieure et intègre bien plus de capteurs (par exemple, le suivi du regard

et des doigts de l'utilisateur). Par exemple le changement de point de vue dans les techniques du Chapitre 5 pourrait se faire par un geste de la main plutôt que de tenir la télécommande du casque Moverio BT-300. Utiliser un casque plus performant mettrait davantage en valeur les atouts des techniques en RM conçues pour aider le travail des chirurgiens.

L'objectif de cette première phase de développement serait de préparer la seconde phase des travaux futurs : l'évaluation de nos contributions par des chirurgiens. En effet, nous avons pour l'instant fait le choix de mener des études expérimentales en laboratoire afin de pouvoir généraliser les résultats expérimentaux à plusieurs contextes d'application potentiels. Les évaluations informelles que nous avons conduit auprès d'internes ont donné des résultats encourageants, mais les techniques d'interaction que nous proposons restent à évaluer dans des conditions plus écologiques.

Pour ce faire, nous identifions deux approches complémentaires. D'une part, mener des études qualitatives et prospectives auprès de chirurgiens permettrait d'étudier en profondeur l'utilisabilité de nos techniques d'interaction et de dégager de nouvelles pistes de recherche à partir des retours des chirurgiens. D'autre part, mener des expérimentations dans un cadre plus proche du bloc opératoire nous permettrait de tester nos techniques d'interaction selon différents scénarios et en conditions pseudo-écologiques. Pour cela, il conviendrait de développer une procédure chirurgicale complète en RM, effectuée avec des instruments chirurgicaux et une maquette de la jambe du patient. Nous mettrions en place cette installation dans la salle d'expérimentation d'Aesculap. Le but serait de faire tester cette procédure chirurgicale à des experts (connaisseurs de l'OrthoPilot®), ainsi qu'à du personnel médical (étudiants en médecine et apprentis chirurgiens).

Techniques de menu

Au cours de nos évaluations expérimentales, les techniques basées sur un menu contrôlé avec la tête (HCM) se sont avérées prometteuses par rapport aux techniques basées sur un curseur virtuel contrôlé avec la tête. Nous n'avons pas observé de différence de performance entre les deux approches, mais les retours qualitatifs des participants soulignent une préférence pour HCM. A partir de ces résultats, deux axes de recherche se dessinent.

Le premier axe concerne la conception du menu pour les techniques HCM. Dans nos travaux, nous avons fait le choix d'utiliser un menu simple, avec un seul niveau hiérarchique et 4 éléments. Ce choix était motivé par la volonté de refléter les menus existants de l'OrthoPilot®, auquel les chirurgiens sont habitués. Les techniques HCM doivent par conséquent être évaluées sur la base de menu plus complexes, notamment avec des menus hiérarchiques. L'un des avantages des techniques HCM, pouvoir voir en amont les éléments du menu, pourrait être limité par l'espace visuel disponible sur l'écran du casque. Dans le cas d'un menu complexe avec plusieurs niveaux, de nombreux éléments pourraient ainsi être masqués au début de l'interaction.

Un second axe de recherche serait d'étudier de manière plus précise l'apparition du menu dans le champ de vision avec une technique HCM. En effet, le protocole expérimental de nos études n'impliquait que des tâches de pointage successives. Dans

un cas concret, l'utilisateur a une tâche principale à effectuer et non uniquement du pointage. L'apparition d'un menu au centre du champ de vision peut perturber cette tâche principale. La question est donc de savoir à quel moment le système peut considérer que l'utilisateur initie une interaction avec un menu donné.

Dans le cadre de la chirurgie augmentée, plusieurs menus contextuels peuvent être présents dans l'environnement mixte à des positions spatiales spécifiques (sur un outil, sur la caméra, etc). Ainsi, le menu pourrait apparaître dans le champ de vision du chirurgien une fois que celui-ci tourne la tête vers une de ces zones. Le moment optimal de l'apparition du menu est alors un compromis entre 1) éviter les déclenchements involontaires qui perturberaient la tâche chirurgicale et 2) faire apparaître le menu au plus tôt pour que ses éléments soient visibles. Une solution alternative au mouvement de la tête vers un menu serait une action de l'utilisateur pour indiquer le mode menu, comme un appui de la pédale. Le risque d'une technique en deux temps (passage explicite en mode HCM, puis sélection d'un élément) est néanmoins d'alourdir l'interaction. Enfin, il serait intéressant de comparer ces approches à des techniques faisant sauter la cible (ici, le menu) vers l'utilisateur [5, 163].

Techniques *Bring2Me*

Les techniques *Bring2Me* décrites au Chapitre 4 ouvrent des perspectives intéressantes sur cette question de l'apparition du menu dans le champ de vision. En effet, elles pourraient être couplées avec des techniques HCM afin de mutualiser les bénéfices des deux approches. L'évaluation des techniques *Bring2Me* a mis en valeur le fait que faire sauter le menu dans le champ de vision est envisageable en RM. Dans notre cas, les techniques *TeleHead* et *TelePad* ont un comportement déterministe : le menu sélectionné est systématiquement téléporté près de la bordure supérieure du champ de vision augmenté. Au final, le sentiment de désorientation [161] semble avoir été limité voir absent dans notre étude. De plus amples recherches sont nécessaires pour étudier l'impact précis d'un contrôle discret (*TelePad* et *TeleHead*) de la position du menu versus un contrôle continu (*Grab*) de cette position.

Une seconde piste de recherche concernant ces techniques serait de les comparer à d'autres techniques de pointage impliquant la main de l'utilisateur. En effet, nous nous sommes concentrés dans notre étude sur la préservation du contexte visuel initial de l'utilisateur lors du pointage. Nous avons comparé nos trois techniques à une technique de référence basée uniquement sur les mouvements de tête qui fait perdre ce contexte visuel. Néanmoins, il est à présent nécessaire de quantifier la différence de performance entre les techniques *Bring2Me* et des techniques de raycasting existantes. Quantifier cette différence de performance permettrait en effet de mieux de mieux cerner le compromis de conception entre préservation du contexte visuel et performances.

Enfin, notre étude se limite pour l'instant à l'interaction avec des menus et plus généralement d'actions discrètes. Nous envisageons d'élargir nos travaux à d'autres types de widgets impliquant des actions continues comme un slider virtuel. Nous voudrions explorer une tâche d'édition de valeurs au cours de laquelle l'utilisateur pourrait rester visuellement concentré sur l'objet édité. Pour le chirurgien, cela consisterait par exemple à tester différentes positions d'une prothèse virtuelle

(coordonnées spatiales, orientation, etc.) pour en évaluer l'effet sur la cinématique de la jambe.

Physicalité des cibles

Au sein du Chapitre 3, nous nous sommes intéressés à la physicalité des menus (menus virtuels, mixtes et physiques). Plus précisément, nous avons étudié l'aspect perceptuel lié à cette physicalité et notamment la visibilité des éléments du menus avec ou sans casque de RM et l'impact sur la flexibilité de l'interface. L'étape suivante de ces travaux est d'approfondir l'impact de la physicalité sur l'apprentissage de l'interaction avec les menus et la mémorisation des éléments. Ainsi, les aspects cognitifs tels que les métaphores d'interaction [131] restent à étudier sous l'angle de la physicalité des menus et widgets.

Dans le cadre de la chirurgie augmentée, l'objectif est toujours de faciliter le travail du chirurgien. Cela passe par deux axes : 1) faciliter la prise en main de l'environnement mixte (le bloc opératoire augmenté) et 2) limiter la charge cognitive liée à l'interaction. Pour cela, une piste prometteuse serait d'approfondir les travaux de Perrault et al. sur la mémorisation via la technique des *Physical loci* [164]. Les auteurs ont proposé une technique de mémorisation basée sur des liens entre des commandes et des objets physiques environnants. Ces liens font appel à la fois à la mémoire spatiale et à des associations sémantiques. Dans le cadre de nos travaux, il serait intéressant d'évaluer l'impact d'une telle technique sur la prise en main de l'environnement mixte. Pour le chirurgien, cela reviendrait à organiser les objets virtuels, mixtes et physiques afin de limiter au maximum la phase d'apprentissage de l'utilisation du système de RM et la charge cognitive lors de l'interaction. Ainsi, nous pouvons imaginer positionner les différents objets virtuels/mixtes/physiques selon des critères comme des liens sémantiques (par exemple, un menu lié à la visibilité des objets qui serait ancré sur le bras de la caméra de l'OrthoPilot®). Enfin, il serait intéressant d'évaluer l'impact de la physicalité des widgets sur la mémorisation.

Perspectives à long terme

Pour nos travaux sur les techniques d'interaction avec un casque de RA, nous avons fait le choix de ne pas introduire de nouveaux capteurs et de nous baser sur le matériel utilisé aujourd'hui en bloc opératoire. En effet nous avons cerné nos travaux au remplacement de l'écran physique distant par le casque de RA. Des techniques exploitant de nouveaux capteurs sont envisageables en particulier parce que le casque permet d'obtenir un retour visuel sur l'interaction dans un espace moteur distinct. Une technique qui nous semble particulièrement prometteuse est de détecter les mouvements de pieds en bloc opératoire [86] puisque le chirurgien travaille debout derrière une table d'opération. L'objectif serait alors de remplacer la pédale physique parfois contraignante et d'exploiter le retour visuel affiché dans le casque pour augmenter la richesse du vocabulaire gestuel au pied.

Organisation de l'environnement mixte en bloc opératoire

Au-delà de nouvelles techniques d'interaction, nos travaux susmentionnés soulèvent la question plus générale et complexe de l'organisation de l'environnement

mixte au sein du bloc opératoire. Dans nos travaux, nous avons étudié l’ancrage du contenu virtuel sur des objets physiques. En considérant la physicalité des widgets et leur position spatiale, nous nous sommes placés dans le repère du bloc opératoire. Il est néanmoins possible d’utiliser un autre repère : celui du chirurgien lui-même. Plutôt que de placer le contenu virtuel dans la pièce (par exemple, en créant des liens sémantiques), il est envisageable de positionner ce contenu relativement au chirurgien (par exemple, les radios à gauche et les menus à droite du chirurgien). Nous pouvons également imaginer un scénario dans lequel le chirurgien personnaliserait lui-même son espace de travail avant l’opération afin d’être le plus efficace possible. Comparer ces approches en menant des évaluations expérimentales auprès de chirurgiens permettrait d’adapter au mieux l’environnement mixte.

Enfin nos travaux n’ont pas traité la dimension multi-utilisateur présente au sein de l’environnement mixte du bloc opératoire. Tout d’abord, nous partons de la coopération et de la coordination fortes au sein de l’équipe médicale physiquement présente en salle d’opération. Avec un système de navigation chirurgicale classique (sans RM), les membres de l’équipe médicale peuvent suivre l’avancée de la procédure sur l’écran distant où est affichée l’interface graphique du système. Cela leur permet notamment d’anticiper les étapes chirurgicales à venir (par exemple, en préparant les outils chirurgicaux appropriés) et de gagner un temps précieux. Avec la RM, seules les personnes portant un casque peuvent percevoir les objets virtuels. Au-delà de la conscience de groupe, se pose également la question de l’interaction. Le chirurgien a certes la possibilité d’interagir avec l’environnement mixte, par exemple grâce aux techniques *Bring2Me* que nous présentons au chapitre 4. Cependant, nous anticipons des scénarios dans lesquels l’équipe médicale pourrait aussi interagir avec les objets virtuels. L’objectif serait alors de gagner du temps et de faciliter la prise de décision du chirurgien, par exemple en préparant à l’avance un groupe de radios virtuelles données ou différentes simulations sur des copies virtuelles de la jambe du patient. Mener des interviews permettrait de mieux cerner combien d’utilisateurs du système de RM serait nécessaire lors d’une opération et quels types de manipulation du contenu virtuel ou mixte sont requises.

La question de la coopération et en particulier de l’interaction multi-utilisateurs se pose également pour le cas d’experts à distance. Certains travaux commencent à étudier comment permettre au chirurgien de travailler avec des experts distants en RM [6]. D’autres recherches se concentrent sur les notions de coopération et d’espace partagé [24], y compris pour la télé-chirurgie [64, 140]. L’organisation de l’espace de travail, qui est intrinsèquement 3D en bloc opératoire, soulève cependant encore de nombreuses questions. Nous pouvons notamment citer :

- La gestion de l’espace partagé, notamment au niveau du patient. Le but de cet espace est principalement de guider le chirurgien en présentiel ou de prendre des décisions à partir de l’anatomie du patient. Se posent alors des enjeux de guidage précis à distance, de gestion des droits sur les objets virtuels du champ opératoire et d’occlusion visuelle.
- La gestion des espaces privés, aussi appelés *apartés*. Un expert distant peut en effet vouloir consulter certains objets virtuels ou l’anatomie du patient indépendamment de l’étape chirurgicale en cours. Or, cela peut poser problème si le point de vue des utilisateurs distant est lié à celui du chirurgien. Se posent alors la question du contrôle du point de vue et de l’existence de plusieurs environnements.

De plus amples recherches sont par conséquent nécessaires dans le cadre d'une opération chirurgicale augmentée menée en coopération par plusieurs experts.

Annexe A

Questionnaire Raw-TLX

Questionnaire Raw-TLX utilisé pendant les études 1 à 5 (en anglais, traduit ici en français).

Demande mentale

Quelle activité mentale et perceptive était nécessaire ? La tâche était-elle facile ou exigeante, simple ou complexe ?

Très Très
faible élevée

Demande physique

Quelle quantité d'activité physique était nécessaire ? La tâche a-t-elle été facile ou exigeante, lâche ou ardue ?

Très Très
faible élevée

Demande temporelle

Quelle pression temporelle avez-vous ressentie en raison du rythme auquel les tâches ou les éléments de tâche se sont déroulés ? Le rythme était-il lent ou rapide ?

Très Très
faible élevée

Performance

Dans quelle mesure avez-vous réussi à exécuter la tâche ? Dans quelle mesure avez-vous été satisfait de votre performance ?

Réussite Échec
parfaite

Effort

À quel point avez-vous dû travailler (mentalement et physiquement) pour atteindre votre niveau de performance ?

Très Très
faible élevée

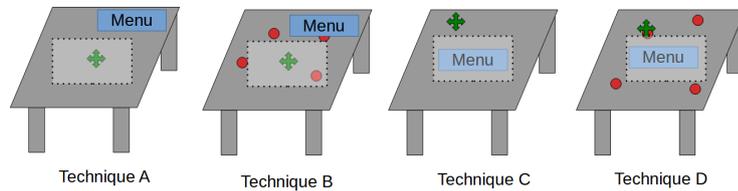
Niveau de frustration

Dans quelle mesure vous êtes-vous senti irrité, stressé et ennuyé pendant la tâche ?

Très Très
faible élevée

Annexe B

Questionnaire final études 1 et 2



Classez de 1 (meilleur) à 4 (pire) les techniques A, B, C et D selon les questions suivantes :

- Technique la plus facile à utiliser
- Technique la plus rapide
- Technique la plus précise
- Technique la moins fatigante
- Technique globalement préférée

	1	2	3	4
Technique A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*5

Quel est votre niveau de fatigue concernant :

- Les yeux
- La nuque
- Les bras et épaules

Pas du tout fatigué	<input type="radio"/>	Très fatigué						
---------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	--------------

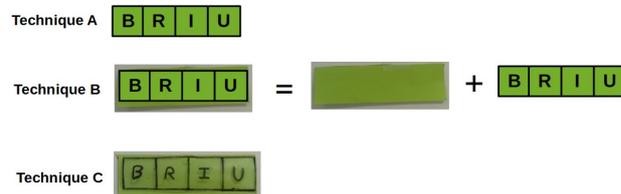
*3

Qu'avez-vous pensé des ancrés ? Pourquoi ?

Un commentaire à ajouter ?

Annexe C

Questionnaire final étude 3



Classez de 1 (meilleur) à 4 (pire) les techniques A, B, C selon les questions suivantes :

- Technique la plus facile à utiliser
- Technique la plus rapide
- Technique la plus précise
- Technique la moins fatigante
- Technique globalement préférée

	1	2	3	4
Technique A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*5

Quelle(s) technique(s) avez-vous préféré et pourquoi ?

Les ancres physiques (Technique B) étaient-elles utiles lors de la tâche ?
Pourquoi/pourquoi pas ?

Pas du tout Tout à fait

Les menus physiques (Technique C) étaient-ils utiles lors de la tâche ?
Pourquoi/pourquoi pas ?

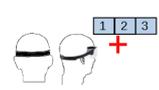
Pas du tout Tout à fait

Avec quelle technique pensez-vous avoir le mieux retenu la position des cibles ? Pour quelle(s) raison(s) ?

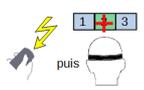
Un commentaire à ajouter ?

Annexe D

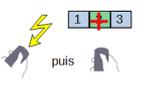
Questionnaire final étude 4



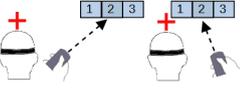
Technique A
Tourner la tête +
 curseur avec la tête



Technique B
Téléporter menu +
 curseur avec la tête



Technique C
Téléporter menu +
 curseur avec la main



Technique D
Attraper le menu et le
 ramener

Classez de 1 (meilleur) à 4 (pire) les techniques A, B, C et D selon les questions suivantes :

- Technique la plus facile à utiliser
- Technique la plus rapide
- Technique la plus précise
- Technique la moins fatigante
- Technique globalement préférée

	1	2	3	4
Technique A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Technique D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*5

Quelle(s) technique(s) avez-vous préféré et pourquoi ?

Ramener le menu (techniques B, C et D) était-il utile par rapport à tourner la tête (technique A) ? Pourquoi/pourquoi pas ?

Pas du tout Tout à fait

Qu'avez-vous pensé de la téléportation instantanée du menu (techniques B et C) ?

Qu'avez-vous pensé du geste continu avec la technique D ?

Un commentaire à ajouter ?

Annexe E

Questionnaire étude 5

Session 1 (Entraînement)

Questionnaire Raw-TLX à remplir en premier.

Comment avez-vous procédé pour explorer l'objet ? Quelle était votre stratégie ?

Quelle approche avez-vous préféré entre Split et Synchro ? Pour quelle(s) raison(s) ?

Si vous pouviez combiner les deux approches (Split et Synchro), comment mèneriez-vous à bien la tâche ?

Session 2 (Mode libre)

Questionnaire Raw-TLX à remplir.

Comment avez-vous procédé pour explorer l'objet ? Quelle était votre stratégie ?

Avez-vous utilisé le mode Synchro ? Pour quelle(s) raison(s) ?

Avez-vous utilisé le mode Split? Pour quelle(s) raison(s) ?

Avez-vous utilisé la Boîte ou le Patron comme référence/objet principal ? Pourquoi ?

Un commentaire à ajouter ?

Bibliographie

- [1] Parastoo ABTAHI et al. “Beyond The Force : Using Quadcopters to Appropriate Objects and the Environment for Haptics in Virtual Reality”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk : Association for Computing Machinery, 2019, p. 1-13.
- [2] Johnny ACCOT et Shumin ZHAI. “Beyond Fitts’ law : models for trajectory-based HCI tasks”. In : *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 1997, p. 295-302.
- [3] AESCULAP. *OrthoPilot®navigation system*. 2020. URL : <https://www.bbBraun.com/en/products-and-therapies/orthopaedic-joint-replacement/orthopilot.html/orthopilot.html>.
- [4] Rajesh AGGARWAL et al. “An evidence-based virtual reality training program for novice laparoscopic surgeons”. In : *Annals of surgery* 244.2 (2006), p. 310.
- [5] David AHLSTRÖM, Rainer ALEXANDROWICZ et Martin HITZ. “Improving menu interaction : A comparison of standard, force enhanced and Jumping Menus”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 2 (2006), p. 1067-1076.
- [6] Daniel ANDERSEN et al. “Virtual annotations of the surgical field through an augmented reality transparent display”. In : *Vis. Comput.* 32.11 (2016), p. 1481-1498.
- [7] Ferran ARGELAGUET et Carlos ANDUJAR. “A survey of 3D object selection techniques for virtual environments”. In : *Comput. Graph.* 37.3 (2013), p. 121-136.
- [8] Sonal ARORA et al. “The impact of stress on surgical performance : a systematic review of the literature”. In : *Surgery* 147.3 (2010), p. 318-330.
- [9] Takeshi ASANO et al. “Predictive interaction using the Delphian Desktop”. In : *UIST Proc. Annu. ACM Symp. User Interface Software Technol.* (2005), p. 133-141.
- [10] Rowel ATIENZA et al. “Interaction techniques using head gaze for virtual reality”. In : *2016 IEEE Region 10 Symposium (TENSYMP)*. IEEE. 2016, p. 110-114.
- [11] Carmen E AU et James J CLARK. “Integrating multiple views with virtual mirrors to facilitate scene understanding”. In : *ACM Transactions on Applied Perception (TAP)* 8.4 (2008), p. 1-14.

- [12] Takumi AZAI et al. “Selection and Manipulation Methods for a Menu Widget on the Human Forearm”. In : *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. Denver, Colorado, USA : Association for Computing Machinery, 2017, p. 357-360.
- [13] Ronald T AZUMA. “A survey of augmented reality”. In : *Presence : Teleoperators & Virtual Environments* 6.4 (1997), p. 355-385.
- [14] Giovanni BADIALI et al. “Augmented reality as an aid in maxillofacial surgery : Validation of a wearable system allowing maxillary repositioning”. In : *J. Cranio-Maxillofacial Surg.* 42.8 (2014), p. 1970-1976.
- [15] Charles BAILLY, François LEITNER et Laurence NIGAY. “Head-Controlled Menu in Mixed Reality with a HMD”. In : *Human-Computer Interaction – INTERACT 2019*. Springer International Publishing, 2019, p. 395-415.
- [16] Charles BAILLY, François LEITNER et Laurence NIGAY. “Bring2Me : Bringing Virtual Widgets Back to the User’s Field ofView in Mixed Reality”. In : *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '20. Salerno, Italy : Association for Computing Machinery, 2020.
- [17] Gilles BAILLY, Eric LECOLINET et Laurence NIGAY. “Flower menus”. In : *Proc. Work. Conf. Adv. Vis. interfaces - AVI '08* (2008), p. 15.
- [18] Michelle Q.Wang BALDONADO, Allison WOODRUFF et Allan KUCHINSKY. “Guidelines for using multiple views in information visualization”. In : *Proc. Work. Adv. Vis. Interfaces* (2000), p. 110-119.
- [19] Marc BALOUP, Thomas PIETRZAK et Géry CASIEZ. “Raycursor : A 3D Pointing Facilitation Technique based on Raycasting”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* (2019).
- [20] Catherine B BARDEN et al. “Effects of limited work hours on surgical training”. In : *Journal of the American College of Surgeons* 195.4 (2002), p. 531-538.
- [21] Andrea BATCH et al. “There is no spoon : Evaluating performance, space use, and presence with expert domain users in immersive analytics”. In : *IEEE transactions on visualization and computer graphics* 26.1 (2019), p. 536-546.
- [22] Michel BEAUDOUIN-LAFON. “Instrumental interaction : An Interaction Model for Designing Post-WIMP User Interfaces”. In : *Proc. 18th Int. Conf. Hum. factors Comput. Syst. - CHI '00* 2.1 (2000), p. 446-453.
- [23] AL BENABID et al. “Computer-driven robot for stereotactic surgery connected to CT scan and magnetic resonance imaging”. In : *Stereotactic and Functional Neurosurgery* 50.1-6 (1987), p. 153-154.
- [24] Steve BENFORD et al. “Shared Spaces : Transportation, Artificiality, and Spatiality”. In : *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*. CSCW '96. Boston, Massachusetts, USA : Association for Computing Machinery, 1996, p. 77-86.
- [25] Hrvoje BENKO, Ricardo JOTA et Andrew WILSON. “MirageTable : freehand interaction on a projected augmented reality tabletop”. In : *Proceedings of the SIGCHI conference on human factors in computing systems*. New York, NY, USA : Association for Computing Machinery, 2012, p. 199-208.

- [26] Francois BERARD et Thibault LOUIS. “The object inside : Assessing 3D examination with a spherical handheld perspective-corrected display”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 2017-May (2017), p. 4396-4404.
- [27] Sylvain BERNHARDT et al. “The status of augmented reality in laparoscopic surgery as of 2016”. In : *Med. Image Anal.* 37.March 2018 (2017), p. 66-90.
- [28] Eric A. BIER et al. “Toolglass and Magic Lenses : The See-through Interface”. In : *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '93. Anaheim, CA : Association for Computing Machinery, 1993, p. 73-80.
- [29] Mark BILLINGHURST, Raphael GRASSET et Julian LOOSER. “Designing augmented reality interfaces”. In : *ACM Siggraph Computer Graphics* 39.1 (2005), p. 17-22.
- [30] Mark BILLINGHURST, Hirokazu KATO et Ivan POUPYREV. “The MagicBook : A transitional AR interface”. In : *Comput. Graph.* 25.5 (2001), p. 745-753.
- [31] Mark BILLINGHURST, Hirokazu KATO, Ivan POUPYREV et al. “Tangible augmented reality”. In : *ACM SIGGRAPH ASIA* 7.2 (2008), p. 1-10.
- [32] Richard A BOLT. ““Put-that-there” : Voice and Gesture at the Graphics Interface”. In : *Proc. 7th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '80* (1980), p. 262-270.
- [33] H. BORGMANN et al. “Feasibility and safety of augmented reality-assisted urological surgery using smartglass”. In : *World J. Urol.* 35.6 (2017), p. 967-972.
- [34] Gerd BRUDER et al. “Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras”. In : *2009 International Conference on CyberWorlds*. IEEE. 2009, p. 43-50.
- [35] Volkert BUCHMANN et al. “FingARTips”. In : *Proc. 2nd Int. Conf. Comput. Graph. Interact. Tech. Australasia Southe East Asia - Graph. '04* 1.212 (2004), p. 212. URL : <http://portal.acm.org/citation.cfm?doid=988834.988871>.
- [36] Xiang CAO, J. LI, JACKY et Ravin BALAKRISHNAN. “Peephole pointing : modeling acquisition of dynamically revealed targets”. In : *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '08*. c. New York, NY, USA : Association for Computing Machinery, 2008, p. 1699-1708.
- [37] Joan CASSEL. “On Control, certitude, and the paranoia of surgeons”. In : *Cult. Med. Psychiatry* 53.9 (1987), p. 1689-1699.
- [38] Yeonjoo CHA et Rohae MYUNG. “Extended Fitts’ law for 3D pointing tasks using 3D target arrangements”. In : *International Journal of Industrial Ergonomics* 43.4 (2013), p. 350-355.
- [39] Isha CHATURVEDI et al. “Peripheral vision : A new killer app for smart glasses”. In : *Int. Conf. Intell. User Interfaces, Proc. IUI Part F1476* (2019), p. 625-636.
- [40] Siyuan CHEN, Julien EPPS et Fang CHEN. “A comparison of four methods for cognitive load measurement”. In : *Proc. 23rd Aust. Comput. Interact. Conf. - OzCHI '11*. 2011.

- [41] Kenny Tsu Wei CHOO, Rajesh Krishna BALAN et Youngki LEE. “Examining Augmented Virtuality Impairment Simulation for Mobile App Accessibility Design”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk : Association for Computing Machinery, 2019, p. 1-11.
- [42] A. COCKBURN et al. “Air pointing : Design and evaluation of spatial target acquisition with and without visual feedback”. In : *Int. J. Hum. Comput. Stud.* 69.6 (juin 2011), p. 401-414.
- [43] Andy COCKBURN. “Anchoring Effects and Troublesome Asymmetric Transfer in Subjective Ratings”. In : *CHI 2019* (2019), p. 1-12.
- [44] Andy COCKBURN, Amy KARLSON et Benjamin B. BEDERSON. “A review of overview+detail, zooming, and focus+context interfaces”. In : *ACM Comput. Surv.* 41.1 (2008), p. 1-31.
- [45] Philip R COHEN et Sharon L OVIATT. “The role of voice input for human-machine communication”. In : *proceedings of the National Academy of Sciences* 92.22 (1995), p. 9921-9927.
- [46] Andrea COLAÇO et al. “Mime”. In : *Proc. 26th Annu. ACM Symp. User interface Softw. Technol. - UIST '13* (2013), p. 227-236. URL : <http://dl.acm.org/citation.cfm?doid=2501988.2502042>.
- [47] Sara CONDINO et al. “Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks”. In : *IEEE Trans. Biomed. Eng.* (2019).
- [48] Maxime CORDEIL et al. “Iatk : An immersive analytics toolkit”. In : *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, p. 200-209.
- [49] Jennifer CRAWFORD et Andrew NEAL. “A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation”. In : *The International Journal of Aviation Psychology* 16.1 (2006), p. 1-19.
- [50] Lawrence D LD CUTLER et al. “Two-handed direct manipulation on the responsive workbench”. In : *Proc. 1997 Symp. Interact. 3D Graph.* Providence (1997), p. 107-114.
- [51] James E CUTTING. “How the eye measures reality and virtual reality”. In : *Virtual Real.* 29.1 (1997), p. 27-36.
- [52] Dragos DATCU et Stephan LUKOSCH. “Free-hands interaction in augmented reality”. In : *Proc. 1st Symp. Spat. user Interact. - SUI '13* (2013), p. 33.
- [53] Vlado DELIĆ et al. “Speech technology progress based on new machine learning paradigm”. In : *Computational intelligence and neuroscience 2019* (2019).
- [54] Usine DIGITALE. *Covid-19 : Pour réduire le temps d'exposition au virus, des médecins s'équipent de casques HoloLens 2*. 2020. URL : <https://www.usine-digitale.fr/article/covid-19-pour-reduire-le-temps-d-exposition-au-virus-des-medecins-s-equipent-de-casques-hololens-2.N967366>.

- [55] Usine DIGITALE. *En Grande-Bretagne, la NHS se réjouit des bénéfices de la formation des soignants en réalité virtuelle*. 2020. URL : <https://www.usine-digitale.fr/article/en-grande-bretagne-la-nhs-se-rejouit-des-benefices-de-la-formation-des-soignants-en-realite-virtuelle>. N976701.
- [56] Usine DIGITALE. *Ikea lance Place, sa nouvelle application en réalité augmentée*. 2020. URL : <https://www.usine-digitale.fr/article/ikea-lance-place-sa-nouvelle-application-en-realite-augmentee>. N586588.
- [57] Usine DIGITALE. *L'Oréal acquiert ModiFace, le leader mondial de l'essayage virtuel*. 2020. URL : <https://www.usine-digitale.fr/article/l-oreal-acquiert-modiface-le-leader-mondial-du-relooking-virtuel>. N667534.
- [58] Usine DIGITALE. *La future Classe S de Mercedes proposera l'affichage tête haute en réalité augmentée*. 2020. URL : <https://www.usine-digitale.fr/article/la-future-classe-s-de-mercedes-proposera-l-affichage-tete-haute-en-realite-augmentee>. N985009.
- [59] Huong Q DINH et al. "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments". In : *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE. 1999, p. 222-228.
- [60] Benjamin J. DIXON et al. "Surgeons blinded by enhanced navigation : The effect of augmented reality on attention". In : *Surg. Endosc. Other Interv. Tech.* 27.2 (2013), p. 454-461.
- [61] Pierre DRAGICEVIC et al. *Fair Statistical Communication in HCI Pierre Dragicevic To cite this version : Fair Statistical Communication in HCI*. 2016, p. 291-330.
- [62] Emmanuel DUBOIS et al. "Classification Space for Augmented Surgery, an Augmented Reality Case Study." In : *Interact.* 1999, p. 353-359.
- [63] Andrew T DUCHOWSKI et al. "The index of pupillary activity : measuring cognitive load vis-à-vis task difficulty with pupil oscillation". In : *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. New York, NY, USA, 2018, p. 282.
- [64] Pieter DUYSBURGH, Shirley A. ELPRAMA et An JACOBS. "Exploring the Social-Technological Gap in Telesurgery : Collaboration within Distributed or Teams". In : *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*. CSCW'14. Baltimore, Maryland, USA : Association for Computing Machinery, 2014, p. 1537-1548.
- [65] Lars C. EBERT et al. "You can't touch this : Touch-free navigation through radiological images". In : *Surg. Innov.* 19.3 (2012), p. 301-307.
- [66] Barrett ENS, David AHLSTRÖM et Pourang IRANI. "Moving Ahead with Peephole Pointing : Modelling Object Selection with Head-Worn Display Field of View Limitations". In : *SUI '16*. Tokyo, Japan : Association for Computing Machinery, 2016.
- [67] Barrett ENS et al. "Characterizing User Performance with Assisted Direct Off-screen Pointing". In : *Proc. 13th Int. Conf. Hum. Comput. Interact. with Mob. Devices Serv. - MobileHCI '11* (2011), p. 485-494.

- [68] EPSON. *Le casque de RA Epson Moverio BT-300*. 2016. URL : <https://www.epson.fr/products/see-through-mobile-viewer/moverio-bt-300>.
- [69] Andéol EVAÏN et al. “Can i think of something else when using a bci? : Cognitive demand of an ssvp-based bci”. In : *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA : ACM, 2017, p. 5120-5125.
- [70] FACEBOOK. *Le casque de VR Oculus Rift*. 2016. URL : https://www.oculus.com/rift/?locale=fr_FR#oui-csl-rift-games=star-trek.
- [71] Pascal FALLAVOLLITA et al. “An augmented reality C-arm for intraoperative assessment of the mechanical axis : a preclinical study”. In : *Int. J. Comput. Assist. Radiol. Surg.* 11.11 (2016), p. 2111-2117.
- [72] Paul M FITTS. “The information capacity of the human motor system in controlling the amplitude of movement.” In : *Journal of experimental psychology* 47.6 (1954), p. 381.
- [73] George W. FITZMAURICE et William BUXTON. “Empirical evaluation of graspable user interfaces : Towards specialized, space-multiplexed input”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 1 (1997), p. 43-50.
- [74] David C. FOYLE, Anthony D. ANDRE et Becky L. HOOEY. “Situation Awareness in an Augmented Reality Cockpit : Design, Viewpoints and Cognitive Glue”. In : *Proc. 11th Int. Conf. Hum. Comput. Interact.* February (2005), p. 10.
- [75] Bruno FRUCHARD, Eric LECOLINET et Olivier CHAPUIS. “Impact of semantic aids on command memorization for on-body interaction and directional gestures”. In : *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. New York, NY, USA : Association for Computing Machinery, 2018, p. 1-9.
- [76] Danilo GASQUES RODRIGUES et al. “Exploring Mixed Reality in Specialized Surgical Environments”. In : *Proc. 2017 CHI Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '17* (2017), p. 2591-2598.
- [77] Atul A GAWANDE et al. “Analysis of errors reported by surgeons at three teaching hospitals”. In : *Surgery* 133.6 (2003), p. 614-621.
- [78] Raphael GRASSET, Julian LOOSER et Mark BILLINGHURST. “Transitional Interface : Concept, Issues and Framework”. In : *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*. ISMAR '06. USA : IEEE Computer Society, 2006, p. 231-232.
- [79] Anthony G GREENWALD. “Within-subjects designs : To use or not to use?” In : *Psychological Bulletin* 83.2 (1976), p. 314.
- [80] Tovi GROSSMAN et Ravin BALAKRISHNAN. “Pointing at Trivariate Targets in 3D Environments”. In : *Proc. 2004 Conf. Hum. factors Comput. Syst. - CHI '04* 6.1 (2004), p. 447-454.
- [81] Tovi GROSSMAN et Ravin BALAKRISHNAN. “A probabilistic approach to modeling two-dimensional pointing”. In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 12.3 (2005), p. 435-459.

- [82] Maxime GUILLON, François LEITNER et Laurence NIGAY. “Static Voronoi-Based Target Expansion Technique for Distant Pointing”. In : *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. AVI '14. Como, Italy : Association for Computing Machinery, 2014, p. 41-48.
- [83] Maxime GUILLON, François LEITNER et Laurence NIGAY. “Target Expansion Lens : It is Not the More Visual Feedback the Better !” In : *Proc. Int. Work. Conf. Adv. Vis. Interfaces - AVI '16* (2016), p. 52-59.
- [84] Eija HAAPALAINEN et al. *Psycho-Physiological Measures for Assessing Cognitive Load*. 2010.
- [85] Sandra G. HART et Lowell E. STAVELAND. “Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research”. In : *Adv. Psychol.* 52.C (1988), p. 139-183.
- [86] Benjamin HATSCHER, Maria LUZ et Christian HANSEN. “Foot interaction concepts to support radiological interventions”. In : *i-com* 17.1 (2018), p. 3-13.
- [87] Zhenyi HE et Xubo YANG. “Hand-based interaction for object manipulation with augmented reality glasses”. In : *Proc. 13th ACM SIGGRAPH Int. Conf. Virtual-Reality Contin. its Appl. Ind. - VRCAI '14* (2014), p. 227-230. URL : <http://dl.acm.org/citation.cfm?doid=2670473.2670505>.
- [88] Florian HEINRICH et al. “Depth perception in projective augmented reality : An evaluation of advanced visualization techniques”. In : *Proc. ACM Symp. Virtual Real. Softw. Technol. VRST* (2019).
- [89] Steven J HENDERSON et Steven FEINER. “Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret”. In : *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2009, p. 135-144.
- [90] Niels HENZE et Susanne BOLL. “Evaluation of an off-screen visualization for magic lens and dynamic peephole interfaces”. In : *Proc. 12th Int. Conf. Hum. Comput. Interact. with Mob. devices Serv. - MobileHCI '10* (2010), p. 191.
- [91] Hans-Günter HIRSCH et David PEARCE. “The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions”. In : *ASR2000-Automatic speech recognition : challenges for the new Millenium ISCA tutorial and research workshop (ITRW)*. 2000.
- [92] David M. HOFFMAN et al. “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue”. In : *Journal of Vision* 8.3 (2008), p. 33.
- [93] Antje HÜBLER et al. “Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining.” In : *CURAC*. 2014, p. 165-168.
- [94] Archie HUGHES-HALLETT et al. “Inattention blindness in surgery”. In : *Surgical endoscopy* 29.11 (2015), p. 3184-3189.
- [95] Louise HULL et al. “The impact of nontechnical skills on technical performance in surgery : a systematic review”. In : *Journal of the American College of Surgeons* 214.2 (2012), p. 214-230.

- [96] Aulikki HYRSKYKARI, Howell ISTANCE et Stephen VICKERS. "Gaze Gestures or Dwell-Based Interaction ?" In : *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California : Association for Computing Machinery, 2012, p. 229-232.
- [97] Shamsi T. IQBAL, Xianjun Sam ZHENG et Brian P. BAILEY. "Task-evoked pupillary response to mental workload in human-computer interaction". In : *CHI 2004*. 2004, p. 1477.
- [98] Sylvia IRAWATI et al. "'Move the Couch Where?' : Developing an Augmented Reality Multimodal Interface". In : *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*. ISMAR '06. USA : IEEE Computer Society, 2006, p. 183-186.
- [99] Hiroshi ISHII. "The tangible user interface and its evolution". In : *Communications of the ACM* 51.6 (2008), p. 32-36.
- [100] *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9*. Standard. International Organization for Standardization, 2000.
- [101] Sakeba N ISSA et Leena SHARMA. "Epidemiology of osteoarthritis : an update". In : *Current rheumatology reports* 8.1 (2006), p. 7-15.
- [102] Robert J. K. JACOB. "What you look at is what you get : eye movement-based interaction techniques". In : *Proc. SIGCHI Conf. Hum. factors Comput. Syst. Empower. people - CHI '90* (1990), p. 11-18.
- [103] Richard J. JAGACINSKI et Donald L. MONK. "Fitts' law in two dimensions with hand and head movements movements". In : *J. Mot. Behav.* 17.1 (1985), p. 77-95.
- [104] Shahram JALALINIYA et al. "Head and eye movement as pointing modalities for eyewear computers". In : *Proc. - 11th Int. Conf. Wearable Implant. Body Sens. Networks Work. BSN Work. 2014* (2014), p. 50-53.
- [105] Yvonne JANSEN, Pierre DRAGICEVIC et Jean-Daniel FEKETE. "Evaluating the efficiency of physical visualizations". In : (2013), p. 2593.
- [106] J. Adam JONES et al. "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception". In : *APGV 2008 - Proc. Symp. Appl. Percept. Graph. Vis.* 1.212 (2008), p. 9-14.
- [107] Ed KAISER et al. "Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality". In : *Proceedings of the 5th international conference on Multimodal interfaces*. 2003, p. 12-19.
- [108] Darko KATIĆ et al. "Context-aware Augmented Reality in laparoscopic surgery". In : *Comput. Med. Imaging Graph.* 37.2 (2013), p. 174-182.
- [109] Darko KATIĆ et al. "A system for context-aware intraoperative augmented reality in dental implant surgery". In : *Int. J. Comput. Assist. Radiol. Surg.* 10.1 (2015), p. 101-108.
- [110] Bonifaz KAUFMANN et David AHLSTRÖM. "Revisiting Peephole Pointing : A Study of Target Acquisition with a Handheld Projector". In : *MobileHCI'12* (2012), p. 211-220.

- [111] Norihiko KAWAI, Tomokazu SATO et Naokazu YOKOYA. “Diminished reality based on image inpainting considering background geometry”. In : *IEEE transactions on visualization and computer graphics* 22.3 (2015), p. 1236-1247.
- [112] Adam KENDON. *Gesture : Visible action as utterance*. Cambridge University Press, 2004.
- [113] Hyunyoung KIM, Céline COUTRIX et Anne ROUDAUT. “Knobslder : Design of a shape-changing UI for parameter control”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 2018-April (2018).
- [114] Yea-Seul KIM et al. “Vocal Shortcuts for Creative Experts”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk : Association for Computing Machinery, 2019, p. 1-14.
- [115] Gregory KRAMIDA. “Resolving the vergence-accommodation conflict in head-mounted displays”. In : *IEEE transactions on visualization and computer graphics* 22.7 (2015), p. 1912-1931.
- [116] Ernst KRUIJFF, J Edward SWAN et Steven FEINER. “Perceptual issues in augmented reality revisited”. In : *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2010, p. 3-12.
- [117] Andre KUNERT et al. “Multi-window 3D interaction for collaborative virtual reality”. In : *IEEE transactions on visualization and computer graphics* (2019).
- [118] Mikko KYTÖ et al. “Pinpointing : Precise Head - and Eye - Based Target Selection for Augmented Reality”. In : *Chi18* April (2018), p. 1-14.
- [119] Joseph J LAVIOLA JR et al. *3D user interfaces : theory and practice*. Addison-Wesley Professional, 2017.
- [120] Claude LE PEN, Camille REYGROBELLET et Isabelle GÉRENTES. “Les conséquences socioéconomiques de l’arthrose en France. Etude COART France”. In : *Revue du rhumatisme* 72.12 (2005), p. 1326-1330.
- [121] Hyeongmook LEE, Mark BILLINGHURST et Woontack WOO. “Two-handed tangible interaction techniques for composing augmented blocks”. In : *Virtual Reality* 15.2-3 (2011), p. 133-146.
- [122] Hyeongmook LEE et Woontack WOO. “Tangible spin cube for 3D ring menu in real space”. In : *Proc. 28th Int. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA '10*. 2010, p. 4147.
- [123] Jae Yeol LEE, Gue Won RHEE et Dong Woo SEO. “Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment”. In : *Int. J. Adv. Manuf. Technol.* 51.9-12 (2010), p. 1069-1082.
- [124] SangYoon LEE et al. “Evaluation of pointing techniques for ray casting selection in virtual environments”. In : (2003), p. 38-44.
- [125] Frank Chun Yat LI, David DEARMAN et Khai N. TRUONG. “Virtual shelves : Interactions with orientation aware devices”. In : *UIST 2009 - Proc. 22nd Annu. ACM Symp. User Interface Softw. Technol.* (2009), p. 125-128.

- [126] Mei Li LIN, Robert G. RADWIN et Gregg C. VANDERHEIDEN. “Gain effects on performance using a head-controlled computer input device”. In : *Ergonomics* 35.2 (1992), p. 159-175.
- [127] David LINDLBAUER et Andy D. WILSON. “Remixed Reality : Manipulating Space and Time in Augmented Reality”. In : *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada : Association for Computing Machinery, 2018, p. 1-13.
- [128] Pedro LOPES et al. “Adding Force Feedback to Mixed Reality Experiences and Games Using Electrical Muscle Stimulation”. In : *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada : Association for Computing Machinery, 2018, p. 1-13.
- [129] T LOUIS et F BERARD. “Superiority of a handheld perspective-coupled display in isomorphic docking performances”. In : *Proc. 2017 ACM Int. Conf. Interact. Surfaces Spaces, ISS 2017* (2017), p. 72-81.
- [130] Irene MA et al. “Leveraging augmented reality training tool for medical education : A case study in central venous catheterization”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 2018-April (2018), p. 1-7.
- [131] Anna MACARANAS, Alissa N. ANTLE et Bernhard E. RIECKE. “Bridging the Gap : Attribute and Spatial Metaphors for Tangible Interface Design”. In : *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. TEI '12. Kingston, Ontario, Canada : Association for Computing Machinery, 2012, p. 161-168.
- [132] I Scott MACKENZIE. *Human-computer interaction : An empirical research perspective*. Newnes, 2012.
- [133] I. Scott MACKENZIE. “Fitts’ Law as a Research and Design Tool in Human-Computer Interaction”. In : *Hum.-Comput. Interact.* 7.1 (1992), p. 91-139.
- [134] Nader MAHMOUD et al. “On-patient see-through augmented reality based on visual SLAM”. In : *Int. J. Comput. Assist. Radiol. Surg.* 12.1 (2017), p. 1-11.
- [135] Chia-Chi MAO, Chung-Chong SUN et Chien-Hsu CHEN. “Evaluate Learner’s Acceptance of Augmented Reality Based Military Decision Making Process Training System”. In : *Proceedings of the 5th International Conference on Information and Education Technology*. ICIET '17. Tokyo, Japan : Association for Computing Machinery, 2017, p. 73-77.
- [136] Diako MARDANBEGI, Dan Witzner HANSEN et Thomas PEDERSON. “Eye-based head gestures movements”. In : *Etra '12* 1.212 (2012), p. 139-146.
- [137] Kim MARRIOTT et al. *Immersive Analytics*. T. 11190. Springer, 2018.
- [138] Marco MASON. “The MIT Museum Glassware Prototype : Visitor Experience Exploration for Designing Smart Glasses”. In : *J. Comput. Cult. Herit.* 9.3 (2016), p. 1-28.
- [139] Larousse MÉDICAL. *Définition de la chirurgie orthopédique*. 2020. URL : https://www.larousse.fr/encyclopedie/medical/chirurgie_orthop%C3%A9dique/11905.

- [140] Helena M. MENTIS, Ahmed RAHIM et Pierre THEODORE. “Crafting the Image in Surgical Telemedicine”. In : *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing*. CSCW’16. San Francisco, California, USA : Association for Computing Machinery, 2016, p. 744-755.
- [141] Helena M. MENTIS et al. “Voice or Gesture in the Operating Room”. In : *Proc. 33rd Annu. ACM Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA ’15* (2015), p. 773-780.
- [142] David E. MEYER et al. “Optimality in Human Motor Performance : Ideal Control of Rapid Aimed Movements”. In : *Psychol. Rev.* 95.3 (1988), p. 340-370.
- [143] MICROSOFT. *Première utilisation de l’Hololens pour une opération chirurgicale à Paris*. 2017. URL : <https://experiences.microsoft.fr/business/realite-mixte-business/chirurgie-assistee-avec-hololens/>.
- [144] MICROSOFT. *Confort d’utilisation pour les appareils holographiques*. 2020. URL : <https://docs.microsoft.com/fr-fr/windows/mixed-reality/comfort>.
- [145] MICROSOFT. *Hololens 2 HMD by Microsoft*. 2020. URL : https://www.microsoft.com/en-us/hololens?icid=SSM_AS_Promo_Devices_HoloLens2.
- [146] Paul MILGRAM et Fumio KISHINO. “A taxonomy of mixed reality visual displays”. In : *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), p. 1321-1329.
- [147] Stefan MITRASINOVIC et al. “Clinical and surgical applications of smart glasses”. In : *Technol. Heal. Care* 23.4 (2015), p. 381-401.
- [148] Shohei MORI, Sei IKEDA et Hideo SAITO. “A survey of diminished reality : Techniques for visually concealing, eliminating, and seeing through real objects”. In : *IPSSJ Transactions on Computer Vision and Applications* 9.1 (2017), p. 1-14.
- [149] Thomas MUENDER et al. “Does It Feel Real? Using Tangibles with Different Fidelities to Build and Explore Scenes in Virtual Reality”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA : Association for Computing Machinery, 2019, p. 1-12.
- [150] Martin NAEF, Oliver STAADT et Markus GROSS. “Spatialized Audio Rendering for Immersive Virtual Environments”. In : *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST ’02. Hong Kong, China : Association for Computing Machinery, 2002, p. 65-72.
- [151] Mathieu NANCEL et al. “Mid-air Pointing on Ultra-Walls”. In : *ACM Trans. Comput. Interact. Artic.* 22.21 (2015).
- [152] Adam NOWAK et al. “What Does the Oscilloscope Say? : Comparing the Efficiency of In-Situ Visualisations during Circuit Analysis”. In : *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA ’20. Honolulu, HI, USA : Association for Computing Machinery, 2020, p. 1-7.

- [153] Kenton O'HARA et al. "Interactional order and constructed ways of seeing with touchless imaging systems in surgery". In : *Comput. Support. Coop. Work* 23.3 (2014), p. 299-337.
- [154] S. K. ONG, M. L. YUAN et A. Y.C. NEE. "Augmented reality applications in manufacturing : A survey". In : *Int. J. Prod. Res.* 46.10 (2008), p. 2707-2742.
- [155] Kaslm ÖZACAR et al. "3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays". In : *Interact. Comput.* 29.4 (2017), p. 579-591.
- [156] Toni PAKKANEN et Roope RAISAMO. "Appropriateness of Foot Interaction for Non-Accurate Spatial Tasks". In : *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '04. Vienna, Austria : Association for Computing Machinery, 2004, p. 1123-1126.
- [157] Assistance Publique - Hôpitaux de PARIS. *AP-HP : 1ère intervention chirurgicale au monde réalisée avec une plateforme collaborative de réalité mixte, en interaction avec des équipes médicales internationales, à l'hôpital Avicenne*. 2017. URL : <https://www.aphp.fr/contenu/ap-hp-1ere-intervention-chirurgicale-au-monde-realisee-avec-une-plateforme-collaborative-de>.
- [158] Le PARISIEN. *Quand le chirurgien place une prothèse d'épaule grâce... à des hologrammes*. 2017. URL : <https://www.leparisien.fr/societe/quand-le-chirurgien-place-une-prothese-d-epaule-grace-a-des-hologrammes-05-12-2017-7435155.php>.
- [159] Robert PATTERSON, Marc D. WINTERBOTTOM et Byron J. PIERCE. "Perceptual issues in the use of head-mounted visual displays". In : *Hum. Factors* 48.3 (2006), p. 555-573.
- [160] Sylvain PAUCHET et al. "GazeForm : Dynamic gaze-adaptive touch surface for eyes-free interaction in airliner cockpits". In : *DIS 2018 - Proc. 2018 Des. Interact. Syst. Conf.* (2018), p. 1193-1206.
- [161] Sebastien PELURSON et Laurence NIGAY. "Multimodal interaction with a bifocal view on mobile devices". In : *ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact.* (2015), p. 191-198.
- [162] Patrick PEREA, Denis MORAND et Laurence NIGAY. "Spotlight on Off-Screen Points of Interest in Handheld Augmented Reality : Halo-Based Techniques". In : *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ISS '19. Daejeon, Republic of Korea : Association for Computing Machinery, 2019, p. 43-54.
- [163] Patrick PEREA, Denis MORAND et Laurence NIGAY. "Target Expansion in Context : The Case of Menu in Handheld Augmented Reality". In : *Proceedings of the International Conference on Advanced Visual Interfaces*. AVI '20. Salerno, Italy : Association for Computing Machinery, 2020.
- [164] Simon T. PERRAULT et al. "Physical Loci : Leveraging Spatial, Object and Semantic Memory for Command Selection". In : *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea : Association for Computing Machinery, 2015, p. 299-308.

- [165] Jeffrey S. PIERCE, Brian C. STEARNS et Randy PAUSCH. “Voodoo Dolls : Seamless interaction at multiple scales in virtual environments”. In : *Proc. Symp. Interact. 3D Graph.* (1999), p. 141-145.
- [166] I POUPYREV et al. “Egocentric Object Manipulation in Virtual Environments : Empirical Evaluation of Interaction Techniques”. In : *Comput. Graph. Forum* 17.3 (1998), p. 41-52.
- [167] Long QIAN et al. “Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display”. In : *Int J CARS* 12 (2017), p. 901-910.
- [168] Yuan Yuan QIAN et Robert J. TEATHER. “The eyes don’t have it : An empirical comparison of head-based and eye-based selection in virtual reality”. In : *SUI 2017 - Proc. 2017 Symp. Spat. User Interact.* (2017), p. 91-98.
- [169] François RAMEAU et al. “A real-time augmented reality system to see-through cars”. In : *IEEE transactions on visualization and computer graphics* 22.11 (2016), p. 2395-2404.
- [170] Ramesh RASKAR, Greg WELCH et Wei-Chao CHEN. “Table-top spatially-augmented reality : bringing physical models to life with projected imagery”. In : *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99)*. IEEE. 1999, p. 64-71.
- [171] H. REGENBRECHT et al. “An Augmented Virtuality Approach to 3D Videoconferencing”. In : *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. ISMAR ’03. USA : IEEE Computer Society, 2003, p. 290.
- [172] Dirk REINERS et al. “Augmented reality for construction tasks : Doorlock assembly”. In : *Proceedings of the international workshop on Augmented reality : placing artificial objects in real scenes : placing artificial objects in real scenes*. AK Peters, Ltd. 1999, p. 31-46.
- [173] Patrick REIPSCHLÄGER et Raimund DACHSELT. “DesignAR : Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays”. In : *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ISS ’19. Daejeon, Republic of Korea : Association for Computing Machinery, 2019, p. 29-41.
- [174] Jonathan C. ROBERTS. “State of the art : Coordinated & multiple views in exploratory visualization”. In : *Proc. - Fifth Int. Conf. Coord. Mult. Views Explor. Vis. C. 2007* (2007), p. 61-71.
- [175] AF ROVERS et HA VAN ESSEN. “Guidelines for haptic interpersonal communication applications : an exploration of foot interaction styles”. In : *Virtual Reality* 9.2-3 (2006), p. 177-191.
- [176] David RUDI et al. “Gaze-based interactions in the cockpit of the future : a survey”. In : *Journal on Multimodal User Interfaces* 14.1 (2020), p. 25-48.
- [177] Mie SATO. “Effects of Auditory Cues on Grasping a Virtual Object with a Bare Hand”. In : (2016), p. 2016-2017.

- [178] Joey SCARR et al. “Dips and ceilings : understanding and supporting transitions to expertise in user interfaces”. In : *Proceedings of the sigchi conference on human factors in computing systems*. New York, NY, USA : Association for Computing Machinery, 2011, p. 2741-2750.
- [179] Maximilian SCHERR. “Multiple and coordinated views in information visualization”. In : *Trends in Information Visualization* 38 (2008), p. 1-33.
- [180] Linda E. SIBERT et Robert J. K. JACOB. “Evaluation of eye gaze interaction”. In : *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '00* (2000), p. 281-288.
- [181] Mark SIMPSON, Jiayan ZHAO et Alexander KLIPPEL. “Take a walk : Evaluating movement types for data visualization in immersive virtual reality”. In : *Proceedings of Workshop on Immersive Analytics (IA)*. 2017.
- [182] Aw Kien SIN et Halimah Badioze ZAMAN. “Tangible interaction in learning astronomy through augmented reality book-based educational tool”. In : *International Visual Informatics Conference*. Springer. 2009, p. 302-313.
- [183] Gurjot SINGH et al. “Depth Judgment Measures and Occluding Surfaces in Near-Field Augmented Reality”. In : *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*. APGV '10. Los Angeles, California : Association for Computing Machinery, 2010, p. 149-156.
- [184] CPias : Réseau national de prévention des infections associées au SOIN. *Les tenues professionnelles dans les établissements de santé*. 2008. URL : http://www.cpias.fr/nosobase/recommandations/cclin_arlin/cclinSudEst/2008_personnel_CCLIN.pdf.
- [185] R. William SOUKOREFF et I. Scott MACKENZIE. “Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts’ law research in HCI”. In : *Int. J. Hum. Comput. Stud.* 61.6 (2004), p. 751-789.
- [186] Oleg ŠPAKOV et Päivi MAJARANTA. “Enhanced Gaze Interaction Using Simple Head Gestures”. In : *UbiComp '12*. Pittsburgh, Pennsylvania : Association for Computing Machinery, 2012, p. 705-710.
- [187] Maximilian SPEICHER, Brian D. HALL et Michael NEBELING. “What is Mixed Reality?” In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk : Association for Computing Machinery, 2019, p. 1-15.
- [188] Jonathan STEUER. “Defining virtual reality : Dimensions determining telepresence”. In : *Journal of communication* 42.4 (1992), p. 73-93.
- [189] Hideyuki SUENAGA et al. “Vision-based markerless registration using stereo vision and an augmented reality surgical navigation system : a pilot study”. In : *BMC Med. Imaging* 15.1 (2015), p. 51.
- [190] Mengu SUKAN et al. “Quick viewpoint switching for manipulating virtual objects in hand-held augmented reality using stored snapshots”. In : *ISMAR 2012 - 11th IEEE Int. Symp. Mix. Augment. Real. 2012, Sci. Technol. Pap.* (2012), p. 217-226.
- [191] J Edward SWAN, Gurjot SINGH et Stephen R ELLIS. “Matching and reaching depth judgments with real and augmented reality targets”. In : *IEEE transactions on visualization and computer graphics* 21.11 (2015), p. 1289-1298.

- [192] J. Edward SWAN et al. “Egocentric depth judgments in optical, see-through augmented reality”. In : *IEEE Trans. Vis. Comput. Graph.* 13.3 (2007), p. 429-442.
- [193] Markus TATZGERN et al. “Exploring real world points of interest : Design and evaluation of object-centric exploration techniques for augmented reality”. In : *Pervasive Mob. Comput.* 18 (2015), p. 55-70.
- [194] Yang TIAN et al. “Virtually-Extended Proprioception : Providing Spatial Reference in VR through an Appended Virtual Limb”. In : *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA : Association for Computing Machinery, 2020, p. 1-12.
- [195] Jocelyne TROCCAZ, Michael PESHKIN et Brian DAVIES. “The use of localizers, robots and synergistic devices in CAS”. In : *CVRMed-MRCAS'97*. Springer. 1997, p. 725-736.
- [196] Philip TUDDENHAM, David KIRK et Shahram IZADI. “Graspables revisited : Multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks”. In : *Conf. Hum. Factors Comput. Syst. - Proc.* 4 (2010), p. 2223-2232.
- [197] Brygg ULLMER et al. “Tangible menus and interaction trays”. In : *Proc. 2nd Int. Conf. Tangible Embed. Interact. - TEI '08*. 2008, p. 209.
- [198] Petr VÁVRA et al. “Recent development of augmented reality in surgery : a review”. In : *Journal of healthcare engineering 2017* (2017).
- [199] Mélodie VIDAL, Andreas BULLING et Hans GELLERSEN. “Pursuits : Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets”. In : *UbiComp '13*. Zurich, Switzerland : Association for Computing Machinery, 2013, p. 439-448.
- [200] Thomas VINCENT. “Handheld augmented reality interaction : spatial relations”. 2014GRENM032. Thèse de doct. 2014. URL : <http://www.theses.fr/2014GRENM032>.
- [201] VIVE. *Casque de VR HTC Vive*. 2020. URL : <https://www.vive.com/fr/product/vive/>.
- [202] Robin VIVIAN, David BERTOLO et Jérôme DINET. “Definition of Tactile Interactions for a Multi-Criteria Selection in a Virtual World”. In : *Int. J. Adv. Comput. Sci. Appl.* 5.12 (2014), p. 65-71.
- [203] Dong WANG et al. “M2asr : Ambitions and first year progress”. In : *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*. IEEE. 2017, p. 1-6.
- [204] Chris WEAVER. “Building highly-coordinated visualizations in improvise”. In : *IEEE Symposium on Information Visualization*. IEEE. 2004, p. 159-166.
- [205] David WEIMER et S Kicha GANAPATHY. “A synthetic visual environment with hand gesturing and voice input”. In : *ACM SIGCHI Bulletin* 20.SI (1989), p. 235-240.

- [206] Matthew WHITAKER et Esther KUKU. “Google Glass : the future for surgical training?” In : *The Bulletin of the Royal College of Surgeons of England* 96.7 (2014), p. 223-223.
- [207] Sean WHITE, David FENG et Steven FEINER. “Interaction and presentation techniques for shake menus in tangible augmented reality”. In : *Sci. Technol. Proc. - IEEE 2009 Int. Symp. Mix. Augment. Reality, ISMAR 2009* July (2009), p. 39-48.
- [208] Andrew WILSON et Steven SHAFER. “XWand : UI for Intelligent Spaces”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. Ft. Lauderdale, Florida, USA : Association for Computing Machinery, 2003, p. 545-552.
- [209] Chadwick A WINGRAVE et Doug A BOWMAN. “Baseline Factors for Raycasting Selection”. In : *Virtual Real. Int.* (2004).
- [210] Markus WIRTH et al. “Evaluation of interaction techniques for a virtual reality reading room in diagnostic radiology”. In : *UIST 2018 - Proc. 31st Annu. ACM Symp. User Interface Softw. Technol.* (2018), p. 867-876.
- [211] Jason WITHER et Tobias HÖLLERER. “Pictorial depth cues for outdoor augmented reality”. In : *Proc. - Int. Symp. Wearable Comput. ISWC 2005* (2005), p. 92-99.
- [212] Viktoria WITKA. “Perceptual Issues in Multi-Display Environments”. In : *Mobile Multi-Display Environments* (2016), p. 1.
- [213] Yukang YAN et al. “Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality”. In : *Proc. CHI* (2018), p. 1-13.
- [214] Yukang YAN et al. “HeadGesture : Hands-Free Input Approach Leveraging Head Movements for HMD Devices”. In : *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.* 2.4 (2018), p. 1-23.
- [215] Shanhe YI et al. “GlassGesture : Exploring head gesture interface of smart glasses”. In : *Proc. - IEEE INFOCOM 2016-Sept* (2016), p. 1017-1018.
- [216] M. L. YUAN, S. K. ONG et A. Y.C. NEE. “Augmented reality for assembly guidance using a virtual interactive tool”. In : *Int. J. Prod. Res.* 46.7 (2008), p. 1745-1767.
- [217] Steven YULE et al. “Non-technical skills for surgeons in the operating room : a review of the literature”. In : *Surgery* 139.2 (2006), p. 140-149.
- [218] Steven YULE et al. “Surgeons’ non-technical skills in the operating room : reliability testing of the NOTSS behavior rating system”. In : *World journal of surgery* 32.4 (2008), p. 548-556.
- [219] Jürgen ZAUNER et al. “Authoring of a mixed reality assembly instructor for hierarchical structures”. In : *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings*. IEEE. 2003, p. 237-246.
- [220] Andre ZENNER et Antonio KRÜGER. “Shifty : A weight-shifting dynamic passive haptic proxy to enhance object perception in virtual reality”. In : *IEEE transactions on visualization and computer graphics* 23.4 (2017), p. 1285-1294.

- [221] Shengdong ZHAO et Ravin BALAKRISHNAN. “Simple vs. compound mark hierarchical marking menus”. In : *Proc. 17th Annu. ACM Symp. User interface Softw. Technol. UIST 04* 6.2 (2004), p. 33.
- [222] Ning-Ning ZHOU et Yu-Long DENG. “Virtual reality : A state-of-the-art survey”. In : *International Journal of Automation and Computing* 6.4 (2009), p. 319-325.