



HAL
open science

Implémentation et bénéfices des systèmes d'interaction haute-fidélité: d'un contrôle plus performant à un apprentissage d'objets 3D plus rapide

Thibault Louis

► To cite this version:

Thibault Louis. Implémentation et bénéfices des systèmes d'interaction haute-fidélité: d'un contrôle plus performant à un apprentissage d'objets 3D plus rapide. Autre [cs.OH]. Université Grenoble Alpes [2020-..], 2020. Français. NNT: 2020GRALM047 . tel-03117757

HAL Id: tel-03117757

<https://theses.hal.science/tel-03117757>

Submitted on 21 Jan 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

Thibault LOUIS

Thèse dirigée par **François BÉRARD**, Maître de Conférence,
Université Grenoble Alpes, et

Amélie ROCHET-CAPELLAN, Chargée de recherche, CNRS, et

Jocelyne TROCCAZ, Directrice de recherche, CNRS.

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

Implémentation et bénéfices des systèmes d'interaction haute-fidélité : d'un contrôle plus performant à un apprentissage d'objets 3D plus rapide

Thèse soutenue publiquement le **22 octobre 2020**,
devant le jury composé de :

Madame Maud MARCHAL

Professeur, Université de Rennes, Rapportrice

Monsieur Emmanuel DUBOIS

Professeur, Université de Toulouse, Rapporteur

Madame Joëlle THOLLOT

Professeur, Université Grenoble-Alpes, Présidente

Monsieur Patrick REUTER

Maître de conférences, Université de Bordeaux, Examineur

Madame Jocelyne TROCCAZ

Directrice de recherche, CNRS, Directrice de thèse

Monsieur François BÉRARD

Maître de conférences, Université Grenoble-Alpes, Co-encadrant (invité)

Madame Amélie ROCHET-CAPELLAN

Chargée de recherche, CNRS, Co-encadrante (invitée)



Résumé

Interagir avec des scènes virtuelles 3D est nécessaire à de nombreuses applications. Parmi elles : la visualisation de données 3D, la conception assistée par ordinateur, les simulateurs d'entraînement ou encore les jeux vidéos. Effectuer cette tâche par l'intermédiaire de systèmes en 2D tels que des ordinateurs de bureau ou des tablettes tactiles peut s'avérer fastidieux. Pour interagir plus efficacement avec des contenus 3D, les *systèmes d'interaction haute-fidélité* tels que les casques de réalité virtuelle tentent de reproduire les modalités d'interaction dont nous disposons dans le monde réel. Ces systèmes offrent un rendu visuel stéréoscopique couplé à la position de l'utilisateur et un contrôle isomorphe des objets 3D. Cependant, la littérature manque de résultats issus d'expériences rigoureuses montrant leurs bénéfices. L'objectif de cette thèse est double. Nous cherchons d'une part à étoffer la littérature par l'intermédiaire d'études utilisateur contrôlées permettant d'apporter des résultats robustes sur les bénéfices de ces systèmes. D'autre part, nous nous efforçons de fournir les moyens d'implémenter les expériences haute fidélité les plus performantes.

Dans ce manuscrit, nous présentons, tout d'abord, un état de l'art des dispositifs haute-fidélité existants et de leurs bénéfices potentiels. Nous introduisons notamment une approche prometteuse appelée appareil à perspective corrigée déplaçable (APCD), que nous avons particulièrement étudiée dans cette thèse.

Nous détaillons ensuite deux contributions qui nous ont permis de quantifier les bénéfices des systèmes haute-fidélité. Afin de témoigner du potentiel applicatif de ces systèmes, nous avons étudié leur apport dans la réalisation de deux tâches mettant en jeu des processus cognitifs très différents. La première étude concerne une tâche de positionnement d'objets 3D dans l'espace. Les deux systèmes haute-fidélité que nous avons testés : un APCD et un casque de réalité virtuelle, se sont respectivement montrés 43% et 29% plus efficaces que le statu quo (un bras articulé utilisé avec un écran plat). La seconde étude concerne une tâche d'apprentissage de la forme d'un objet 3D. Pour cette tâche, les deux mêmes systèmes haute-fidélité ont permis d'améliorer de 27% les performances de reconnaissance de l'objet par rapport à un apprentissage via une tablette tactile.

Nous présentons ensuite deux contributions apportant des solutions pour faciliter l'implémentation matérielle et logicielle des systèmes d'interaction haute-fidélité. Nous proposons une méthode permettant d'évaluer l'impact de différents paramètres techniques sur le sentiment de présence éprouvée au cours d'une expérience interactive : un sentiment qui témoigne de la fidélité de l'expérience relativement à la réalité simulée. Mettre en jeu cette méthode dans une étude utilisateur nous a permis d'identifier que la stabilité du système de suivi et la fréquence d'affichage sont les paramètres les plus critiques avec l'APCD que nous avons testé. Nous proposons enfin une panoplie de techniques d'interaction permettant l'implémentation d'applications adaptées aux APCD sphériques et à tout autre système fournissant un écran manipulable à deux mains. Les interactions que nous proposons tirent parti du contrôle efficace des rotations de l'appareil et se sont montrées intuitives et efficaces lors d'un test qualitatif dans une application d'apprentissage d'anatomie.

Abstract

Interacting with 3D virtual scenes is essential for numerous applications. Among others : 3D data visualization, computer assisted design, training simulators and video games. Performing this task through 2D systems like desktop computers or multi-touch tablets can be tedious. To interact more efficiently with 3D contents, *high fidelity interactive systems* such as virtual reality head-mounted displays try to reproduce the interactive modalities available in real life. Such systems offer a stereoscopic head-coupled rendering and an isomorphic control of 3D objects. However, there is a lack of rigorous studies that showed their benefits in the literature. This thesis has two purposes. We want to enrich the literature through controlled user studies that bring robust results on high fidelity systems' benefits. We also seek to provide the means to implement the most efficient high fidelity experiences.

In this manuscript, we start by presenting a state of the art of existing high fidelity devices and their potential benefits. We especially introduce a promising approach called handheld perspective corrected displays (HPCD), that we particularly studied through this thesis.

We then present two contributions that allowed us to quantify high fidelity systems benefits. We studied two tasks involving very different cognitive processes in order to attest the variety of applications that could benefit from those systems. The first study concerns a 6D docking task. The two high fidelity systems that we tested, an HPCD and a virtual reality head mounted display, performed respectively 43% and 29% more efficiently than the status quo (an articulated arm used alongside a flat screen). The second study focuses on the task of learning the shape of an unknown 3D object. Regarding this task, the two previously studied high fidelity systems allowed to enhance by 27% the object's recognition performances when compared to the use of a multi-touch tablet.

We then present two other contributions that bring solutions to ease both hardware and software implementation of high fidelity systems. We provide a method to evaluate the impact of several technical parameters on the presence felt during an interactive experience, which is a feeling that testifies to the experience's fidelity with regard to the simulated reality. Using this method in a user study allowed us to identify the fact that, with the tested HPCD, the tracking stability and the rendering frame rate were the most critical parameters concerning presence. We finally suggest a suit of interacting techniques that enable the implementation of applications well suited for spherical HPCD, and any other devices that provide a manipulable screen held with both hands. The proposed interactions take advantage of the efficient control of the device rotations and appeared to be both intuitive and efficient during a qualitative test in an anatomy learning application.

Remerciements

Je remercie les membres de mon jury de thèse qui se sont montrés très réactifs malgré la situation sanitaire incertaine, et dont les rapports et l'évaluation de qualité ont permis de valider sérieusement un travail au demeurant farfelu.

Je remercie mes directrices et mon directeur de thèse pour avoir réussi à m'encadrer tout au long de ce travail, nous permettant d'aboutir à de jolis résultats de recherche. Je remercie François pour son esprit critique, Amélie pour ses qualités humaines et Jocelyne pour sa capacité à prendre du recul, et pour avoir mené le projet An@tomy 2020 qui a financé cette thèse.

Merci aux copains/copines docteurs et futurs docteurs, aux collègues qui font des blagues parfois douteuses et à ceux qui ont des abdos en béton, pour l'ambiance agréable durant ces quelques années que nous avons passées ensemble.

Merci à mes amis : l'apocalypse, les grenoblois, les tagazous et tous les autres, et à ma famille et ma belle-famille, pour tous les bons moments que nous passons ensemble, indispensables pour se ressourcer entre deux publications difficiles. Merci à Agnès pour l'immense travail de relecture de ce manuscrit.

Enfin, merci à Corélie pour avoir toujours été là pour moi, et pour tout le reste.

Merci également à toutes celles et ceux qui ont lu et liront ce manuscrit. J'espère que votre lecture sera plaisante et qu'elle vous apportera les réponses aux questions que vous vous posez.

Table des matières

Introduction	19
I État de l’art	24
1 Des systèmes d’interaction haute-fidélité	25
Introduction	25
1.1 Définitions	26
1.1.1 Différentes réalités	26
1.1.2 Fidélité d’un système interactif	26
1.1.3 Des réalités fidèles	27
1.1.4 Le sentiment de présence	28
1.2 Les systèmes HF existants	29
1.2.1 Performances reportées et capacités humaines	29
1.2.2 Les systèmes de réalité virtuelle	30
1.2.3 Les systèmes de réalité augmentée	37
1.2.4 Bilan	44
1.3 Un APCD haute-fidélité	45
1.3.1 Les premiers prototypes d’APCD	45
1.3.2 Une implémentation HF	46
1.3.3 Les Avantages des APCD	49
Conclusion	52
2 Les bénéfices attendus de la haute fidélité	54
Introduction	54
2.1 Un contrôle d’objets plus performant	56
2.1.1 Placer des objets dans l’espace	56
2.1.2 Périphériques permettant un contrôle 6D d’objets virtuels	56
2.1.3 Bénéfices d’un contrôle isomorphe	61
2.1.4 Bilan	62
2.2 Un Apprentissage plus efficace	63
2.2.1 Adapter l’interaction aux besoins de la tâche d’apprentissage	63
2.2.2 Apprendre un objet 3D	64
2.2.3 Des techniques d’interaction plus fidèles pour mieux apprendre un objet 3D	67
2.2.4 Bilan	70
2.3 Un meilleur sentiment de présence	71
2.3.1 Les définitions de la présence dans la littérature	71
2.3.2 Les applications de la présence	72
2.3.3 Comment produire un haut niveau de présence	75
2.3.4 Mesurer la présence	77
2.3.5 Bilan	79

Conclusion	80
II Pourquoi	81
3 Supériorité de l'APCD dans une tâche de placement 6D	82
Introduction	82
3.1 Un indice de difficulté pour les tâches de placement 6D	84
3.1.1 Intérêt de la loi de Fitts	84
3.1.2 Adaptation à une tâche 6D	84
3.2 Étude utilisateur	88
3.2.1 Détails technique	88
3.2.2 Conditions expérimentales	89
3.2.3 Protocole expérimental	91
3.2.4 Résultats	93
3.3 Discussion	97
3.3.1 Validité de l'indice de difficulté	97
3.3.2 Supériorité des techniques haute-fidélité	97
3.3.3 Supériorité de l'APCD sur le casque de RV	99
Conclusion	102
4 Supériorité de la haute fidélité dans une tâche d'apprentissage d'objets 3D	103
Introduction	103
4.1 Conception de l'étude utilisateur	105
4.1.1 Vue d'ensemble	105
4.1.2 Protocole expérimental	106
4.1.3 Conditions expérimentales	108
4.1.4 Installation technique	111
4.2 Résultats et Analyses	112
4.2.1 Méthodologie	112
4.2.2 Résultats de reconnaissance	113
4.2.3 Résultats des VMRT	115
4.2.4 Analyse de la quantité de rotation	116
4.2.5 Résultats subjectifs	117
4.3 Discussion	118
4.3.1 Supériorité des groupes HF	118
4.3.2 Apprentissage d'objets 3D et capacités de rotation mentale	119
4.3.3 Un protocole expérimental ambitieux	120
Conclusion	122
III Comment	123
5 Mesurer l'impact des paramètres techniques d'un système HF sur la présence	124
Introduction	124
5.1 Une nouvelle méthode pour évaluer la présence	126
5.1.1 Une mesure relative plutôt qu'une échelle absolue	126
5.1.2 Présentation de la méthode	127
5.2 Conception de l'étude utilisateur	129
5.2.1 Choix des paramètres expérimentaux	129

5.2.2	Protocole expérimental	131
5.3	Résultats de l'étude	134
5.3.1	Méthodologie	134
5.3.2	Effet de l'entité	134
5.3.3	Résultats des comparaisons	135
5.3.4	Résultats subjectifs	136
5.4	Discussion	138
5.4.1	Analyse des résultats	138
5.4.2	Compromis de performance	140
5.4.3	Pertinence de la méthode d'évaluation de la présence présentée	140
	Conclusion	142
6	Comment implémenter une application entièrement interactive avec un APCD	143
	Introduction	143
6.1	Analyse du contexte d'utilisation ciblé	145
6.1.1	Un hublot volumique, tangible et portable	145
6.1.2	Interagir avec des environnements virtuels 3D	146
6.1.3	S'adapter aux contraintes physiques de l'approche	147
6.1.4	Compromis présence-efficacité	149
6.2	Conception d'une panoplie interactive pour APCD	150
6.2.1	Sélection d'objets	150
6.2.2	Manipulation d'objets	154
6.2.3	Zoom	155
6.2.4	Conteneurs de Widgets	157
6.2.5	Widgets	159
6.3	Démonstration d'usage de la panoplie interactive	164
6.3.1	Une application pour apprendre l'anatomie	164
6.3.2	Évaluation qualitative	166
6.3.3	Perspectives	168
	Conclusion	170
	Conclusion	171
A	Implémentation du prototype d'APCD de Bérard et Louis	175
A.1	Calibration	175
A.2	Implémentation logicielle	179

Table des figures

1.1	Schéma proposé par Cruz et al. pour illustrer la rétro-projection dans une CAVE à quatre faces (source : [41]).	31
1.2	À gauche : Utilisation d'un bras articulé par Pontonnier et al. pour ajouter un retour haptique dans une CAVE (source : [114]). À droite : utilisation d'un cube tangible par Zielinski et al. pour manipuler des entités virtuelles dans une CAVE (source : [171]).	31
1.3	La CAVE Immersia à Rennes, de près de 10m de long.	32
1.4	À gauche : Le système d'aquarium de RV développé par Arthur et al. (source : [4]). À droite : Le système d'aquarium de RV à 360° développé par Bimber et al. (source : [18]).	33
1.5	À gauche : L'écran volumétrique utilisé par Grossman et al. (source : [59]). À droite : le système d'aquarium de RV pour deux utilisateurs de Zhou et al. (source : [170]).	34
1.6	À gauche : le casque de RV testé par Patrick et al., utilisant un Visette Pro pour l'affichage et un Ascention SpacePad pour la capture de position (source : [111]). Au milieu : le Vive de HTC sorti en 2016, avec ses contrôleurs isomorphes et les bases permettant la capture optique. À droite : la sphère tangible utilisée par Englmeier et al. (source : [50]).	35
1.7	À gauche : l'occulus Quest, les caméras permettent la capture de mouvement sans nécessiter de dispositif externe. À droite : le Xtal de Vrgineers, offrant une résolution 5K et un champ de vision de 180°.	36
1.8	À gauche : Le DigitalDesk de Wellner (source : [162]). Au milieu : utilisation de plusieurs projecteurs par Raskar et al. pour augmenter une surface 3D (source : [121]). À droite : le système de RA spatialisée mobile de Cao et Balakrishnan (source : [29]).	37
1.9	A gauche : le système de Jones et al. détectant automatiquement les briques tangibles utilisées pour interagir (source : [79]). À droite : le système de Hilliges et al. utilisant un miroir semi-transparent (source : [68]).	38
1.10	À gauche : illustration du jeu Pokemon Go de Niantic, exploitant la RA mobile (2016). À droite : illustration de la librairie logicielle pour la RA ArCore de Google. Des points apparaissent au niveau des surfaces détectées automatiquement et permettent d'y ancrer des objets virtuels.	40
1.11	À gauche : différences d'affichage d'une tablette tactile en fonction de la mise en place d'un couplage à la position de la tablette (à gauche) ou d'un couplage à la position de l'utilisateur (à droite) (source : [7]). À droite : illustration de l'utilisation du Phab 2 Pro de Lenovo dans un contexte de RA mobile.	41
1.12	À gauche : le système complet de RA mobile en extérieur de Höllerer et al. (source : [73]). À droite : schéma montrant le fonctionnement du prototype de Maimone et Fuchs permettant un affichage multi-focal opaque d'objets virtuels (source : [98]).	42

1.13	À gauche : le casque de RA meta 2 de metavision. La TV en arrière-plan affiche la vue de l'utilisateur alors qu'il déplace l'objet virtuel avec sa main (2017). À droite : illustration de casque de RA Hololens 2 de Microsoft représentant le panneau virtuel vu par l'utilisatrice (2020).	43
1.14	A gauche : le premier prototype d'APCD : le pCubee de Stavness et al. (source : [142]). À droite : le nouveau pCubee de Tang et al. (source : [148]).	45
1.15	À gauche : un utilisateur manipulant l'APCD sphérique dans un jeu de suivi de chemin avec une bille. À droite : deux APCD peuvent être utilisés en même temps, ici la position du soleil influe sur l'éclairage de la planète. Les deux images sont présentées en vue monoscopique adaptée au point de vue de l'appareil photo à des fins d'illustration.	46
1.16	Configuration du projecteur et du système de suivi optique. L'utilisateur porte des lunettes à obturateur. Des marqueurs passifs permettent le suivi des lunettes et de l'APCD. (source : [25]).	48
1.17	Différents APCD conçus par Bérard et Louis. À gauche : le petit APCD sphérique de 14 cm de diamètre. Au milieu : le grand APCD sphérique de 30 cm de diamètre. À droite : l'APCD cylindrique de 27 cm de long. Les trois images sont présentées en vue monoscopique adaptée au point de vue de l'appareil photo à des fins d'illustration.	49
1.18	Illustration de Issartel et al. montrant comment un APCD peut être utilisé pour sélectionner et manipuler des objets virtuels au sein d'une large scène 3D (source : [74]).	50
2.1	Widgets utilisés dans le logiciel blender pour translater un objet (à gauche) ou pour orienter un objet (à droite). Sélectionner avec la souris l'un des axes ou des arcs de couleur permet de contraindre le mouvement suivant le seul degré de liberté sélectionné.	56
2.2	Périphériques de contrôle 6D électromagnétiques. À gauche : le Bat de Ware (source : [160]). À droite : le Finger Ball de Zhai et al. (source : [168]). . .	57
2.3	Des périphériques de contrôle 6D fixe. A Gauche : le Space Mouse de DLR (source : [167]). Au milieu : le GlobeFish de Froehlich et al. (source : [54]). A droite : le Phantom Omni de Geomatic.	58
2.4	Techniques d'interaction tactile multi-touch. À gauche : un I-pad Air de Apple. À droite : des rotations 2D mappées sur les déplacements 2D d'un doigt lorsqu'un autre doigt reste fixe, technique introduite par Liu et al. (source : [91]).	59
2.5	Techniques d'interaction à main levée avec systèmes de capture optique. À gauche : le contrôleur du Vive de HTC, avec plusieurs boutons. Au milieu : interagir en manipulant une représentation physique de l'objet virtuel (source : [157]). À droite : manipuler en bougeant sa main, sans périphérique d'entrée. (source : [157]).	60
2.6	À gauche : une image d'IRM de genou. À droite : un dessin technique au trait.	65
2.7	Un élément du test de rotation mentale de Vandenberg (VMRT). Le participant doit trouver deux images parmi les quatre de droite qui représentent le même objet que celui qui est affiché sur l'image de gauche, vu à partir de différents points de vue.	65

2.8	Des exercices proposés par Contero et al. pour entraîner les capacités de RM en utilisant un système de RA sur ordinateur. L'utilisateur manipule une feuille sur laquelle le modèle 3D est ajouté virtuellement. À gauche, les prismes marqués des lettres A à E doivent être identifiés sur une vue 2D au trait. À droite, les vues 2D correspondant au modèle 3D manipulé doivent être identifiées parmi un ensemble de vues disponibles (source : [39]).	66
2.9	Une visualisation statique proposée par Berney et al., affichant cinq images du mouvement de l'omoplate. Ces images sont extraites d'une visualisation dynamique affichant la totalité du mouvement dans une vidéo de 118 secondes (source : [13]).	67
2.10	À gauche : le dispositif de RV utilisé par Schnabel et al. et un volume cubique à analyser (source : [126]). Au milieu : L'environnement de RV de dissection de grenouille utilisé par Lee et al. (source : [87]). À droite : le dispositif de RV (Dextroscope) utilisé par Jang et al. (source : [77]).	68
2.11	À gauche : La visualisation de RA sur ordinateur utilisée par Chen et al. (source : [34]). À droite : Le dispositif de RA utilisé par Shin et al., implémenté avec un casque opaque et une caméra faible latence (source : [131]).	69
2.12	Un utilisateur manipulant intuitivement un crâne virtuel, qu'il perçoit avec un fort sentiment de présence dans la sphère physique. Il peut retourner la sphère pour voir l'arrière du crâne.	72
2.13	Un utilisateur est virtuellement exposé à son acrophobie.	73
2.14	Des pilotes sont entraînés par l'intermédiaire d'un simulateur de vol.	74
2.15	Interface pour personnifier son personnage dans le jeu vidéo <i>Monster Hunter World</i> développé par Capcom (2018).	75
2.16	À gauche : un système très haute fidélité incluant un casque opaque couvrant la totalité du champ de vision et produisant des retours auditifs 3D, une combinaison complète à retour haptique, et une plateforme permettant de simuler un déplacement naturel, imaginé dans le film <i>Ready Player One</i> (2018). À droite : l'effet du ventriloque qui produit l'illusion que la voix provient de la marionnette dont les lèvres bougent.	76
3.1	Illustration de la tolérance d'erreur avec l'exemple de la boîte à formes (illustrée à gauche). Au milieu : cas où toute la tolérance d'erreur (TE) est utilisée pour une erreur de rotation. À droite : cas où toute la tolérance d'erreur est utilisée pour une erreur de translation.	85
3.2	Paramètres d'une tâche de placement 6D d'amplitude de translation A et d'amplitude de rotation α . L'objet (forme grise avec un contour plein épais) doit être placé dans une cible (expansion rouge englobant le volume de l'objet d'une marge TE). Dans cet exemple, les composantes d'erreurs de translation (en bleu) et de rotation (en vert) sont explicitées. La tolérance d'erreur globale TE peut être approximée comme la somme de tolérances d'erreurs flottantes de translation TE_{trans} et d'orientation TE_{rot} . D_{max} est la plus grande distance entre deux points de l'objet, elle permet la conversion d'une erreur d'orientation en une erreur spatiale (détails dans le texte).	86
3.3	Les conditions expérimentales. a) le bras articulé (PHANTOM). b) l'appareil tangible (PROP). c) le casque de RV (RV). d) l'APCD (APCD).	88
3.4	La scène vue au travers du casque de RV. Seul un plan texturé au niveau du sol est ajouté en plus de la chaise et de la cible semi-transparente.	90

3.5	Visualisation de la scène lors d'une tâche de placement. Cette capture d'écran est éditée à des fins d'illustration : l'arrière-plan était noir et non pas blanc pendant l'expérience, la cible était blanche en semi-transparence et non pas grise, et une seule chaise était affichée et contrôlée en temps réel. Les parties de la chaise à l'intérieur de la cible deviennent vertes. La flèche rouge pointant vers la cible n'est affichée que lorsque la cible est éloignée.	92
3.6	En haut : Le <i>temps d'acquisition</i> dans chaque condition, moyenné sur les participants et avec intervalles de confiance à 95%. bas : ratio des <i>temps d'acquisition</i> des participants pour chaque paire de conditions, moyenné sur les participants et avec intervalle de confiance à 95%.	93
3.7	Le <i>temps d'acquisition</i> en fonction de l'ID des cibles pour chaque condition, avec intervalle de confiance à 95%. Les régressions linéaires sont représentées avec l'ID de 6 ignoré pour la condition PHANTOM (détails dans le texte).	94
3.8	Évaluations subjectives des participants, avec intervalle de confiance à 95%.	95
3.9	Correction d'erreurs en fonction du temps écoulé. Les essais sont segmentés en deux phases (ligne verticale) : une phase balistique suivie d'une phase corrective.	98
3.10	Temps passé dans la phase balistique (en haut) par rapport au temps passé dans la phase corrective (en bas), au cours des tâches de placement, moyenné sur les participants et les ID, avec intervalles de confiance à 95%.	99
4.1	Les trois techniques d'interaction testées : la technique HF avec le casque de RV (à gauche), la technique HF avec l'APCD (au milieu), et la technique basse fidélité avec la tablette (à droite).	105
4.2	Le sphinx (à gauche) et les trois distracteurs (à droite).	105
4.3	La chronologie de l'expérience par jour.	106
4.4	Configuration expérimentale montrant la tablette (A), le dispositif sphérique (B), et les marqueurs pour la validation par tap (C), pour tous les groupes, lors d'une session d'entraînement à manipulation dictée.	109
4.5	Une tâche de reconnaissance. Cette tâche est identique pour tous les groupes.	110
4.6	Taux de réussite (en haut) et temps de réponse moyen (en bas) au cours des trois sessions de reconnaissance, par groupe, avec les intervalles de confiance à 95%.	114
4.7	Score moyen des participants regroupés par technique d'interaction et session, avec les intervalles de confiance à 95%.	115
4.8	Scores moyens des pré-tests et post-tests de VMRT, et la différence entre les deux, par groupe, avec les intervalles de confiance à 95%.	115
4.9	Notes subjectives moyennes des participants de chaque groupe, avec les intervalles de confiance à 95%.	117

5.1	Représentation schématique d’une comparaison. Un bouton permet à l’utilisateur de basculer à tout moment entre les deux représentations. Un autre bouton permet d’identifier la représentation courante comme la plus présente. Pour simplifier, on considère dans cet exemple trois paramètres : la résolution, la stéréoscopie et la fréquence de rafraîchissement (icônes d’écran, de casque de RV et de compteur de vitesse respectivement). Dans la représentation A, la résolution est altérée au niveau orange. Dans la représentation B, la stéréoscopie est altérée au niveau rouge. La fréquence de rafraîchissement n’est pas altérée dans ces deux représentations. Le niveau vert correspond au niveau optimal de chaque paramètre. La scène présentée (un papillon) ne doit pas changer entre les deux représentations d’une comparaison.	127
5.2	Les deux entités utilisées. À gauche : l’os statique. À droite : le papillon animé quand il vole à l’intérieur de la sphère. Pour mieux comprendre l’illustration, la stéréo a été désactivée sur cette image.	131
5.3	Une participante explorant une représentation de l’entité fémur (à gauche) et de l’entité papillon (à droite). La lumière du jour était couverte pendant l’expérience pour maximiser le contraste du projecteur.	132
5.4	Des marqueurs sont attachés aux chaussures de la participante pour détecter les “taps”.	133
5.5	Effet de l’entité : probabilité de l’hypothèse nulle de différence de préférence moyenne en fonction de l’entité, pour chaque comparaison. Plus une cellule est claire, plus un effet statistique de l’entité est probable pour la comparaison correspondante.	134
5.6	Préférences moyennes des utilisateurs par comparaison, avec les intervalles de confiance à 95%. Une valeur de 1 signifie que la condition définie par la ligne de la cellule a toujours été préférée à la condition définie par la colonne de la cellule. -1 signifie le contraire. À Gauche : résultats pour l’os. À droite : résultats pour le papillon. La teinte de la cellule code la moyenne et la saturation code la taille de l’intervalle de confiance.	135
5.7	format=empty	136
5.8	Plage de présence de toutes les représentations explorées par entité, représentée par les utilisateurs. Les côtés gauche et droit des cases représentent respectivement la moyenne des limites minimale et maximale de la plage. Les intervalles de confiance à 95% sont représentés par les segments horizontaux colorés.	137
5.9	Plaisir subjectif moyen et difficulté de la tâche, par entité, avec les intervalles de confiance à 95%.	137
5.10	Altérations ordonnées par la force de leur impact négatif sur la présence, du plus préjudiciable (en haut) au moins préjudiciable (en bas). Les rectangles regroupent les altérations qui appartiennent à la même classe d’équivalence. L’effet des altérations de résolution dépend de l’entité.	138
6.1	Un utilisateur manipulant un APCD.	143
6.2	Couper la scène. L’APCD est représenté en gris clair, le cône de vue en jaune clair. Les parties visibles des objets de la scène sont rouges, les parties clippées sont blanches. À gauche : clipping complet . Au milieu : clipping avant . À droite : sans clipping	147
6.3	Faire tourner l’APCD, vu de <i>dessus</i> . Positions initiales des mains en blanc, positions finales en rouge.	148

6.4	À Gauche : utilisation du picking pour sélectionner un objet en dehors de la zone accessible confortablement. À droite : l'utilisateur pointe vers une cible éloignée (l'os) avec le curseur. Les images du curseur et de l'os vues par son œil dominant sont alignées (à droite), mais pas les images vues par son œil non dominant (à gauche). L'objet sélectionnable est mis en évidence par un contour jaune.	153
6.5	Manipulation d'un objet après une sélection par picking. À gauche : avec un curseur standard, l'objet tourne autour du centre de la sphère. Au milieu : avec le curseur proxy , l'objet tourne autour du point de sélection. À droite : avec le curseur magnétique , l'objet est attiré vers le curseur. .	155
6.6	Rotations de la sphère effectuant des changements d'échelle de la scène centrées autour du centre de la sphère.	156
6.7	Pointage dans le marking menu : la translation de la sphère déplace le pointeur 2D dans le marking menu fixe (le point rouge représente le pointer 2D, la croix rouge représente le curseur 3D).	159
6.8	Pointage dans le panneau de contrôle : la rotation de la sphère déplace le pointeur 2D dans le panneau qui reste toujours centré dans la sphère. À des fins d'illustration, la stéréo est désactivée dans toutes les images du système réel, sauf dans la figure 6.1.	160
6.9	Interagir avec les widgets. À gauche : retour visuel progressif lors de l'activation d'un bouton. À droite : interaction par franchissement avec un slider.	161
6.10	Un menu radial et un slider dans le panneau de contrôle. À gauche : tous les sous-menus affichent uniquement des aperçus réduits. À droite : le nœud "A" est sélectionnée et son sous-menu est développé (révélant tous ses nœuds fils), tandis que les sous-menus de ses nœuds frères ("B" et "C") restent réduits.	163
6.11	L'application de démonstration. À gauche : identifier et sélectionner une structure anatomique à partir de sa forme. Au milieu : choisir une pièce à partir de son nom. À droite : la structure choisie dans le menu hiérarchique apparaît dans la scène. Le contour jaune indique la partie actuellement pointée.	164
6.12	Adapter l'échelle de la scène à la tâche. À gauche : les muscles nécessitent un petit facteur de zoom. À droite : les ligaments nécessitent un facteur de zoom important.	165
6.13	Notes subjectives des participants concernant l'intuitivité (en haut) et l'efficacité (en bas) de l'interface, avec les intervalles de confiance à 95%. . . .	168
6.14	Ressentis des participants, avec les intervalles de confiance à 95%.	168
A.1	À gauche : une grille de marqueurs est placée dans le champ de projection pour la calibration semi-automatique. À droite : la caméra fixée sous le projecteur voit les marqueurs qui reflètent la lumière envoyée par le projecteur.	175
A.2	Deux étapes de l'identification par code de Gray de la position des marqueurs dans le référentiel du projecteur. À gauche : les images projetées. À droite : la vue caméra correspondante montrant les marqueurs illuminés. .	176
A.3	Une étape de raffinement de l'identification de la position des marqueurs en projetant des lignes localement. À gauche : une image projetée. À droite : la vue caméra correspondante. Les carrés rouges indiquent les zones de l'image caméra correspondant à la position des marqueurs ; ils sont affichés dans un but de visualisation uniquement.	177

A.4 Répartition des marqueurs rétro-réflexifs sur l'APCD et sur les lunettes à obturateurs.	178
A.5 Affichage à perspective corrigée avec mappage de texture. Étape 1 : l'image de texture (base de la pyramide) est créée à partir d'une caméra virtuelle placée à la position de l'œil et visant le centre de l'APCD. Étape 2 : un maillage représentant la forme de l'APCD est texturé en projetant les sommets de chaque triangle dans le plan de projection de la caméra. À des fins d'illustration, un seul triangle d'un maillage sphérique grossier est texturé. La scène virtuelle (le cube multicolore) est affichée derrière l'APCD, mais l'approche se généralise à n'importe quelle position, par exemple à l'intérieur de l'appareil (source : [25]).	179

Liste des tableaux

1.1	Résumé des performances techniques des différentes approches HF existantes en 2020.	52
1.2	Résumé de caractéristiques générales des différentes approches HF existantes en 2020.	53
3.1	Indices de difficulté utilisés dans l'expérience. Ils sont calculés en résolvant numériquement le système d'équation 3.9. A et TE sont exprimés pour les conditions RV et APCD. Ces grandeurs sont toutes deux réduites d'un facteur six pour la condition PHANTOM, et d'un facteur deux pour la condition PROP, sans que cela n'impacte l'ID.	92
3.2	Paramètres des régressions linéaires prédisant le <i>temps d'acquisition</i> en fonction des ID pour chaque condition.	95
4.1	Coefficients de corrélation de Spearman (et p-valeur correspondantes) entre les résultats au pré-test de VMRT des participants et leur score de reconnaissance du jour 2, séparés par groupes ou tous regroupés.	116
4.2	Coefficient de corrélation de Spearman entre le temps de réponse et la quantité de rotation effectuée pendant un essai, séparé par type de session et par jour. Toutes les p-valeurs sont $< 2e-16$	116
5.1	La valeur de chaque paramètre RLFT pour chaque représentation comparée. Les valeurs en bleu sont les valeurs optimales du paramètre donné. . .	132

Introduction

*Quand j'étais enfant, j'ai vu disparaître les souris à boules, qui devaient être récurées au fur et à mesure qu'elles récoltaient des quantités de poussière dont je ne soupçonnais même pas l'existence sur mon bureau. Un jour, j'ai pu voir ma première vidéo sur internet. Ce jour là, heureusement, personne n'a appelé mes parents sur la ligne fixe : une communication téléphonique aurait interrompu la connexion ADSL. Le soir même, j'interrompais toutes mes activités de jeune adolescent pour ne pas rater "minority report" à la télévision : un film de science-fiction qui mettait en jeu des **techniques d'interaction** totalement fantaisistes, telles que des murs d'écrans et des interactions tactiles ou à main levée. J'imaginai alors le futur. J'imaginai un monde de science-fiction où l'information digitale serait disponible partout, tout le temps. Un monde où ces informations seraient parfaitement **intégrées dans notre réalité**, et où interagir avec elles serait tellement **intuitif** que les enfants ne sauraient plus comment tourner les pages d'un livre papier.*

*J'entrais en études supérieures quand mes amis avant-gardistes ont eu leur premier smartphone. Alors qu'ils nous montraient qu'ils pouvaient aller sur internet depuis n'importe où, nous soutenions, jalousement, que c'était un gadget inutile. Et de toute façon on voyait bien que la connexion n'était pas performante du tout, ça nous rappelait le 56k : l'ancêtre de l'ADSL. Sans oublier de rester studieux, je regardais toujours les nouveaux films de science-fiction à l'affiche. Cela m'amusait de voir l'imagination farfelue des réalisateurs. Cette année-là, Iron-man était à l'affiche, avec son cortège d'interactions de **réalité augmentée**. Mais ça ne battait toujours pas le plus improbable de tous les films, Matrix, et son interaction en **réalité virtuelle si immersive** que les utilisateurs ne se sentaient plus **présents** que dans cette réalité digitale, et non plus dans leur réalité physique.*

*Aujourd'hui je n'ai plus de téléphone fixe chez moi. Parfois, je regarde des vidéos sur mon smartphone parce que la 4G marche mieux que ma box de fibre optique. Le soir, je sors parfois avec des amis combattre des aliens lors d'une partie de laser-game en réalité virtuelle. La journée, je prends en main une sphère en polystyrène, j'appuie sur une touche, et la Terre apparaît entre mes mains. Une autre touche, et un papillon resplendissant volette dans ma sphère. C'est drôle : quand je tapote sur sa cage, il s'approche joyeusement. Je joue un peu avec lui, puis il disparaît au profit d'un énorme tas de Kaplas virtuels, bien plus que je n'en ai jamais eu enfant. Je bâtis des structures impressionnantes avec des centaines de Kaplas. Parfois, j'en fais tomber un et il brise une partie de ma structure, me forçant à la reconstruire. Mais je m'en sors bien, c'est dire que je **profite de l'expérience** qui me reste de mon enfance sans internet. Il est temps de passer aux choses sérieuses : j'appuie sur un dernier bouton et un modèle hyper réaliste de jambe humaine s'affiche. Je ne suis pas médecin, mais je sais où s'ancrent les ligaments croisés sur le fémur et le tibia. Ce sont des connaissances en 3D, de notre monde 3D, que j'ai **apprises en 3D**.*

*Toutes ces entités qui s'affichent dans ma sphère, elles sont virtuelles. Enfin, je crois. Mais je n'en suis plus si sûr : **elles paraissent tellement réelles** ! Ainsi, immergé dans ma réalité augmentée, j'imagine les techniques d'interaction farfelues de demain.*

Contexte

Depuis de nombreuses années, nous avons pris l'habitude d'interagir avec du contenu digital en utilisant des interfaces en 2 dimensions (2D) telles que les ordinateurs de bureau (écran, clavier et souris) et les smartphones et tablettes tactiles. Pourtant, comme dans la vie de tous les jours, de nombreuses utilisations nécessitent d'interagir avec des scènes virtuelles 3D. Parmi elles : la visualisation de données 3D, la conception assistée par ordinateur, les simulateurs d'entraînement ou encore les jeux vidéos. Il est possible de visualiser des scènes 3D via des interfaces en 2D grâce à des rendus en perspective et un contrôle du point de vue, mais positionner et orienter des objets dans l'espace peut requérir un contrôle à 6 degrés de liberté (3 en position + 3 en rotation), qui est fastidieux à réaliser avec un dispositif d'entrée à deux dimensions.

C'est pourquoi de nombreux chercheurs et auteurs d'œuvres de science-fiction des dernières décennies imaginent des techniques d'interaction permettant d'interagir avec des scènes virtuelles aussi intuitivement et efficacement que nous interagissons avec les objets de notre réalité physique (c.à.d. le monde réel). Pour ce faire, une première approche consiste à immerger l'utilisateur dans une réalité virtuelle (RV). Ainsi immergé dans un environnement virtuel, l'utilisateur pourra interagir avec les objets virtuels présents autour de lui comme s'il faisait partie de leur réalité. Cette approche tend à couper l'utilisateur de son environnement physique, ce qui peut être problématique dans le cas où l'utilisateur aurait également des besoins d'interaction ou de communication avec son environnement physique (par exemple discuter avec d'autres personnes ou se déplacer dans des lieux précis). Une seconde approche, nommée réalité augmentée (RA), consiste au contraire à apporter l'information virtuelle dans le contexte physique de l'utilisateur.

Ce n'est que récemment que la barrière technologique permettant la réalisation d'appareils de RV et de RA crédibles a été franchie, notamment depuis l'émergence des casques de RV commerciaux produits par Oculus, HTC et Sony en 2016, et des casques de RA tels que le Microsoft HoloLens qui a suivi peu après. Ces appareils sont capables de fournir un rendu visuel stéréoscopique et un rendu auditif 3D, couplés en temps réel à la position de l'utilisateur, de sorte que l'affichage de l'environnement s'adapte en permanence à son point de vue. Ils fournissent également un contrôle 6D des objets virtuels par l'intermédiaire de manettes traquées par le système, ou en analysant les gestes de l'utilisateur. En fin de compte, ces appareils procurent des capacités d'interaction hautement fidèles aux capacités interactives dont nous disposons dans notre réalité physique. Nous parlerons de techniques d'interaction 3D haute-fidélité (HF).

De nos jours, le nombre d'études empiriques ayant permis d'évaluer les bénéfices des systèmes interactifs 3D haute-fidélité est limité. Ce manque de littérature s'explique par les difficultés d'implémentation de prototypes HF causées par les limitations technologiques actuelles, mais également par un manque de méthodologie favorisant la comparaison des différents travaux existants. Des premiers résultats suggèrent que les systèmes HF permettent à l'utilisateur de se sentir présent dans l'environnement virtuel, favorisent le transfert des connaissances acquises dans la réalité physique, et réduisent la charge cognitive générée par l'utilisation de l'appareil interactif. De nombreux défis restent à relever pour comprendre et optimiser l'efficacité des systèmes HF ; en particulier pour identifier et implémenter les applications qui bénéficieraient au mieux de ces systèmes. Ce n'est qu'une fois que ces défis auront été relevés que les systèmes HF pourront être exploités à leur plein potentiel.

Objectif

L'objectif de cette thèse est double : répondre à la question "**pourquoi** utiliser des systèmes HF ?" en apportant des preuves empiriques de leurs bénéfices, et répondre à la question "**comment** implémenter des systèmes HF ?" en contribuant à l'élaboration de lignes directrices qui aident à concevoir les systèmes HF les plus performants. Nous avons cherché à identifier et à mesurer aussi bien les bénéfices fondamentaux de ces systèmes, comme la manipulation d'objets virtuels ; que les bénéfices applicatifs, comme l'apprentissage de formes 3D. Nous avons également cherché à fournir les moyens d'implémenter et d'évaluer les systèmes et les applications qui optimisent ces bénéfices.

Nous espérons que ces travaux contribueront à une meilleure compréhension des avantages que peuvent apporter les systèmes d'interaction HF pour visualiser, manipuler et apprendre avec des objets 3D.

Démarche

Nous avons commencé par étudier l'état de l'art des systèmes haute-fidélité existants et de leurs bénéfices potentiels. En particulier, nous introduisons une approche prometteuse appelée "appareil à perspective corrigée déplaçable" (APCD) : une technique d'interaction HF qui consiste à afficher des objets virtuels "à l'intérieur" d'un appareil manipulable, comme si les objets étaient physiquement présents à l'intérieur d'un conteneur transparent. Nous avons mis en jeu cette technique dans les quatre contributions de cette thèse.

Dans notre recherche de bénéfices, nous nous sommes tout d'abord intéressés à une tâche fondamentale pour interagir avec des scènes 3D : le positionnement d'objets dans l'espace. Cette tâche a été largement explorée dans le cadre des interactions 2D, mais beaucoup moins dans un contexte de positionnement 6D dans l'espace. En particulier, très peu de systèmes HF ont pu être évalués et comparés aux autres systèmes existants. Nous avons donc commencé par élaborer une étude utilisateur allant dans ce sens (première contribution). Cette étude nous a permis de mettre en évidence un bénéfice des systèmes HF lié aux capacités motrices de l'utilisateur. Nous nous sommes alors intéressés à une tâche mettant en jeu un processus cognitif plus complexe : l'apprentissage. Apprendre à reconnaître la forme d'objets 3D semble être un besoin commun à de nombreux enseignements scientifiques, et la littérature semble convaincue du potentiel des systèmes haute-fidélité pour aider dans cette tâche. Mettre en place une étude contrôlée qui permette d'apporter des preuves empiriques des bénéfices d'une technique d'interaction sur une tâche aussi complexe que de l'apprentissage constituait un véritable défi (deuxième contribution).

Ces études nous ont permis de fournir des résultats empiriques sur les bénéfices des systèmes d'interaction HF dans deux tâches mettant en jeu des processus cognitifs très différents. Après avoir montré deux bénéfices robustes de ces systèmes, nous présentons nos contributions pour leur implémentation.

De nos jours, les performances des systèmes HF dépendent des limites des technologies disponibles. Concevoir un système HF idéal nécessite de produire des stimuli dont les performances égalent les capacités de perception humaine, ce qui demande entre autres une latence très faible et de très hautes résolution et fréquence d'affichage. Certaines de ces caractéristiques techniques jouent probablement un rôle plus important que d'autres sur la fidélité du système telle qu'elle est perçue par l'utilisateur. Partant de ce postulat, nous avons mené une étude visant à déterminer quelles caractéristiques matérielles méritent particulièrement d'être priorisées (troisième contribution). Malgré tout, il ne suffit pas de proposer un matériel capable d'afficher fidèlement des objets 3D pour pouvoir produire des applications fonctionnelles ; il est également nécessaire de pouvoir attraper ou relâcher

des objets, zoomer la scène et gérer des paramètres discrets ou continus. Implémenter des solutions intuitives et efficaces pour fournir toutes ces fonctionnalités demande de s'adapter aux contraintes spécifiques du système utilisé. Ce travail n'a pas encore été effectué dans le contexte des APCD. Nous avons donc cherché à proposer des solutions d'implémentation qui permettent de concevoir des applications autonomes pour ces systèmes (quatrième contribution).

Ces deux études nous ont permis de fournir des lignes directrices aussi bien matérielles que logicielles permettant de concevoir des expériences haute-fidélité.

Le projet An@tomy2020

Cette thèse s'intègre dans le cadre du projet ANR An@tomy2020 qui vise à fournir des outils d'apprentissage "incarnés", c'est-à-dire intégrant une expérience corporelle, pour l'enseignement de l'anatomie. Il s'agira par exemple de mobiliser des structures musculaires lors de certains mouvements pour mieux en comprendre les activations. Le projet intègre donc une composante de capture des mouvements de l'apprenant, l'animation d'un modèle sur cette base, la connexion à une base de connaissances anatomiques et des leçons faisant appel à un rendu des informations et à une interaction avec l'utilisateur. Une hypothèse du projet est que cette expérience peut être facilitée en utilisant des systèmes de réalité augmentée et des modalités d'interaction spatiale et/ou mobile. L'apprenant doit en effet pouvoir se créer une bonne représentation mentale en 3D du corps humain. Les théories de la cognition incarnée suggèrent que l'utilisation de systèmes d'interaction qui sont capables de fournir une expérience sensorimotrice "naturelle"[76] lors de l'apprentissage peut favoriser l'acquisition et la mémorisation de ce type de représentation mentale.

Notre apport dans ce projet a été de développer des techniques d'interaction haute-fidélité et de montrer leur potentiel dans différentes tâches en 3D, y compris dans une tâche d'apprentissage. Ces résultats ont contribué au développement d'un prototype d'enseignement incarné de l'anatomie qui doit être testé prochainement avec une cohorte d'étudiants en médecine.

Structure du document

Ce document s'organise en trois parties.

Dans la première partie, nous présentons l'état de l'art. Nous commençons par définir la haute fidélité avant de présenter les systèmes d'interaction HF existants. Nous introduisons les APCD, et nous discutons les avantages et inconvénients de chaque approche (chapitre 1). Nous faisons ensuite le point concernant les bénéfices attendus des systèmes HF en rapport avec nos contributions (chapitre 2).

Dans la seconde partie, nous présentons nos contributions qui apportent des éléments de réponse à la question "**pourquoi** utiliser des systèmes HF ?". Notre première contribution (chapitre 3) consiste en une étude utilisateur testant l'efficacité de deux systèmes d'interaction HF et d'un système basse fidélité dans une tâche de placement d'objets dans l'espace. Notre seconde contribution (chapitre 4) met en jeu les deux mêmes techniques d'interaction HF dans une tâche visant à aider l'utilisateur à apprendre la forme d'un objet 3D. Nous comparons ces techniques avec une technique basse fidélité dans une étude utilisateur.

Dans la troisième partie, nous présentons nos contributions en rapport avec la question : "**comment** implémenter des systèmes HF ?". Notre troisième contribution (chapitre 5) présente et met en jeu une approche permettant de mesurer l'influence de différents paramètres techniques sur le sentiment de présence ressenti par les utilisateurs au

cours d'une expérience interactive HF. Enfin, notre quatrième contribution (chapitre 6) vise à fournir les moyens d'implémenter une application autonome qui bénéficie au mieux des avantages des APCD.

Première partie

État de l'art

Chapitre 1

Des systèmes d'interaction haute-fidélité

Introduction

Dans ce chapitre, nous commençons par définir le concept de fidélité des systèmes d'interaction 3D. Nous détaillons ensuite l'état de l'art des systèmes de réalité virtuelle et de réalité augmentée permettant une interaction haute-fidélité avec des scènes virtuelles. Enfin, nous présentons une nouvelle approche prometteuse, appelée appareils à perspective corrigée déplaçables (APCD), et nous détaillons les performances du prototype d'APCD haute-fidélité que nous avons utilisé dans nos travaux.

Sommaire

Introduction	25
1.1 Définitions	26
1.1.1 Différentes réalités	26
1.1.2 Fidélité d'un système interactif	26
1.1.3 Des réalités fidèles	27
1.1.4 Le sentiment de présence	28
1.2 Les systèmes HF existants	29
1.2.1 Performances reportées et capacités humaines	29
1.2.2 Les systèmes de réalité virtuelle	30
1.2.3 Les systèmes de réalité augmentée	37
1.2.4 Bilan	44
1.3 Un APCD haute-fidélité	45
1.3.1 Les premiers prototypes d'APCD	45
1.3.2 Une implémentation HF	46
1.3.3 Les Avantages des APCD	49
Conclusion	52

1.1 Définitions

1.1.1 Différentes réalités

Dans cette thèse, nous définissons une **réalité** (physique ou virtuelle) comme un environnement autonome complet, régi par des lois fondamentales, ainsi que tous les objets et créatures sociales qui évoluent et interagissent les uns avec les autres dans cet environnement. Nous utilisons le terme **entités** pour regrouper toutes sortes d’acteurs évoluant dans une réalité (c’est-à-dire des objets, des êtres sociaux ou des individus). En tant qu’êtres humains, nous évoluons dans la réalité physique. Afin d’interagir avec d’autres entités physiques de cette réalité, notre cerveau analyse les données perçues par notre système sensoriel en utilisant une connaissance préalable des lois fondamentales de cette même réalité. Toutes les expériences d’interaction que nous avons eues au cours de notre vie nous permettent de prédire le comportement d’entités que nous n’avons jamais vues auparavant, d’autant plus si l’on voit quelqu’un d’autre appréhender l’entité, ou si on nous la décrit. De la texture, des reflets et de la forme d’un objet, nous pouvons déduire sa matière, son poids et sa fonction.

Il y a plus de cinquante ans [145], la recherche dans le domaine de la réalité virtuelle a commencé à aborder la création d’autres réalités, numériques, avec leurs propres entités, règles et métaphores d’interaction. En s’immergeant dans une réalité virtuelle (RV), les utilisateurs peuvent interagir avec des entités et des environnements qui n’existent pas ou ne peuvent pas exister dans la réalité physique. La qualité de ces interactions dépend notamment de la fidélité du système.

1.1.2 Fidélité d’un système interactif

Le concept de *fidélité* d’un simulateur a été introduit pour la première fois en 1969 par Gerathewohl comme le degré de précision avec lequel le simulateur reproduit les sensations expérimentées dans la situation réelle [56]. Dans le domaine de la réalité virtuelle (RV), ce concept a été relié au concept d’*immersion*. Bowman et McMahan définissent l’immersion comme “le degré objectif de fidélité sensorielle que fournit un système de RV” (“the objective level of sensory fidelity a VR system provides”) [21]. De son côté, Slater définit l’immersion d’un système comme le degré avec lequel les informations sensorielles capturées et affichées par le système en réponse à une expérience virtuelle sont fidèles aux informations sensorielles expérimentées dans le monde réel au cours d’une expérience équivalente [133].

Dans cette thèse, nous ferons référence à la **fidélité** d’un système interactif comme le degré de similarité entre les capacités d’interaction offertes à l’utilisateur par le système et les capacités d’interaction que l’utilisateur possède dans sa réalité physique. En particulier on qualifiera un système de **haute-fidélité (HF)** s’il permet de percevoir ou de contrôler une entité virtuelle d’une manière comparable à la manière dont l’utilisateur percevrait ou contrôlerait une entité similaire dans sa réalité physique. Un tel système permet donc une manipulation très *naturelle* des entités virtuelles au sens de Bérard et Rochet-Capellan, et favorise des transferts de compétences sensori-motrices entre la réalité physique et la réalité virtuelle [27].

Il est à noter que la notion de fidélité peut dépendre de la tâche envisagée. Dans une tâche d’assemblage qui vise à entraîner un technicien sur une manipulation donnée, il semble primordial que le retour tactile ainsi que les forces physiques jouant sur l’assemblage (poids, forces de friction, ...) soient fidèlement reproduits, tandis que les détails visuels fins ne seront que très peu pertinents. A l’inverse, dans une tâche d’observation d’entités hors de portée de l’utilisateur, l’absence de rendu tactile ne sera pas préjudiciable, et un rendu visuel de qualité sera indispensable. Idéalement, un système HF

devrait pouvoir produire des stimuli fidèles pour tous les canaux sensoriels humains, mais certains efforts de rendu peuvent être évités en fonction du contexte d'utilisation sans porter préjudice à la fidélité perçue du système.

Prérequis pour une perception visuelle fidèle

La perception visuelle est le sens le plus utile et le plus exploité en interaction Homme-machine. Permettre une perception visuelle fidèle semble donc être primordial pour les systèmes HF, même si ce n'est pas suffisant pour produire une expérience parfaitement immersive. Dans la réalité physique, nous sommes capables de contrôler notre point de vue en temps réel grâce aux mouvements de notre tête. Le changement de point de vue est en permanence couplé à l'activation des muscles mis en jeu et à un retour instantané cohérent d'informations fournies par notre proprioception, notre oreille interne, et nos yeux. Ce couplage permanent nous permet de nous situer mentalement dans notre environnement. Conserver ce couplage semble indispensable pour percevoir fidèlement un environnement virtuel. Ainsi, un système d'interaction HF doit permettre à l'utilisateur de contrôler son point de vue de manière cohérente en bougeant librement la tête. Dans cette thèse, nous ferons référence à cette fonctionnalité en parlant d'**affichage couplé à la position de l'utilisateur**.

De plus, nos deux yeux nous fournissent une vision stéréoscopique nous permettant une perception de profondeur particulièrement utile pour interagir avec les entités proches de nous. Fournir un affichage stéréoscopique semble être un second prérequis pour pouvoir qualifier un système de haute-fidélité.

Compte tenu de ces prérequis, un ordinateur de bureau classique ne peut donc pas être considéré comme un système HF, à moins d'offrir un affichage stéréoscopique, et de l'utiliser avec un système de capture de mouvements permettant de coupler l'affichage à la position de l'utilisateur.

Prérequis pour un contrôle fidèle

Nous utilisons principalement nos mains pour contrôler les entités qui nous entourent. Notre compréhension du comportement physique de notre réalité nous permet d'effectuer instinctivement et sans y prêter attention une séquence de mouvements permettant de déplacer et d'orienter des objets en fonction de nos besoins. Un système HF doit permettre de manipuler des objets sans ajouter d'indirection entre l'intention de l'utilisateur et les mouvements qu'il doit mettre en place pour y parvenir. Pour cela, les périphériques d'entrée utilisés doivent permettre un contrôle d'objets *isomorphe* : le mouvement appliqué par l'utilisateur au périphérique d'entrée est reproduit "tel quel" sur l'objet virtuel sélectionné. On parlera dans cette thèse de **périphérique de contrôle isomorphe**.

Ainsi, utilisée sans gain dynamique, une souris d'ordinateur classique pourrait être considérée comme un périphérique de contrôle isomorphe pour un mouvement contraint dans un plan 2D. En revanche, ce périphérique ne permet pas un contrôle isomorphe d'objets dans l'espace 3D.

1.1.3 Des réalités fidèles

Avec des systèmes de RV haute-fidélité qui reproduisent les lois, les stimuli visuels et les potentialités interactives de la réalité physique, les utilisateurs peuvent être immergés dans des réalités où ils peuvent interagir intuitivement avec des entités virtuelles, grâce à l'expérience emmagasinée au cours de leur vie. Cependant, lorsque l'environnement physique convient aux besoins de l'utilisateur, il peut sembler inutile de le reproduire dans une réalité virtuelle dans le seul but d'y ajouter une entité non physique. Il peut

être plus pertinent de créer une réalité augmentée (RA) où seule l'entité virtuelle est immergée dans le monde physique. La RV et la RA partagent le même objectif d'offrir une interaction intuitive en exploitant un transfert des compétences et des connaissances acquises dans la réalité physique.

Idéalement, avec une réalité virtuelle ou une réalité augmentée très haute fidélité, les utilisateurs ne seront pas en mesure de dire si l'entité avec laquelle ils interagissent est physique ou virtuelle. Lorsque cela se produit, l'utilisateur ressent l'entité virtuelle comme entièrement *présente* dans sa réalité : plus une entité est présente, plus il est difficile pour l'utilisateur d'identifier si l'entité est virtuelle ou physique.

1.1.4 Le sentiment de présence

Slater décrit la présence comme le sentiment éprouvé par l'utilisateur en réponse à une exposition à un système de RV suffisamment immersif. Il ajoute qu'un moyen d'éprouver un sentiment de présence est de construire un système qui a un tel niveau de fidélité envers la réalité qu'il devient indistinguable de la réalité elle-même. On retrouve des concepts similaires dans différentes thématiques de recherche allant de la télé-coopération à la RA : présence physique, présence sociale, coprésence et télé-présence. Ces différentes définitions, que nous détaillerons dans la section 2.3.1, tendent à subdiviser le concept de présence en fonction des besoins spécifiques de chaque thématique. Pourtant, elles décrivent toutes un sentiment psychologique d'appartenir à la même réalité qu'une ou plusieurs entités.

Lee [89] a proposé une définition de la présence qui unifie les approches de RV et de RA : il définit la présence comme "a psychological state in which virtual (para-authentic or artificial) objects/social actors/selves are experienced as actual objects in either sensory or nonsensory ways", que nous pouvons traduire par : un état psychologique dans lequel les objets virtuels, les acteurs sociaux virtuels et les représentations virtuelles de soi-même sont ressentis par l'utilisateur comme des objets/acteurs/soi réels, que ce soit via les canaux sensoriels ou non sensoriels. "Virtuels" renvoie ici à "para-authentiques" et "artificiels". Nous reprenons cette définition, et utilisons le terme "entités" pour désigner tous les acteurs possibles, des objets inanimés au soi. Nous reconnaissons également que la présence est une métrique continue plutôt qu'une valeur binaire, comme cela est évalué dans de nombreuses études [138, 118, 35], et que la présence dépend de la tâche. Par exemple, un objet virtuel représentant un lourd cube métallique peut sembler présent tant que les utilisateurs n'interagissent pas avec, mais la présence peut rapidement être brisée si l'objet ne pèse rien dans les mains de l'utilisateur. En résumé, nous définissons la **présence** comme une métrique continue de l'état psychologique correspondant au degré auquel les entités virtuelles sont ressenties par l'utilisateur comme des entités réelles pour réaliser une tâche spécifique, que ce soit via les canaux sensoriels ou non sensoriels.

Reconnaissant l'absence de consensus sur la définition de la présence et de ses concepts associés, nous notons que notre définition peut également être apparentée à la *plausibilité* des entités virtuelles [135].

1.2 Les systèmes HF existants

Dans cette section, nous présentons l'état de l'art des systèmes haute-fidélité existant en 2020. Nous présentons les différentes approches qui ont été mises en œuvre en RV et en RA. Pour chaque approche, nous présentons un historique rapide de l'évolution des technologies et des travaux qui ont permis la conception de systèmes de plus en plus fidèles. Nous nous attardons sur les différents aspects du rendu visuel, qui a été le plus traité dans la littérature et qui nécessite encore beaucoup de travail d'optimisation. Nous mettons en valeur les avantages et les inconvénients de chaque approche, et nous présentons les performances actuelles des systèmes les plus récents.

1.2.1 Performances reportées et capacités humaines

Dans les sections qui suivent, nous rapportons à titre comparatif les performances techniques des systèmes présentés concernant certains paramètres primordiaux pour la réalisation d'interaction et de rendu visuel HF. Nous présentons ici la liste de ces paramètres en précisant les limites des capacités humaines pour chacun d'entre eux, afin de mettre en contexte les chiffres qui suivent. Cette liste n'est pas exhaustive, mais présente les paramètres qui ont été les plus étudiés dans la littérature.

Précision de capture

Nous désignons par précision de capture la plus petite unité de déplacement que le système peut mesurer de manière stable (non bruitée). Les capteurs utilisés par les systèmes de suivi optique ou électromagnétique produisent souvent un signal bruité par un bruit blanc Gaussien. Nous considérons alors la précision de capture comme le pas de mesure le plus petit permettant de couvrir le bruit de mesure. Bérard et al. ont mesuré une précision cinématique des utilisateurs avec le doigt de 0.17 mm sur une surface tactile [11]. La précision de capture doit être optimisée afin de pouvoir traiter les mouvements exacts de l'utilisateur sans qu'ils soient détériorés.

Latence

La latence de bout en bout d'un système est le délai entre le moment où l'utilisateur effectue une action et le moment où le système fournit le retour sensoriel correspondant. Ce paramètre doit être optimisé afin que l'utilisateur ait le sentiment que les retours qu'il observe sont effectivement une réponse à ses actions, et ce, quel que soit le canal sensoriel considéré. Concernant le retour visuel lors d'une action de toucher direct, Ng et al. ont mesuré une capacité de détection moyenne de l'ordre de 6 ms pour les utilisateurs qu'ils ont testés [107].

Résolution d'affichage

La résolution d'affichage constitue le nombre de points de couleurs différents que le système est capable d'afficher simultanément. Cette résolution est souvent ramenée en nombre de points par pouce (dpi : "dots per inches"), afin de comparer des affichages de différentes tailles. L'œil humain est, quant à lui, basé sur une vision angulaire, ce qui signifie que plus une image est proche de l'œil, plus elle couvre un large angle de vision, et plus ses détails sont discernables. La résolution angulaire maximum de l'œil humain est de l'ordre de 0.017° , ce qui correspond à une résolution de 174 dpi à une distance de 50 cm, et à une résolution de 44 dpi à une distance de 2m.

Champ de vision

Le champ de vision désigne la portion angulaire de l'espace visible par un individu restant immobile. Pour les êtres humains, le champ de vision horizontal binoculaire est de l'ordre de 120° , et il dépasse les 190° en incluant la vision monoculaire périphérique. Ce paramètre doit être optimisé afin de procurer un rendu visuel immersif.

Fréquence d'affichage

Elle désigne le nombre d'images différentes que le système est capable d'afficher en une seconde. Plusieurs études ont mesuré que l'œil humain est capable de percevoir des fréquences d'affichage maximum allant jusqu'à 90 Hz, lors de l'observation d'une source stable, mais est capable de détecter des artefacts visuels à une fréquence de 500 Hz grâce au mécanisme inconscient des saccades oculaires [44]. Ce paramètre doit être optimisé pour générer un rendu stable.

1.2.2 Les systèmes de réalité virtuelle

Les dispositifs de réalité virtuelle sont souvent regroupés en trois catégories principales en fonction de leur technologie d'affichage : les CAVE (Cave Automatic Virtual Environment), les aquariums de RV (Fish-tank VR), et les casques de RV. Les CAVE permettent aux utilisateurs de se déplacer naturellement dans des volumes immersifs. Les aquariums de RV permettent d'interagir avec l'environnement virtuel contenu dans l'aquarium comme s'il était physiquement contenu dans une boîte transparente. Les casques permettent une immersion totale à moindre coût.

Les CAVE

Les CAVE ont pour la première fois été introduites en 1992 par Cruz-Neira et al. [41]. L'idée générale est de rétro-projeter des images sur un maximum de faces d'une salle dans laquelle l'utilisateur peut se déplacer librement, comme l'illustre la figure 1.1. Chaque face se transforme ainsi en un écran géant sans bord, et en affichant sur trois murs et le sol, il est déjà possible de couvrir la totalité du champ de vision de l'utilisateur quand il fait face au mur du fond. Les images à afficher par chaque projecteur sont recalculées en temps réel en fonction de la position de l'utilisateur par rapport aux plans de projection (les faces de la salle). Il est important de rétro-projeter les images afin d'éviter les ombres causées par l'utilisateur. En utilisant des projecteurs stéréoscopiques, ce système permet d'immerger l'utilisateur dans un environnement virtuel dans lequel des entités peuvent être affichées à l'intérieur comme à l'extérieur de l'espace disponible dans la salle physique. En effet l'œil humain est capable de converger à une distance différente de la distance d'accommodation. De cette manière les murs physiques disparaissent complètement lorsque l'utilisateur focalise sur les entités virtuelles affichées, permettant un très haut niveau d'immersion. Malgré tout, la littérature déconseille d'afficher des entités virtuelles trop loin des murs physiques, que ce soit vers l'intérieur ou l'extérieur de la pièce, pour limiter le conflit accommodation-divergence, qui a tendance à causer un inconfort visuel et à fausser l'estimation des distances, pouvant provoquer des collisions de l'utilisateur avec le mur physique [24].

Un avantage important de ce dispositif est de permettre à l'utilisateur de se déplacer naturellement dans l'environnement virtuel (tant qu'il reste dans les limites physiques de la pièce évidemment), tout en conservant la vision de ses membres, ce qui favorise le sentiment d'être présent dans la scène virtuelle. Tant que les entités affichées sont hors

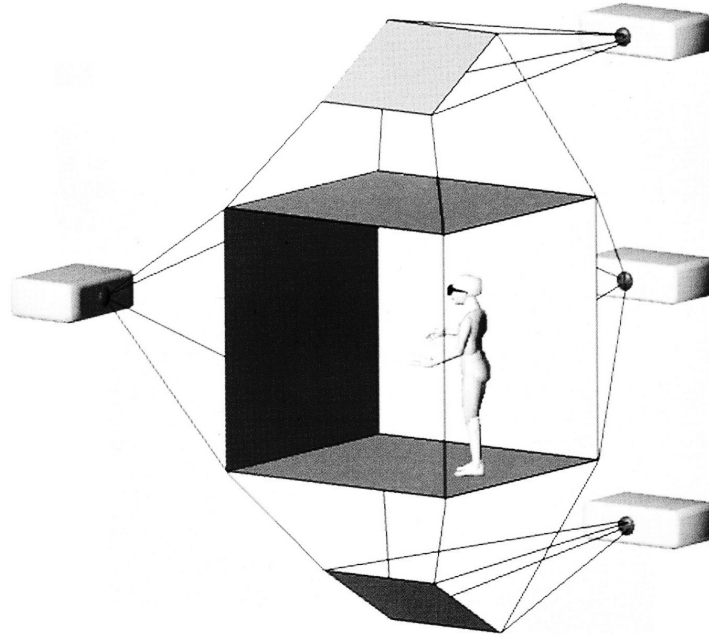


FIGURE 1.1: Schéma proposé par Cruz et al. pour illustrer la rétro-projection dans une CAVE à quatre faces (source : [41]).

de portée, les membres de l'utilisateur s'intègrent parfaitement dans la scène virtuelle. En revanche, si l'utilisateur essaie d'attraper une entité avec ses mains, les bras cacheront les parties de l'entité qui ne devraient pas être cachées. Ceci, ajouté à l'absence de tangibilité des entités virtuelles, brise l'illusion de présence. Pour pallier ce manque de tangibilité, Pontonnier et al. ont testé l'utilisation d'un bras articulé à retour de force Virtuose de Haption dans une tâche d'assemblage [114]. Le bras utilisé imposait aux utilisateurs de manipuler des objets via une poignée physique (voir figure 1.2, à gauche), ce qui a complexifié la tâche d'assemblage et a été subjectivement jugé peu fidèle par les participants. Une autre approche consiste à utiliser des gants avec pseudo retour de force simulé par des vibrations ciblées et/ou des impulsions piézoélectriques. Sorgini et al. ont montré la

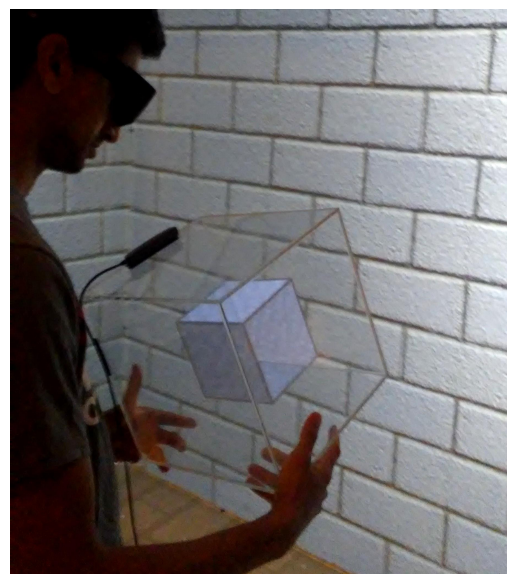
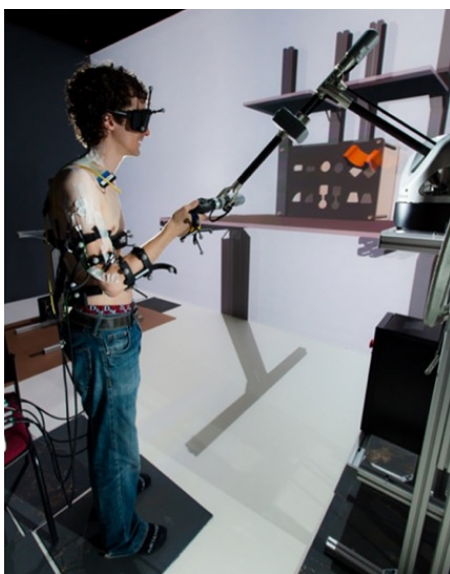


FIGURE 1.2: À gauche : Utilisation d'un bras articulé par Pontonnier et al. pour ajouter un retour haptique dans une CAVE (source : [114]). À droite : utilisation d'un cube tangible par Zielinski et al. pour manipuler des entités virtuelles dans une CAVE (source : [171]).

faisabilité de tels gants et ont proposé, sans la tester, leur utilisation dans le contexte des CAVE [139]. Zielinski et al. ont proposé d'utiliser un cube tangible transparent pour manipuler des entités virtuelles dans une CAVE [171]. Ce système est illustré dans la figure 1.2, à droite. Ils ont testé l'utilisation du cube par rapport à l'utilisation de deux contrôleurs à main levée dans une tâche de rotation 3D. Le cube tangible s'est avéré plus rapide et plus précis que les contrôleurs pour tourner des objets virtuels. Les utilisateurs ont néanmoins rapporté que le poids du cube, plus de 2 kg, rendait son utilisation inconfortable. Le cube était également câblé à l'utilisateur.

L'inconvénient principal de ces systèmes est qu'ils sont difficiles à déployer, compte tenu de la surface au sol importante qui doit leur être allouée – celle-ci correspond à peu près au double de la surface utilisable pour permettre la rétro-projection-, et de leur coût très important qui peut atteindre plusieurs millions d'euros. Ainsi, ces systèmes ont été principalement déployés dans des contextes industriels massifs tels que l'automobile ou l'architecture. Dans ces contextes, les CAVE permettent une visualisation réaliste du rendu 3D final d'un prototype, à l'échelle 1 :1, tout en évitant des coûts de déploiement importants.

Performances techniques :

Le premier système de CAVE de Cruz-Neira et al. permettait d'interagir dans une salle de trois mètres de côté, dont trois murs ainsi que le sol sont utilisés pour la projection. Pour chaque face, un projecteur de résolution 1280x512 pixels (10.8 dpi) produisait un affichage stéréoscopique à 120 Hz (60 Hz par œil). Le système de capture de mouvements électromagnétiques permettait une précision de l'ordre de quelques millimètres mais la latence totale du système devait dépasser les 250 ms d'après les estimations de Brooks [23].

L'un des systèmes récents les plus performants est la CAVE à quatre faces Immersia à Rennes, illustrée dans la figure 1.3. Cette CAVE propose une salle immersive de 9.6m par 3m et de 3.1m de hauteur. L'affichage stéréoscopique à 120 Hz se fait à l'aide de quatorze projecteurs de résolution 2560x1600 pixels (21 dpi). Le système de capture de mouvement optique offre une précision de l'ordre du dixième de millimètre. La latence totale du système n'a pas été documentée mais est probablement inférieure à 60 ms compte tenu des technologies utilisées.

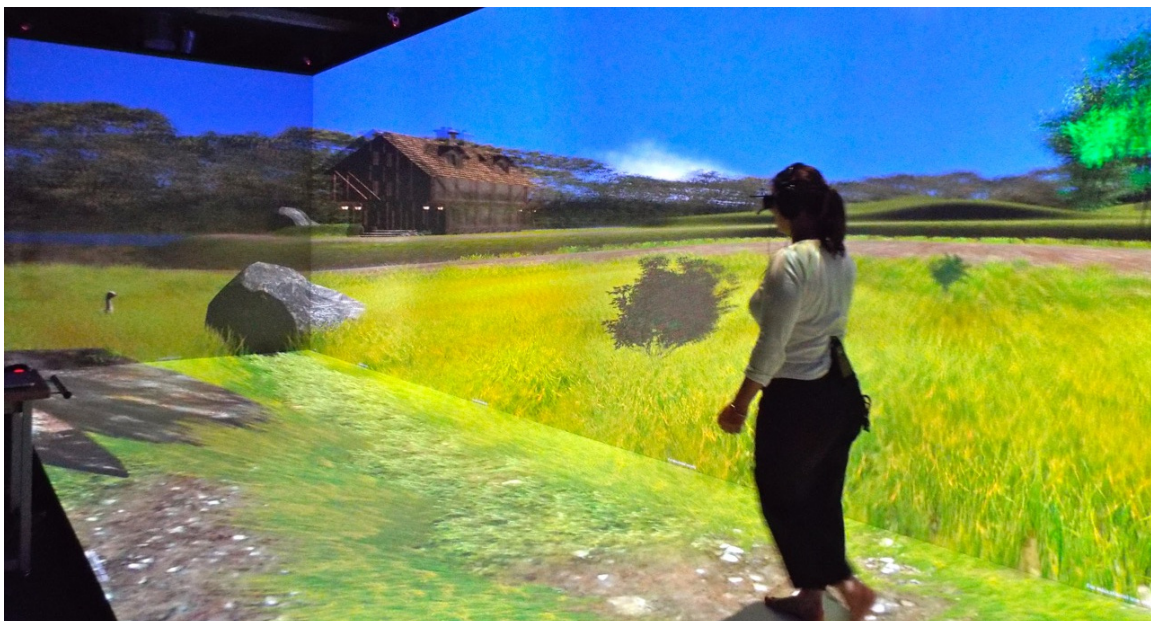


FIGURE 1.3: La CAVE Immersia à Rennes, de près de 10m de long.

Les aquariums de RV

Le concept d'aquarium de RV est apparu pour la première fois en 1992, lorsque Deering a montré qu'il était possible de fournir un affichage stéréoscopique couplé à la position de l'utilisateur avec un écran d'ordinateur de bureau [45]. À l'époque des écrans cathodiques, les écrans eux-mêmes occupaient un espace volumique de la taille d'un aquarium. Arthur et al. se sont rendu compte qu'avec cette technologie, il était possible de créer l'illusion de la présence d'objets 3D virtuels à l'intérieur du volume de l'écran. Ils ont montré que les aquariums de RV permettaient d'améliorer les performances de perception d'objets 3D par rapport à l'utilisation standard de l'écran [4]. Leur système est illustré dans la figure 1.4, à gauche. Ils ont également montré que les indices de profondeur apportés par le couplage à la position de l'utilisateur étaient plus utiles que les indices apportés par la stéréoscopie dans une tâche de suivi de graphes 3D.

Le système de Bimber et al. est l'un des premiers à utiliser des miroirs pour réaliser un aquarium de RV dont le contenu virtuel peut être observé à 360° et non simplement à travers une seule face [18]. Il est illustré dans la figure 1.4, à droite.

Grossman et Balakrishnan ont comparé deux aquariums de RV dans une tâche de perception de graphes 3D. L'une de leurs conditions utilise un écran cathodique (une seule face affichable), tandis que l'autre condition utilise un écran volumétrique sphérique dont l'affichage 3D offre passivement un rendu stéréoscopique couplé à la position des utilisateurs [60]. Cette approche, illustrée dans la figure 1.5 à gauche, permet l'observation du contenu virtuel de l'aquarium à 360°, par plusieurs utilisateurs simultanément, et sans nécessiter de porter des lunettes spécifiques. Cependant, ce dispositif offre une qualité d'affichage limitée (198 images de 768x768 pixels affichées à 24 Hz) pour un coût nettement supérieur à l'approche avec écran de bureau standard. Bien que les auteurs aient mesuré une perception de profondeur comparable avec les deux conditions, l'écran d'ordinateur a obtenu de meilleures performances que l'écran volumétrique dans la tâche de perception de graphes 3D. Les auteurs estiment que la différence de performance est due aux limitations techniques de leur prototype d'écran volumétrique, en termes de résolution et de fréquence d'affichage.

Néanmoins, la littérature s'est penchée sur l'utilisation d'aquariums de RV sphériques plutôt que cubiques, tels que le Snowglobe de Bolton et al. [19]. L'utilisation d'une sphère permet d'éviter la présence des arêtes et des coins, qui produisent des discontinuités dans l'affichage, perturbant l'illusion de présence des objets virtuels à l'intérieur de l'aquarium. L'absence de bord opaque évite également toute occultation de contenu. Pour arriver à

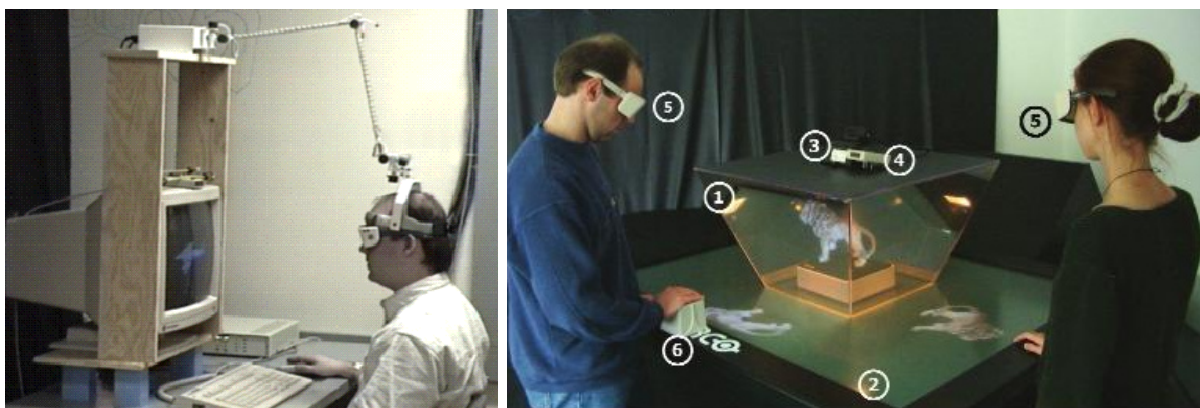


FIGURE 1.4: À gauche : Le système d'aquarium de RV développé par Arthur et al. (source : [4]). À droite : Le système d'aquarium de RV à 360° développé par Bimber et al. (source : [18]).

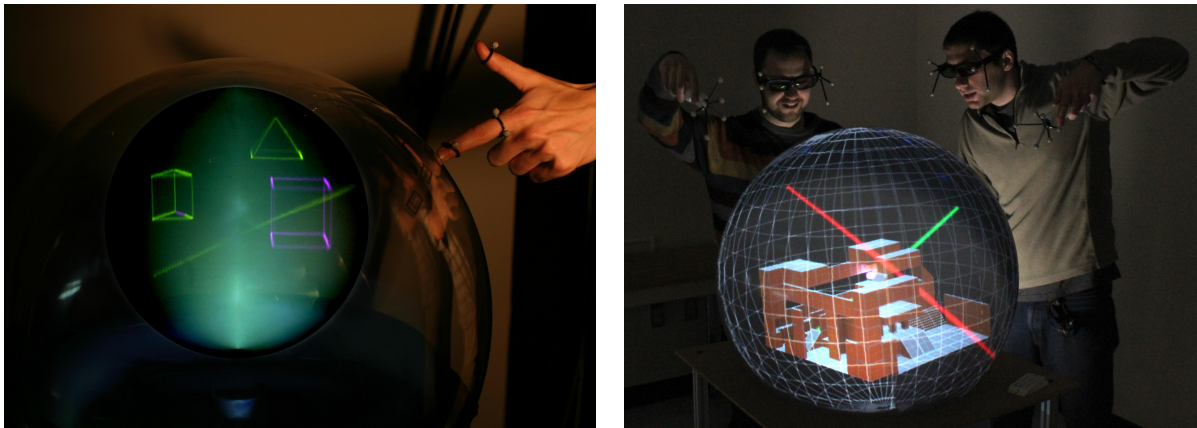


FIGURE 1.5: À gauche : L'écran volumétrique utilisé par Grossman et al. (source : [59]). À droite : le système d'aquarium de RV pour deux utilisateurs de Zhou et al. (source : [170]).

produire un affichage sphérique, Bolton et al. projettent sur un miroir hémisphérique une image déformée de manière adaptée pour produire l'image finale correcte. Bien qu'offrant un affichage couplé à la position de l'utilisateur, le Snowglobe, comme de nombreux autres prototypes exploratoires des dix dernières années, ne permet pas un affichage stéréoscopique. Un prototype très récent de Zhou et al. utilise quatre projecteurs stéréoscopiques pour pouvoir afficher sur toute la surface d'une sphère de 60 cm de diamètre sans angle mort [170]. Plutôt que de multiplexer les images adaptées à chaque œil de l'utilisateur pour permettre un affichage stéréoscopique, le système multiplexe les images adaptées à la position de deux utilisateurs, réalisant deux affichages couplés à la position des utilisateurs simultanément. Ce prototype est illustré dans la figure 1.5.

Si la stéréoscopie est moins importante que le couplage de l'affichage à la position de l'utilisateur pour percevoir la profondeur, elle semble utile pour améliorer les performances d'interaction avec du contenu 3D virtuel. Fafard et al. ont en effet montré que l'absence de rendu stéréoscopique avec un aquarium de RV pouvait augmenter de 20% à 91% le temps nécessaire pour percevoir et interagir dans des tâches de visualisation et de sélection avec des nuages de points [51].

D'après Fafard et al., l'un des avantages majeurs des aquariums de RV est qu'ils sont situés à l'intérieur de la réalité physique, permettant à l'utilisateur des transitions fluides entre les informations de sa réalité physique et les informations de la RV. De plus, le système offre une frontière tangible entre les deux réalités qui renforce l'illusion de présence des entités virtuelles à l'intérieur de l'aquarium physique. En contrepartie, cette frontière empêche les utilisateurs de plonger leurs mains ou un quelconque contrôleur à l'intérieur de l'aquarium pour aller saisir directement des objets. Par conséquent, ces systèmes doivent proposer une technique de sélection à distance. Grossman et Balakrishnan ont testé plusieurs techniques de sélection d'objets à l'intérieur de leur affichage volumétrique sphérique de 25 cm de diamètre [61]. Ils ont montré la supériorité d'une technique de lancer de rayons par rapport à un curseur ponctuel 3D.

Les aquariums de RV constituent l'approche la plus à même d'offrir un affichage très haute résolution pour un coût limité, grâce aux avancées technologiques récentes permettant le développement de systèmes de capture de mouvements très précis, et d'afficheurs très haute définition. Cette approche est particulièrement adaptée pour l'exploration de modèles 3D et la visualisation de données volumétriques telles que celles des scanners médicaux.

Performances techniques :

Le prototype d'Arthur et al. offrait déjà un affichage de résolution 960x680 pixels à 120 Hz, et la précision du système de capture était de l'ordre de 5 mm. La plus grosse limitation était donc la latence, comparable à celle des anciennes CAVE : plus de 250 ms.

Aujourd'hui, il existe des écrans plats et des projecteurs grand public permettant des affichages 4K (3840x2160 pixels) à 120 Hz. Utilisés avec un bon système de capture optique et une carte graphique récente, ils permettent de proposer un aquarium de RV à une face avec une précision de capture inférieure au dixième de millimètre et une latence inférieure à 50 ms. En effet, de nombreux écrans conçus pour le marché du jeu vidéo offrent des latences d'affichage de l'ordre de 15 ms ou moins. De telles performances ne sont pas encore accessibles pour des aquariums de RV sphériques produisant un affichage stéréoscopique. Les meilleures solutions techniques actuelles consistent à projeter des images correctement distordues sur un écran sphérique, mais cela a pour résultat une qualité d'affichage inégale en fonction de l'angle d'incidence du projecteur et de la distorsion appliquée.

Les casques de RV

Bien qu'imaginés par Sutherland et par de nombreuses œuvres de science-fiction dès les débuts de l'histoire de la réalité virtuelle dans les années soixante [145], les casques de RV n'étaient pas exploitables pour un usage réel jusqu'aux années 2000 à cause de l'insuffisance des performances technologiques. Les quelques prototypes réalisés dans les années 90 offraient des résolutions d'affichage maximum de 320x480 pixels par œil et des couleurs en 16-bits, ainsi que des champs de vision ne dépassant pas les 45° [23]. Mais le principal problème était le manque de précision et de stabilité des capteurs de mouvement électromagnétiques : avec un casque de RV, la moindre erreur ou instabilité dans le calcul de la position de l'utilisateur a des impacts énormes sur l'affichage, et peut provoquer nausées et maux de tête. Ces inconforts ont été fréquemment rapportés [129] et constituaient une véritable limitation pour le déploiement de cette technologie. La majorité des expériences mises en jeu à cette époque ne parvenaient pas à identifier empiriquement les bénéfices des casques de RV. Patrick et al. ont par exemple montré que l'utilisation d'un grand écran et d'un projecteur remplaçait avantageusement l'utilisation d'un casque de RV dans une tâche d'exploration d'environnement virtuel [111]. Le casque de RV qu'elles ont utilisé, incluant le dispositif de capture électromagnétique, est illustré dans la figure 1.6, à gauche. Qi et al. ont montré que l'utilisation d'un aquarium de RV était plus performante que l'utilisation d'un casque dans une tâche de visualisation de nuages de points 3D [119].



FIGURE 1.6: À gauche : le casque de RV testé par Patrick et al., utilisant un Visette Pro pour l'affichage et un Ascention SpacePad pour la capture de position (source : [111]). Au milieu : le Vive de HTC sorti en 2016, avec ses contrôleurs isomorphes et les bases permettant la capture optique. À droite : la sphère tangible utilisée par Englmeier et al. (source : [50]).

C'est seulement en 2016 que ces systèmes connaissent un renouveau avec la sortie de la nouvelle génération de casques de RV incluant l'occulus Rift, et le HTC Vive illustré dans la figure 1.6, au milieu. Il aura fallu plus de quinze ans pour que les avancées technologiques en termes de capture optique, de miniaturisation d'écrans LCD et de puissance de calcul des cartes graphiques permettent le développement de casques suffisamment fidèles et précis pour un déploiement en contexte écologique. Leur coût financier abordable (inférieur à 1000 euros), a permis une commercialisation massive, et par la même occasion, un regain d'attention pour la recherche dans le domaine de la réalité virtuelle. De nos jours, ces systèmes sont principalement déployés dans l'industrie du divertissement, où l'immersion des utilisateurs procure un gain important. L'industrie du jeu vidéo développe à la fois des solutions logicielles telle que l'API SteamVR de Valve (facilitant le déploiement de jeux en RV), et des solutions matérielles telles que le playstation VR de Sony. Les casques de RV sont aussi utilisés pour déployer des versions virtuelles de laser games ou de jeux d'évasion grandeur nature, permettant l'intégration de contenus surréalistes, et la démultiplication des scénarios proposés dans un même lieu physique. Ces systèmes sont également déployés dans le cadre de simulateurs d'entraînement, tels que des simulateurs de conduite ou de pilotage.

Les casques de RV immergent totalement les utilisateurs dans un environnement virtuel. Pour ce faire, les utilisateurs doivent être complètement isolés de leur réalité physique, y compris de la vue de leurs propres membres. Cette absence de retour visuel a des effets qui peuvent cependant être négatifs sur la présence, les performances et la charge cognitive de l'interaction. Ces effets peuvent être compensés grâce à l'intégration d'un avatar virtuel reproduisant les mouvements du corps de l'utilisateur [143]. Aussi, ces systèmes nécessitent le port d'un dispositif encombrant sur la tête et un affichage très proche des yeux. Ils sont donc mentalement et musculairement fatigants à utiliser, surtout lors de longues sessions d'utilisation (plus d'une heure). Un autre inconvénient des casques de RV est, comme pour les CAVE, le manque de tangibilité lors de l'interaction avec les entités virtuelles, qui a tendance à briser le sentiment de présence de l'utilisateur. Pour dépasser cette limitation, Englmeier et al. proposent l'utilisation d'un accessoire tangible pour interagir avec des objets dans un environnement de RV immersif [50]. Ils se sont inspirés des aquariums sphériques et ont démontré que l'utilisation d'une sphère tangible améliore l'évaluation analytique d'une visualisation sphérique. Ils ont testé deux sphères, de 25 cm et de 40 cm de diamètre respectivement. Les sphères contiennent un contrôleur permettant de connaître leur position dans l'espace et de faire le lien entre l'espace physique et



FIGURE 1.7: À gauche : l'occulus Quest, les caméras permettent la capture de mouvement sans nécessiter de dispositif externe. À droite : le Xtal de Vrgineers, offrant une résolution 5K et un champ de vision de 180°.

l'espace virtuel. La large sphère est illustrée dans la figure 1.6, à droite.

Avec des progrès technologiques constants, de nouveaux casques offrent un rendu de plus en plus fidèle, et sont accompagnés de périphériques d'entrée permettant un contrôle isomorphe des entités virtuelles. Des solutions autonomes, légères et sans fil apparaissent, telles que l'oculus Quest illustré dans la figure 1.7, à gauche. Ces dispositifs analysent en temps réel les images fournies par des caméras embarquées dans le casque afin de calculer la position de celui-ci par rapport à la position de points d'intérêt fixes dans la pièce (comme des motifs avec de forts contrastes). Avec cette approche, il n'est pas utile d'équiper la pièce de capteurs externes.

Performances techniques :

Les casques de RV les plus récents offrent une résolution d'affichage de 1600x1400 pixels par œil, une fréquence de rafraîchissement comprise entre 70 Hz et 90 Hz, et un champ de vision de 110°. La latence totale de ces systèmes est inférieure à 40 ms et peut descendre en dessous de 30 ms avec une carte graphique assez puissante. La précision du système de capture est de l'ordre de quelques dixièmes de millimètres. Ces performances continuent d'évoluer très vite. Le Xtal de Vrgineers, disponible en 2020, offre une résolution d'affichage de 2560x1440 pixels par œil pour un champ de vision de 180°, mais est plus lourd et encombrant que ses concurrents. Le Xtal est illustré dans la figure 1.7, à droite.

1.2.3 Les systèmes de réalité augmentée

La plupart des systèmes de RA peuvent être regroupés en trois catégories : la RA spatialisée, la RA mobile et les casques de RA. La RA spatialisée consiste à afficher des informations virtuelles directement dans l'espace de travail physique de l'utilisateur, souvent à l'aide d'un projecteur. La RA mobile consiste à observer une augmentation de son environnement au travers d'un appareil mobile tel qu'un smartphone. Cette approche est particulièrement répandue. Les casques de RA fonctionnent comme les casques de RV mais l'écran utilisé est semi-transparent, permettant une superposition de l'information virtuelle avec l'environnement physique de l'utilisateur.

RA spatialisée

L'un des premiers systèmes de RA imaginé est le DigitalDesk de Wellner (illustré dans la figure 1.8, à gauche), qui permet à l'utilisateur d'interagir naturellement avec des contenus virtuels 2D projetés sur un bureau physique, grâce à l'utilisation d'outils

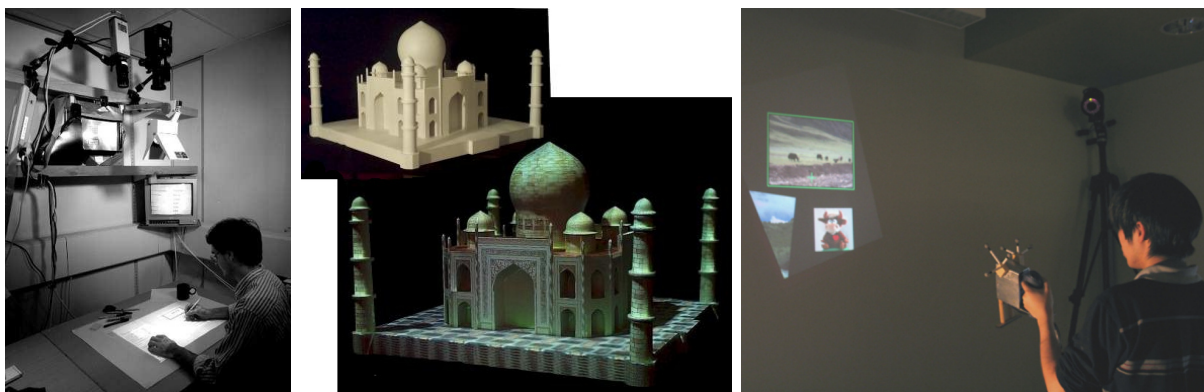


FIGURE 1.8: À gauche : Le DigitalDesk de Wellner (source : [162]). Au milieu : utilisation de plusieurs projecteurs par Raskar et al. pour augmenter une surface 3D (source : [121]). À droite : le système de RA spatialisée mobile de Cao et Balakrishnan (source : [29]).

physiques tels que des stylos ou directement les mains [162]. Raskar et al. ont élargi le concept à l'augmentation d'objets physiques 3D de formes complexes en utilisant plusieurs projecteurs afin de mapper la texture d'un modèle virtuel de même forme que l'objet physique sur ce dernier [121]. Ce système est illustré dans la figure 1.8, au milieu. Bimber et Raskar qualifient ces systèmes de réalité augmentée spatialisée [17]. Dans les années 2000, ils constituaient la seule alternative de RA haute résolution et convaincante. Ces dispositifs sont adaptés pour augmenter des surfaces fixes, y compris la surface d'objets 3D, en permettant à plusieurs utilisateurs de profiter de l'augmentation simultanément sans porter de dispositif particulier. En revanche, ils nécessitent souvent une longue phase de calibration pour s'intégrer dans un environnement précis, sont sujets à des problèmes d'occultation par les utilisateurs eux-mêmes, et souffrent d'un manque de mobilité. Pour pallier ce manque de mobilité, des solutions portables sont étudiées : Cao et Balakrishnan utilisent un projecteur portable qui agit comme une lampe de poche pour afficher des informations virtuelles 2D sur les murs ou sur des tables [29], comme c'est illustré dans la figure 1.8, à droite. Cette solution est un bon compromis entre la RA spatialisée et la RA mobile, mais le manque de stabilité et la latence de leur prototype ne permettent pas de fournir une augmentation HF.

Une autre limitation de ces systèmes est qu'ils se contentent d'augmenter des objets déjà existant dans le monde physique : ils ne permettent pas d'inclure des objets 3D purement virtuels dans l'environnement physique des utilisateurs. Jones et al. proposent par exemple de construire les objets physiques à augmenter à l'aide d'un assemblage de briques blanches, et d'utiliser un système projecteur-caméras pour que le système déduise automatiquement la forme des objets à augmenter à partir d'une projection de lumière structurée [79]. Ce dispositif est illustré dans la figure 1.9, à gauche. Les systèmes projecteur-caméras sont régulièrement utilisés dans le contexte de la RA spatialisée car les caméras RGB et caméras de profondeur permettent de détecter les mouvements de plusieurs utilisateurs simultanément sans que ces derniers aient besoin de porter des équipements supplémentaires. Les caméras permettent également une calibration précise et automatique des projecteurs [80], y compris pour des systèmes à plusieurs projecteurs dont les surfaces d'affichage se recoupent par endroits comme le RoomAlive de Jones et al. [78]. Ces systèmes utilisent souvent une Kinect de Microsoft (sortie en 2012 sous Windows), qui procure une carte de profondeur en plus d'une image RGB pour un coût

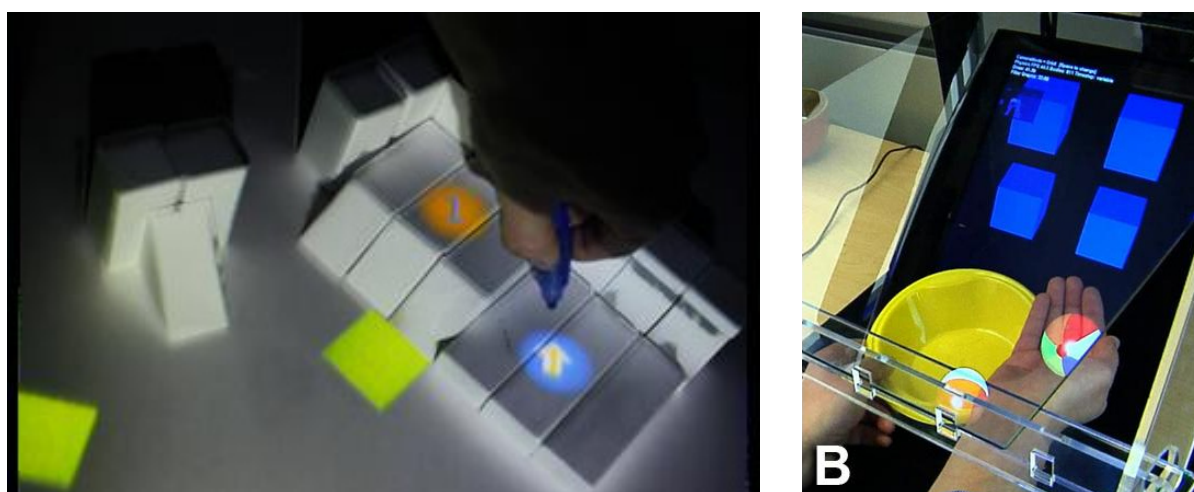


FIGURE 1.9: A gauche : le système de Jones et al. détectant automatiquement les briques tangibles utilisées pour interagir (source : [79]). À droite : le système de Hilliges et al. utilisant un miroir semi-transparent (source : [68]).

particulièrement faible (moins de 200 euros). Aujourd’hui, plusieurs autres systèmes, encore moins chers et tout aussi performants, sont disponibles pour le grand public. Un cas d’application particulièrement adapté à ces systèmes est l’augmentation de salles de réunion ou de conférence. L’augmentation de murs ou de bureaux permet d’offrir de larges surfaces affichables et favorisent le transfert d’informations et la collaboration des participants. Pour cette application, de hautes performances techniques ne sont pas nécessaires et seules des surfaces sont augmentées pour afficher du contenu 2D.

En utilisant un projecteur stéréoscopique et en couplant l’affichage à la position de l’utilisateur, il serait possible d’afficher des objets 3D virtuels, mais les objets apparaîtraient toujours en arrière-plan, sans gestion correcte des occultations possibles (comme dans une CAVE). Pour gérer ce problème, il est possible d’utiliser un miroir semi-transparent afin de superposer des images digitales par-dessus l’espace de travail physique. Cette solution permet de gérer les occultations numériquement en adaptant les images à afficher en fonction du contexte. Hilliges et al. ont implémenté un tel système en utilisant une Kinect pour la détection de profondeur et un écran LCD pour l’affichage [68]. Leur système fixe, illustré figure 1.9 à droite, permet une gestion correcte des occultations entre les mains de l’utilisateur et les entités virtuelles et ainsi des interactions fidèles avec celles-ci.

Performances techniques :

Dans un contexte d’augmentation de surfaces avec systèmes projecteur-caméras fixes, les performances sont essentiellement limitées par les capacités des caméras de profondeur. L’un des systèmes les plus performants actuellement est la Kinect Azure, sortie en 2020, qui procure une image 4K et une carte de profondeur de résolution 1024x1024 pixels à une fréquence d’affichage de 30 Hz. La précision de capture de position est de 12 mm à 1 m de la caméra pour une latence hardware moyenne de 40 ms. L’inconvénient de cette approche est qu’il est nécessaire d’appliquer un algorithme coûteux pour déduire la position de l’utilisateur en fonction des données des caméras. Ces algorithmes ajoutent généralement une latence supérieure à 60 ms. L’alternative est d’utiliser un système de capture de position optique, mais ces systèmes nécessitent de placer des marqueurs passifs sur les objets à détecter et sont significativement plus coûteux.

En ce qui concerne les projecteurs, il existe de plus en plus de projecteurs 4K à courte focale permettant d’excellents niveaux de contraste. Certains projecteurs récents possèdent une latence d’affichage très faible (nous avons mesuré une latence inférieure à 20 ms pour le Barco F50 WQXGA), bien que cette caractéristique soit rarement optimisée par les constructeurs. Aujourd’hui, la puissance de calcul des CPU et GPU sur le marché est suffisante pour permettre l’affichage de scènes complexes à haute résolution sans perdre en fréquence d’affichage. La latence totale d’un système de RA projetée tourne autour de 140 ms pour des systèmes projecteur-caméras, et peut descendre en dessous de 30 ms pour un système de capture optique avec marqueurs.

Les performances de systèmes de RA spatialisée avec miroirs semi-transparentes tels que celui de Hilliges et al. sont comparables aux performances de systèmes projecteur-caméras. En effet, les écrans LCD les plus récents offrent une résolution 4K à 60 Hz ou 120 Hz pour une latence d’affichage inférieure à 15 ms.

RA mobile

La RA mobile est essentiellement réalisée avec des smartphones et des tablettes qui fournissent un écran tactile, une caméra et des capteurs de mouvements internes. Les premiers smartphones tactiles ont été inventés dans les années 90, mais ils ne possédaient ni

la puissance de calcul ni les capteurs permettant de réaliser de la RA. Il faudra attendre la révolution des smartphones en 2007 pour envisager de la RA mobile exploitable pour un usage réel, avec l'arrivée de l'iPhone de Apple : le premier appareil disposant à la fois d'un écran tactile et d'une caméra de résolutions raisonnables (320x480 pixels pour l'écran, 1600x1200 pixels pour la caméra), mais également d'accéléromètres et d'un CPU performants. Dans leur revue des technologies de RA existant en 2010, Van Krevelen et Poelman présentent les premiers travaux exploratoires de RA mobile et concluent que l'arrivée des smartphones constitue la piste la plus prometteuse pour déployer la RA, grâce à leur faible coût et leur portabilité [153]. La RA mobile a effectivement eu beaucoup de succès durant cette dernière décennie : il s'agit du sujet qui a été le plus traité par les articles de la conférence internationale sur les réalités mixtes et augmentées (ISMAR) depuis 2008 [81]. Elle a également eu beaucoup de succès auprès du grand public, notamment grâce au jeu mobile Pokémon Go, qui affiche des créatures dans l'environnement du joueur en utilisant de la RA, comme l'illustre la figure 1.10, à gauche. Ce jeu, sorti en 2016, a totalisé plus de 500 millions de téléchargements en moins de trois mois. Ces systèmes font également leur apparition dans l'industrie, notamment pour la formation et l'assistance de techniciens pour la maintenance de machines complexes. Ces applications bénéficient de l'importante mobilité procurée par cette solution.

La RA mobile consiste à regarder son environnement physique à travers l'appareil mobile comme s'il s'agissait d'une lentille magique. Ce concept initialement introduit pour des applications graphiques sur ordinateur de bureau consiste à modifier numériquement l'image vue à travers la lentille [16]. Ainsi, l'appareil mobile affiche la vue de sa caméra arrière en y intégrant des entités virtuelles. Afin de créer l'illusion que les entités virtuelles ont une place dans l'environnement physique, les informations virtuelles à intégrer sont calculées en fonction de la position de l'appareil dans son environnement. Cette solution repose donc sur la connaissance de la position de l'appareil dans l'espace [166], rendue possible soit en exploitant des informations de déplacement relatif obtenues à l'aide d'accéléromètres ou de gyroscopes ; soit en analysant les images caméra afin de reconnaître des points d'intérêt fixes de l'environnement. L'analyse des images caméra permet un calcul de position plus fiable et plus stable mais augmente la latence totale du système, et ses performances peuvent varier en fonction de l'aspect visuel de l'environnement (détails identifiables, luminosité, ...). Les systèmes récents appliquent souvent une méthode hybride qui utilise les informations des capteurs internes disponibles pour accélérer l'analyse d'images.



FIGURE 1.10: À gauche : illustration du jeu Pokémon Go de Niantic, exploitant la RA mobile (2016). À droite : illustration de la librairie logicielle pour la RA ArCore de Google. Des points apparaissent au niveau des surfaces détectées automatiquement et permettent d'y ancrer des objets virtuels.

Depuis 2018, Google et Apple fournissent des bibliothèques logicielles de développement de RA mobile, nommées ARCore et Tango pour Google et ARKit pour Apple. Ces deux bibliothèques exploitent des méthodes hybrides pour connaître la position de l'appareil dans l'environnement physique. Les deux bibliothèques détectent également automatiquement les surfaces planes dans l'environnement physique observé, comme illustré dans la figure 1.10, à droite.

Dans un contexte de RA mobile sur surface tactile, l'interaction avec des entités se fait généralement au niveau de l'écran, en cliquant sur les images projetées des entités virtuelles à atteindre. Avec de petits écrans et l'instabilité causée par le tremblement de la main, l'interaction souffre d'une faible précision. La littérature a proposé des techniques permettant de gagner en précision en tirant parti de la surface multi-touch. Vincent et al., par exemple, ont réussi à améliorer la précision de pointage en RA mobile en contrôlant finement un curseur avec des mouvements relatifs du pouce sur l'écran [155]. En combinant un smartphone avec un Leap motion, Qian et al. ont réussi à fournir une manipulation intuitive à main levée, réduisant la charge cognitive nécessaire pour sélectionner et déplacer les objets virtuels [120].

Grâce à la quantité d'informations contextuelles et d'indices de profondeur disponibles dans la réalité physique, le manque de stéréoscopie est moins préjudiciable pour les systèmes de RA mobiles que pour les systèmes de RV, mais cela limite quand même les performances 3D. De plus, avec ces systèmes, l'affichage est couplé à la position de l'appareil et non pas directement à la position de l'utilisateur. Baricevic et al. ont montré qu'un couplage à la position de l'appareil limite les performances de sélection de cibles avec une tablette 11 pouces, mais pas avec un smartphone 3.5 pouces, comparé à un couplage à la position de l'utilisateur [7]. La différence d'affichage en fonction du couplage implémenté est illustrée dans la figure 1.11, à gauche. Les auteurs proposent également un prototype démontrant la faisabilité d'un système de RA mobile dont l'affichage est couplé à la position de l'utilisateur. Les surfaces tactiles récentes possédant déjà des capteurs de profondeur et une caméra à l'avant de l'appareil, il est raisonnable d'imaginer le développement de techniques de couplage utilisateur dans les prochaines années.

Performances techniques :

L'un des smartphones les plus performants dans un contexte de RA mobile est le Lenovo Phab 2 Pro, illustré dans la figure 1.11, à droite. Il possède une caméra de profondeur en plus d'une caméra 4K, et implémente la solution logicielle Tango de Google. Sa résolution d'affichage est de 2560x1440 pixels à 60 Hz. La précision de capture de position est

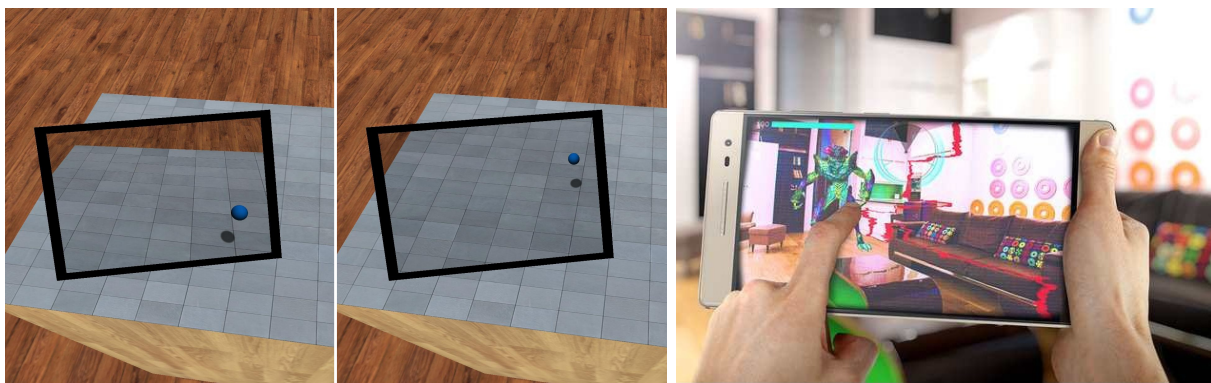


FIGURE 1.11: À gauche : différences d'affichage d'une tablette tactile en fonction de la mise en place d'un couplage à la position de la tablette (à gauche) ou d'un couplage à la position de l'utilisateur (à droite) (source : [7]). À droite : illustration de l'utilisation du Phab 2 Pro de Lenovo dans un contexte de RA mobile.

de l'ordre de quelques centimètres à une distance de 1m de la caméra et elle diminue au cours du temps d'utilisation. Nous n'avons pas trouvé d'évaluation de la latence avec ce système mais au vu des technologies employées, nous estimons qu'elle dépasse 100 ms. En 2019, plusieurs constructeurs de smartphone incluant Sony et Honor ont développé des appareils possédant plusieurs caméras arrière et un capteur de profondeur. Ces appareils et leurs successeurs devraient permettre d'améliorer la précision et la latence des systèmes de RA mobile, afin d'atteindre un haut niveau de fidélité.

Casques de RA

Dans leur état de l'art de la RA en 2001, Azuma et al. indiquent qu'aucun système de RA semi-transparent existant ne permet un affichage assez fidèle et assez stable pour être véritablement exploitable [5]. Le casque de RA le plus convaincant à l'époque est le prototype de Höllerer et al., qui permet un affichage de résolution 400x300 pixels par œil, pour un champ de vision de 35° [73]. Ce système de capture de position hybride composé d'un GPS et d'un capteur inertiel a une précision de l'ordre du centimètre. Il s'agit du premier prototype mobile fonctionnant en extérieur. Ce système est illustré dans la figure 1.12, à gauche. Malgré les limites technologiques, Fuchs et al. ont su montrer le potentiel des casques de RA pour l'assistance de chirurgiens dans une tâche de biopsie à l'aiguille [55]. Ils ont simulé un casque de RA en augmentant en temps réel les images filmées par des caméras placées à l'avant d'un casque opaque, permettant l'affichage de données échographiques volumétriques co-localisées avec le corps du patient sans gêner les mouvements du praticien. Si leur prototype n'était pas assez performant pour une utilisation clinique, ils ont montré un potentiel applicatif fort des casques de RA.

Au cours des années 2000, de nombreux travaux de recherche ont été menés pour résoudre les problèmes de capture de position et d'affichage des casques de RA, notamment parmi les publications à la conférence internationale ISMAR, recensées par Zhou et al. en 2008 [169]. Pendant cette période, les avancées des systèmes de capture optique ont permis d'atteindre les niveaux de précision nécessaires à une visualisation stable dans un casque de RA. Cependant, les casques de RA restent trop limités dans leurs capacités d'affichage, principalement à cause de trop faibles résolutions et de champs de vision pour des dispositifs trop lents et trop lourds. D'après Zhou et al., ces systèmes sont également trop chers pour permettre un déploiement grand public. Aussi, Kruijff et al. recensent d'autres problèmes de perception impactant la fidélité des casques de RA, parmi eux : le conflit

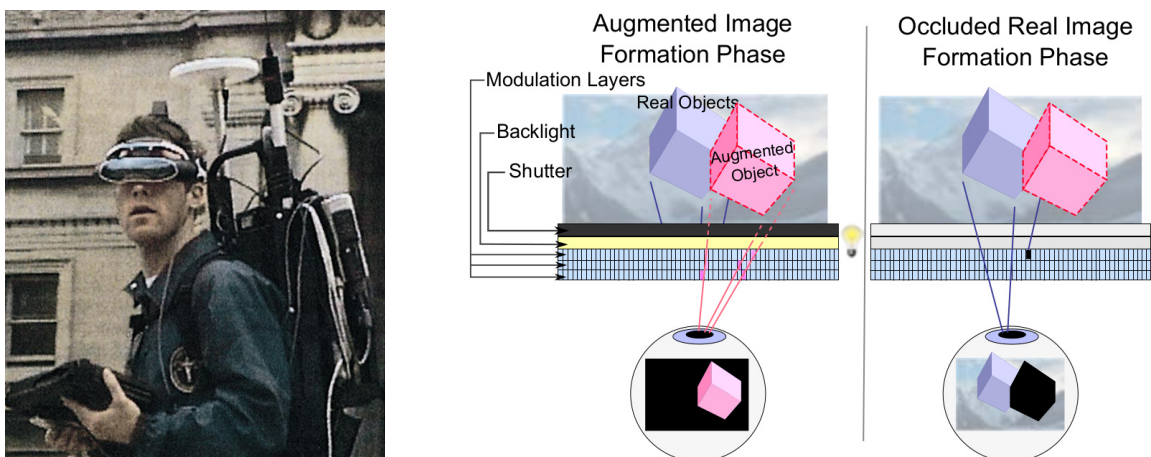


FIGURE 1.12: À gauche : le système complet de RA mobile en extérieur de Höllerer et al. (source : [73]). À droite : schéma montrant le fonctionnement du prototype de Maimone et Fuchs permettant un affichage multi-focal opaque d'objets virtuels (source : [98]).

convergence-accommodation, la mauvaise gestion des occultations, le manque d'opacité et de tangibilité [84]. Maimone et Fuchs présentent un prototype basé sur l'utilisation de plusieurs couches d'écrans LCD rétroéclairés permettant un affichage multi-focal avec un large champ de vision [98]. Leur solution permet également de masquer l'image de la réalité physique là où les objets virtuels sont supposés cacher la vue, permettant une excellente gestion d'opacité. Cette méthode est illustrée dans la figure 1.12, à droite. Ce prototype résout trois problèmes majeurs d'affichage identifiés, mais n'a pas encore été implémenté dans un casque de RA.

Au début des années 2010, l'arrivée des smartphones a fait passer la RA mobile au premier plan, et l'absence de casque de RA suffisamment performant a impacté l'intérêt de la recherche dans ce domaine. Suivant la révolution des casques de RV, une première génération de casques de RA convaincants voit le jour en 2017, avec la sortie du HoloLens de Microsoft, du Meta 2 de metavision (illustré dans la figure 1.13, à gauche) et des Smart-Glasses de DAQRI. Ces casques souffrent encore de problèmes de conflits convergence-accommodation, de manque d'opacité, et de faible champ de vision pour deux d'entre eux, mais ils permettent néanmoins une interaction haute-fidélité avec du contenu augmenté. Ces casques proposent tous des possibilités d'interaction isomorphe à main levée via des caméras analysant les mouvements des mains de l'utilisateur. En termes d'applications, le traitement par l'exposition de phobies de petits animaux est particulièrement adapté à l'utilisation de casques de RA, qui permettent une forte colocalisation du stimulus et des membres physiques du patient. Ces systèmes favorisent également la collaboration ainsi que l'assistance à la réalisation de tâches critiques telles que la chirurgie, car les utilisateurs perçoivent entièrement leur environnement physique et naviguent à l'intérieur sans devoir tenir un appareil.

Toutefois, les interactions à main levée ont tendance à générer une fatigue musculaire dans le bras, appelée effet de bras de gorille [69]. D'autres techniques sont envisagées pour éviter cette fatigue. Kyto et al. ont étudié des techniques d'interaction basées sur les mouvements de la tête et les mouvements oculaires [86]. Les techniques basées sur les yeux étaient les plus rapides mais ont engendré beaucoup d'erreurs. Les techniques de mouvement de la tête étaient à la fois plus rapides et plus précises que du pointage à main levée ou utilisant un contrôleur relatif (une souris gyroscopique), et ont démontré

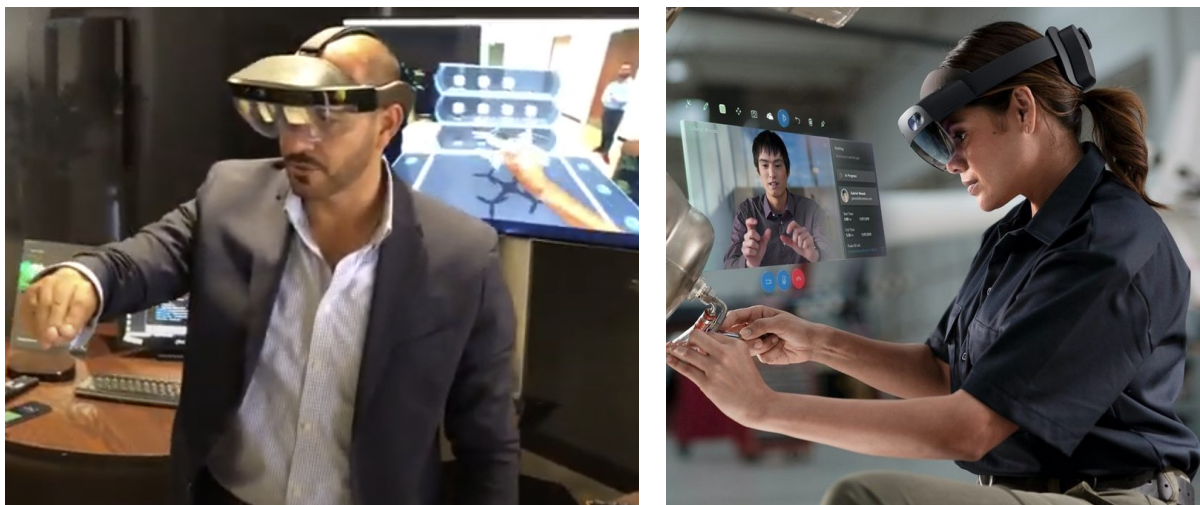


FIGURE 1.13: À gauche : le casque de RA meta 2 de metavision. La TV en arrière-plan affiche la vue de l'utilisateur alors qu'il déplace l'objet virtuel avec sa main (2017). À droite : illustration de casque de RA HoloLens 2 de Microsoft représentant le panneau virtuel vu par l'utilisatrice (2020).

le potentiel de cette approche. Sadri et al. ont trouvé un résultat similaire en comparant des mouvements de la tête par rapport à une interaction à main levée pour manipuler un modèle anatomique 3D visualisé avec un Hololens [124].

Performances techniques :

Le Meta 2 est le seul casque de RA avec un champ de vision élevé : 90° , mais le reste de ses performances est limité par rapport aux autres casques de RA. Il offre un affichage de résolution 1280×1440 pixels par œil à 70 Hz. Nous avons mesuré la précision du système de capture de position à 1.7 mm et une latence totale de 60 ms. En comparaison, le Microsoft Hololens 2 (sorti en 2020, illustré dans la figure 1.13, à droite) et le Magic Leap One (sorti en 2019) offrent tous deux un champ de vision horizontal de l'ordre de 40° et un affichage à 120 Hz. L'Hololens 1 offre une précision de capture de 0.2 mm et une latence totale inférieure à 20 ms. La précision et la latence de l'Hololens 2 sont comparables. Enfin, les appareils de Microsoft possèdent tous les deux un affichage haute résolution de 1920×1080 pixels par œil.

1.2.4 Bilan

Les technologies permettant de développer des systèmes d'interaction haute-fidélité sont arrivées sur le marché lors de cette dernière décennie. Les CAVEs et la RA spatialisée ont été les premières approches à offrir une interaction HF dans certains contextes, mais leur coût de mise en place et leur manque de mobilité les rendent difficilement déployables. L'année 2016 a marqué un tournant technologique dans les domaines de la RV et de la RA, avec la commercialisation des premiers casques de RV et de RA HF, et avec le déploiement de smartphones de plus en plus adaptés à l'utilisation de la RA mobile. À l'heure actuelle, les technologies évoluent rapidement et des nouvelles générations de systèmes HF de plus en plus performants arrivent sur le marché tous les deux ou trois ans.

1.3 Un APCD haute-fidélité

Les appareils à perspective corrigée déplaçables (APCD) ont été introduits par Stavness et al. en 2010 [142]. Ces appareils manipulables, souvent de forme cubique ou sphérique, permettent d'interagir avec des entités virtuelles 3D grâce à un affichage couplé à la fois à la position de l'utilisateur et à la position de l'appareil lui-même. Cela permet de créer l'illusion qu'une entité virtuelle est présente à l'intérieur de l'appareil déplaçable, comme si l'utilisateur tenait un objet transparent contenant une entité physique à l'intérieur. Les APCD sont à mi-chemin entre les aquariums de RV et la RA spatialisée. Ils sont proches de la RA dans le sens où ils créent l'illusion de la présence d'une entité virtuelle dans un objet physique qui appartient à l'environnement de l'utilisateur à l'intérieur duquel il est déplaçable. Ils sont également proches de la RV, dans le sens où l'entité virtuelle semble résider dans l'appareil plutôt que dans l'environnement général de l'utilisateur, comme pour les aquariums de RV.

Un APCD doit être mobile, et la totalité de sa surface doit être affichable. Un APCD haute-fidélité doit être assez réactif et précis pour conserver un affichage fidèle, même lorsque l'utilisateur manipule l'objet rapidement. A notre connaissance, le seul APCD haute-fidélité existant est le prototype de Bérard et Louis [25], que nous avons mis en œuvre dans les différentes contributions de cette thèse. Dans cette section, nous détaillons les performances techniques de ce prototype après avoir introduit l'historique des APCD, puis nous discutons brièvement des avantages de cette approche.

1.3.1 Les premiers prototypes d'APCD

Le premier prototype d'APCD est le pCubee de Stavness et al. [142], illustré dans la figure 1.14, à gauche. Il se présente sous la forme d'un appareil cubique de 15 cm de côté et de 12 cm de hauteur, dont cinq faces sont recouvertes d'écrans LCD. Ce prototype utilise un système de capture de mouvements électromagnétiques Polhemus pour suivre la position du cube et de l'utilisateur à 60 Hz. L'affichage des cinq faces du cube reste alors couplé à la position de l'utilisateur, quelles que soient la position et l'orientation de l'appareil. En affichant un objet virtuel dans le référentiel de l'APCD, l'utilisateur a accès à un contrôle tangible et isomorphe de l'objet virtuel simplement en déplaçant le cube. Les écrans utilisés ont une résolution de 640x480 pixels (160 dpi) et un taux de rafraîchissement de 60 Hz. Ce prototype ne fournit pas un affichage stéréoscopique.

Stavness et al. ont testé leur prototype dans une tâche de visualisation de graphes complexes, en comparant deux conditions d'affichage (cube ou écran plat), croisées avec

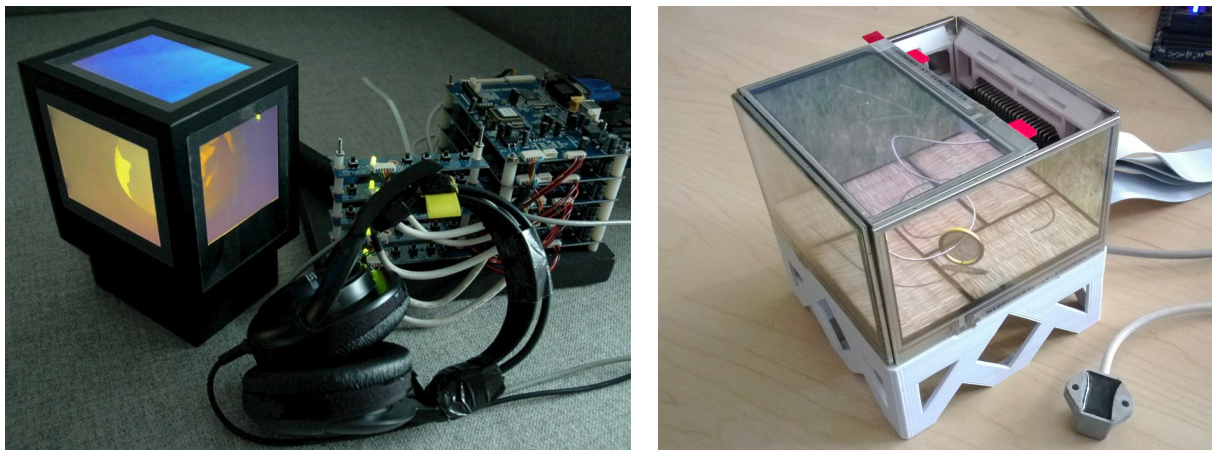


FIGURE 1.14: A gauche : le premier prototype d'APCD : le pCubee de Stavness et al. (source : [142]). À droite : le nouveau pCubee de Tang et al. (source : [148]).

deux conditions de manipulation (cube ou souris). Ils ont montré que l’affichage offert par le cube réduisait le taux d’erreurs de visualisation et améliorait le temps de réponse quand il était utilisé avec un contrôle à la souris. En revanche, la manipulation isomorphe du cube s’est avérée très peu performante par rapport à une interaction à la souris. Cependant, les limitations techniques et ergonomiques du prototype pCubee peuvent avoir eu de forts effets sur les résultats obtenus. Le cube était connecté à un ordinateur à l’aide d’un fil épais et pesait 1.3 kg, et les systèmes de capture électromagnétique ne permettent pas une précision en dessous du centimètre. Cela peut avoir empêché des rotations rapides et sans contrainte de l’appareil. De plus, les panneaux LCD avaient des bordures d’environ 2.3 cm d’épaisseur qui obstruaient une partie importante du contenu virtuel de l’écran, et la stéréo n’était pas utilisée. Dans une tâche d’exploration, ces larges bordures rendent le suivi visuel difficile en cas de rotation du cube.

Quatre ans plus tard, Tang et al. ont présenté une extension du travail précédent appelé le nouveau pCubee [148], illustré dans la figure 1.14, à droite. Ce second prototype met en jeu des écrans OLED tactiles et possède des bordures beaucoup plus fines (0.3 cm). Il est également plus léger mais son poids exact n’est pas mentionné. Ce prototype est plus performant que le précédent, mais il utilise toujours un système de capture de mouvement électromagnétique, et l’affichage n’est toujours pas stéréoscopique.

1.3.2 Une implémentation HF

Bérard et Louis ont implémenté un prototype d’APCD haute performance, léger et sans bordure [25]. L’affichage stéréoscopique est réalisé avec un projecteur 120 Hz à faible latence au lieu d’écrans LCD. Un système de capture de mouvement optique à 240 Hz très précis (0.022 mm est utilisé) pour atteindre des performances haute fidélité. Ce système fournit une combinaison d’indices de profondeur similaires à ceux du monde physique : projection en perspective, affichage stéréoscopique, parallaxe de mouvement et ombrage. La figure 1.15 illustre la qualité de rendu de ce prototype. La stéréoscopie est désactivée dans cette figure à des fins d’illustration.

Lien entre ce prototype et notre travail

Ce prototype a été implémenté en 2016 par François Bérard et moi-même. Il a été un des fondements de nos travaux, notamment parce que nous avons pu le mettre en jeu dans nos différentes contributions. Cependant, nous considérons que son implémentation fait partie de notre état de l’art, au même titre que les autres prototypes HF présentés dans

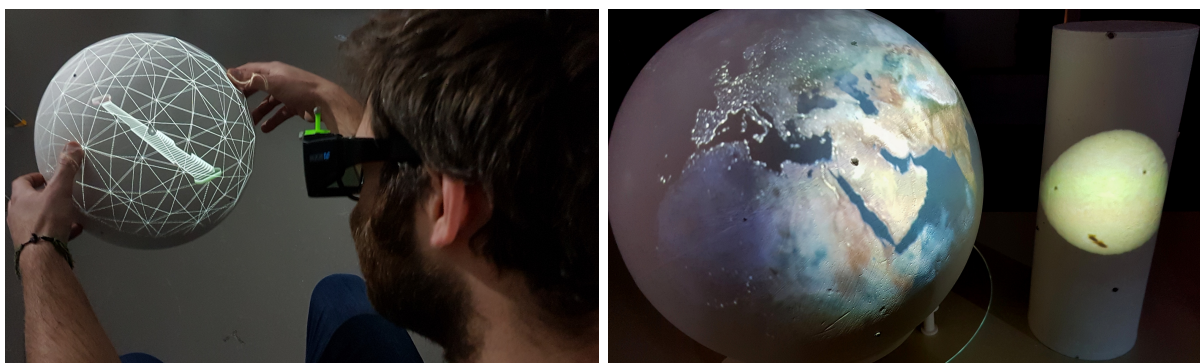


FIGURE 1.15: À gauche : un utilisateur manipulant l’APCD sphérique dans un jeu de suivi de chemin avec une bille. À droite : deux APCD peuvent être utilisés en même temps, ici la position du soleil influence sur l’éclairage de la planète. Les deux images sont présentées en vue monoscopique adaptée au point de vue de l’appareil photo à des fins d’illustration.

ce chapitre. Les détails de l'implémentation de ce prototype sont néanmoins disponibles dans l'annexe A. Ce prototype a été présenté pour la première fois dans la conférence CHI 2017 [25], a reçu la récompense de meilleure démonstration de la conférence ISMAR en 2018 [26], et a été continuellement amélioré depuis lors, notamment dans le cadre de cette thèse. Les performances présentées dans cette section sont celles qui ont été atteintes en 2020.

Projection externe versus Écrans couvrants

Recouvrir un appareil d'écrans LCD comme un aquarium de RV n'est pas la seule solution pour rendre toute sa surface affichable. Il est aussi possible d'utiliser un objet blanc sur lequel projeter les images voulues avec un ou plusieurs projecteurs externes, comme pour différents prototypes de RA spatialisée, présentés en section 1.2.3. L'utilisation de plusieurs projecteurs permet d'éviter la présence d'angles morts.

La projection externe rend possible l'utilisation de n'importe quel objet passif en tant qu'APCD, et permet ainsi de résoudre plusieurs problèmes inhérents aux limitations technologiques des écrans LCD actuels. Le seul critère restant dépendant de l'objet tenu étant sa surface, il est possible d'utiliser des objets creux et donc légers et très confortables à déplacer, même pour de longues sessions. Les appareils étant passifs, ils n'ont pas besoin d'être alimentés par un câble qui pourrait bloquer des rotations amples. De plus, cette approche offre une grande flexibilité dans le choix de la forme de l'appareil, et permet de produire plusieurs APCD, de formes et de tailles différentes, avec la même installation. Enfin, cette approche évite toute bordure, ce qui permet un affichage lisse sur toute la surface non occultée de l'appareil.

En contrepartie, l'utilisation de projection externe apporte les problèmes de l'occultation causée par le corps de l'utilisateur et de la limitation de la mobilité du dispositif au volume de projection. Ces problèmes sont atténués dans le cas d'une utilisation assise : la position la plus confortable pour de longues sessions de manipulation. Dans cette position, toute la zone à portée de bras de l'utilisateur peut être contenue à l'intérieur du volume de projection, y compris avec un seul projecteur placé au-dessus de sa tête comme cela est illustré dans la figure 1.16. Décaler légèrement le projecteur sur le côté permet également d'éviter que l'APCD ne se retrouve dans l'alignement du projecteur et de la tête de l'utilisateur. Enfin, l'occultation causée par les mains n'est pas gênante car c'est le comportement attendu. À l'inverse, le fait de voir l'image projetée sur ses doigts pourrait gêner, mais l'image y est trop distordue et trop mal réfléchi sur la peau pour que les doigts soient perçus comme transparents.

Matériel et performances techniques

Le prototype de Bérard et Louis utilise un projecteur Barco F-50 de 2560x1600 pixels à 120 Hz, fournissant une résolution moyenne de 90 dpi au centre de la sphère (la résolution varie en fonction de la distance sphère-projecteur), et un taux de rafraîchissement de 60 Hz par œil. Des lunettes à obturateur LCD sont utilisées pour l'affichage stéréoscopique. Les auteurs calibrent manuellement la position de mise au point du projecteur afin que l'image soit la plus nette sur les APCD quand l'utilisateur est en position de repos. Avec ce projecteur, l'image reste parfaitement nette dans une gamme de distances au projecteur de ± 15 cm par rapport à la distance de mise au point. Le projecteur est positionné au-dessus de la tête des participants, pointant vers le bas avec un angle d'environ 60° par rapport à l'horizontale, comme illustré sur la figure 1.16. La pose du projecteur dans le référentiel du système de capture ainsi que ses paramètres intrinsèques sont calibrés en utilisant une méthode semi-automatique. Le logiciel développé par les auteurs utilise ces paramètres pour déterminer en temps réel les images que le projecteur doit afficher en

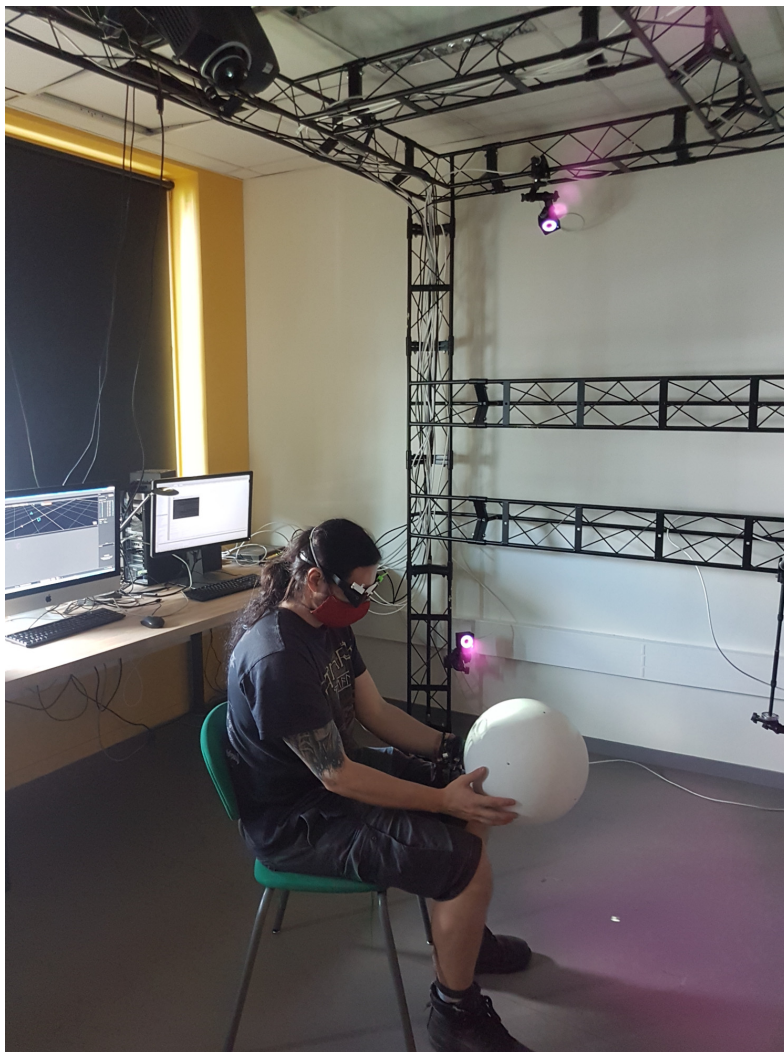


FIGURE 1.16: Configuration du projecteur et du système de suivi optique. L'utilisateur porte des lunettes à obturateur. Des marqueurs passifs permettent le suivi des lunettes et de l'APCD. (source : [25]).

fonction des poses de l'APCD et des lunettes à obturateur. La méthode de calibration ainsi que le logiciel gérant l'affichage sont décrits dans l'annexe A. Le logiciel est exécuté sur un ordinateur de bureau Intel Xeon 3.7 GHz avec une carte graphique NVIDIA RTX 2080 Ti. La capture de mouvement est effectuée par un système Optitrack à dix caméras Prime 13 à 240 Hz, offrant une précision de capture de 0.022 mm. L'utilisation d'une dizaine de marqueurs sur les APCD permet un suivi robuste : le système ne perd jamais leur trace, même lorsque les utilisateurs cachent plusieurs marqueurs en manipulant la sphère.

Les auteurs ont estimé la latence totale de ce système à 27 ms en utilisant l'approche prédictive simple de Cattani et al. [30]. Cette faible latence est quasiment imperceptible, sauf lorsqu'on secoue rapidement l'APCD. Un algorithme de prédiction linéaire élimine complètement ce reliquat de délai. Cet algorithme produit l'effet parasite d'amplifier le bruit blanc du système de capture. Il est possible de mettre en place cette prédiction progressivement en fonction de la vitesse de déplacement de l'appareil afin d'éviter d'impacter la stabilité du système au repos.

Les utilisateurs n'ont signalé aucune gêne oculaire lors de l'utilisation de l'appareil. De plus, ils ne portent que des lunettes à obturateur légères qui ne créent aucune fatigue musculaire au niveau du cou.

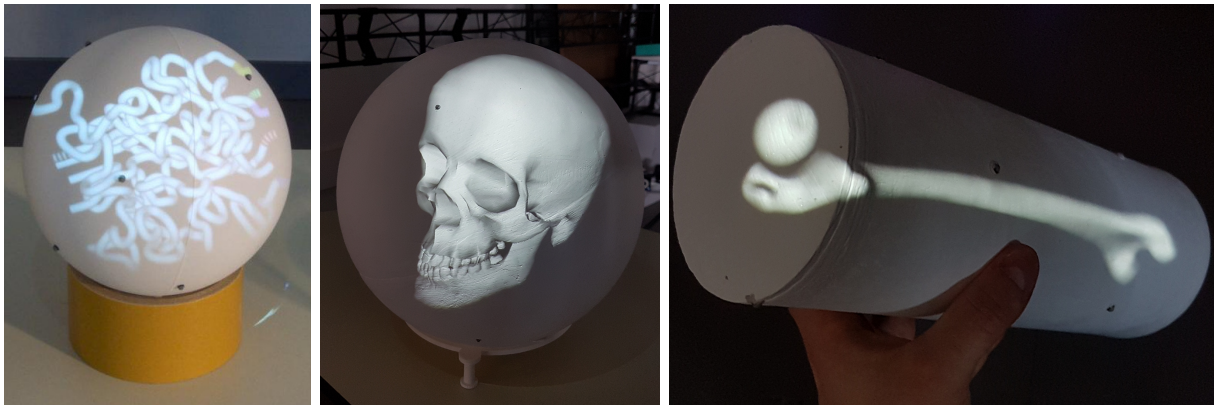


FIGURE 1.17: Différents APCD conçus par Bérard et Louis. À gauche : le petit APCD sphérique de 14 cm de diamètre. Au milieu : le grand APCD sphérique de 30 cm de diamètre. À droite : l'APCD cylindrique de 27 cm de long. Les trois images sont présentées en vue monoscopique adaptée au point de vue de l'appareil photo à des fins d'illustration.

Différentes formes d'APCD

La première forme d'APCD testée par Bérard et Louis était un cube tel que celui qu'utilisent Stavness et al. [142]. Cependant, les bords très anguleux entre les faces produisent des discontinuités dans l'affichage. Cet effet est d'autant plus marqué lors de l'utilisation d'un seul projecteur, où des faces larges toujours visibles par l'utilisateur peuvent se retrouver dans des orientations depuis lesquelles il est impossible de projeter des images nettes.

Les auteurs ont ainsi choisi d'utiliser des objets qui ne présentent aucun angle abrupt. Ils ont construit deux objets sphériques. Le premier (illustré figure 1.17 à gauche) fait 14 cm de diamètre et a une masse d'environ 100 g, il est composé d'une sphère en plastique peinte en blanc, et se manipule confortablement à une main. Le second (illustré figure 1.17 au milieu) est une large sphère de 30 cm de diamètre en polystyrène recouverte de peinture blanche, il pèse 200 g et nécessite les deux mains pour des manipulations confortables. Ils ont également construit un APCD cylindrique adapté pour la manipulation d'objets de forme oblongue. Il est illustré dans la figure 1.17, à droite. Il a été confectionné à partir d'un étui de whisky de 27 cm de long et de 9 cm de diamètre peint en blanc. Cet APCD pèse environ 0.1 kg et se manipule très bien à une et deux mains. Dans ce contexte d'utilisation, les auteurs précisent que les discontinuités au niveau des extrémités du cylindre sont moins gênantes qu'avec le cube car elles ne perturbent pas l'affichage lors d'une rotation autour de l'axe principal de l'appareil ; mais l'utilisation d'une capsule à la place d'un cylindre permettrait d'éviter entièrement ce problème.

1.3.3 Les Avantages des APCD

Bien qu'ils n'offrent qu'un champ de vision limité, comparés à la plupart des systèmes HF existants, les APCD présentent de nombreux avantages. La manipulation physique de l'appareil lui-même offre une interaction tangible plus intuitive que beaucoup d'interactions qui cherchent à reproduire de manière virtuelle ce genre de manipulations physiques. Les APCD ne coupent pas les utilisateurs de leur réalité physique, contrairement aux systèmes de RV. Ils offrent une expérience interactive confortable car ils peuvent toujours être amenés dans la zone de confort des utilisateurs. Contrairement aux casques de RV et de RA, ils ne nécessitent pas de porter des équipements encombrants sur la tête : les lunettes à obturateur sont notoirement plus légères que les casques et ne sont pas reliées par un câble. Les images sont formées à une distance naturelle des yeux, ce qui permet

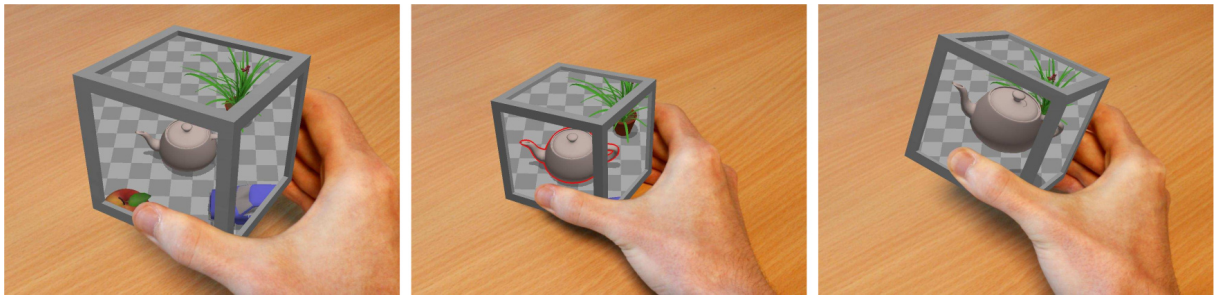


FIGURE 1.18: Illustration de Issartel et al. montrant comment un APCD peut être utilisé pour sélectionner et manipuler des objets virtuels au sein d’une large scène 3D (source : [74]).

de longues sessions d’utilisation grâce à la réduction de la fatigue oculaire et des maux de tête.

Une spécificité majeure des APCD est leur nature portable, il faut donc adapter leur taille afin qu’ils puissent être tenus aisément, soit d’une seule main, soit à deux mains. Proposer une conception d’APCD qui se tient à deux mains permet de profiter de trois avantages : obtenir un champ de vision plus large sur la scène virtuelle en tenant un APCD plus grand, manipuler l’APCD avec plus de précision et de stabilité qu’avec une seule main, et réduire l’effort de maintien de l’APCD dans l’air (c’est-à-dire réduire l’effet de bras de gorille) en répartissant la charge sur les deux bras. Comme pour les aquariums de RV, l’APCD représente une frontière tangible entre le monde physique où l’utilisateur est présent et le monde virtuel à l’intérieur de l’appareil. Cela permet d’atteindre un niveau élevé de présence qu’il serait difficile d’égaliser avec des casques de RA ou de RV. Lorsqu’un objet virtuel est rigidement “attaché” à l’intérieur d’un APCD, ses rotations sont directement contrôlées par les rotations de l’appareil lui-même. Ware et Rose ont montré qu’il était important d’avoir la main au même endroit qu’un objet virtuel manipulé pour effectuer des rotations efficaces de celui-ci, mais qu’il n’est pas nécessaire que la forme du contrôleur corresponde à la forme de l’objet virtuel [161]. Par conséquent, l’utilisation d’APCD semble être une approche prometteuse pour explorer et manipuler efficacement des entités 3D, bien que les APCD n’aient pas la même forme que les entités contrôlées.

Les APCD semblent adaptés pour l’analyse d’objets 3D complexes ou de données volumétriques, en vue de se former un modèle mental des objets. Des domaines applicatifs potentiels sont l’enseignement de notions 3D et l’aide à la planification d’opérations chirurgicales. En revanche, les APCD ne permettent pas de produire des expériences totalement immersives (contrairement aux CAVE et aux casques de RV), ni de superposer de l’information virtuelle avec un objet d’intérêt physique (contrairement à la majorité des systèmes de RA).

Comme les recherches sur les APCD sont récentes, les efforts ont été dirigés vers l’amélioration de leur mise en œuvre et vers les éléments fondamentaux de l’interaction 3D, tels que la sélection et la manipulation d’objets. Issartel et al. proposent que les APCD puissent être utilisés comme des portails volumiques permettant de naviguer au sein de larges scènes virtuelles [74]. Les auteurs ont utilisé un système de RA mobile avec une tablette pour simuler un APCD en raison de la complexité de la mise en œuvre d’un écran “entièrement portable et autonome”. Ils ont testé l’application d’une pression sur les côtés de l’écran comme métaphore pour saisir des objets virtuels, et ont pu observer que la métaphore était spontanément comprise par la majorité des utilisateurs. L’APCD offre ainsi une solution naturelle et tangible pour positionner et orienter des entités dans la scène virtuelle, comme l’illustre la figure 1.18.

Ces efforts poussent les APCD à devenir des options viables pour mettre en œuvre des applications autonomes réalistes dans des domaines tels que la modélisation 3D ou les simulateurs d'entraînement. Pour cela, il est nécessaire qu'un ensemble complet d'interactions soit disponible en plus du contrôle du point de vue et du pointage spatial. Mine a identifié cinq formes d'interactions fondamentales dans les environnements 3D virtuels [103] : le *mouvement* de l'utilisateur, la *manipulation* des entités, la *sélection*, la *mise à l'échelle* et la possibilité d'interagir via des *menus/widgets*. Il est important que des solutions soient apportées pour chacun de ces besoins interactifs afin de permettre une exploitation des APCD en-dehors des laboratoires de recherche. Nous présenterons dans le chapitre 6 notre contribution pour répondre à ce besoin.

Conclusion

Dans ce chapitre, nous avons présenté l'état de l'art des systèmes interactifs existants qui fournissent des capacités d'interaction **haute-fidélité** avec des scènes virtuelles 3D. On considère qu'un système est haute-fidélité s'il offre des moyens d'interagir avec des entités virtuelles qui sont comparables aux moyens d'interaction dont nous disposons dans notre réalité physique. Cela implique entre autres qu'il puisse fournir un affichage stéréoscopique couplé à la position de l'utilisateur, ainsi qu'un contrôle isomorphe des entités virtuelles.

Nous avons vu que malgré un intérêt de la recherche en RV et en RA sur le sujet depuis plus de trente ans, les progrès technologiques n'ont permis l'implémentation de prototypes HF que depuis les dix dernières années, et tout particulièrement depuis 2016. Aujourd'hui les systèmes convainquants les plus répandus sont les casques de RV et de RA, mais ils possèdent encore des défauts qu'il reste à corriger. Enfin, nous avons détaillé une approche novatrice, à mi-chemin entre la RV et la RA, qui paraît particulièrement adaptée pour fournir une interaction HF de qualité : les APCD. Nous avons présenté le prototype d'APCD HF de Bérard et Louis, que nous avons mis en jeu dans les différentes contributions de cette thèse, et dont les performances techniques en termes de précision et de latence sont comparables à celles des meilleurs dispositifs commerciaux actuels.

Le tableau 1.1 résume l'état de l'art en termes de performances techniques des différentes approches HF existant en 2020. La résolution (résol.) des casques et le champ de vision des autres appareils reportés dans ce tableau ont été calculés en fonction des distances utilisateur-affichage reportées dans la première colonne du tableau, qui correspondent à des cas d'usage typique des différents systèmes. Pour les aquariums de RV et les systèmes projecteur-caméras, le compromis résolution-taille de l'affichage est à optimiser en fonction du contexte. Nous donnons dans ce tableau des valeurs pour des tailles minimales et maximales pertinentes. Pour ces deux approches, la valeur minimale rapportée dans la colonne de résolution correspond à la valeur maximale rapportée dans la colonne de champ de vision, et vice-versa. Il est à noter que plus l'appareil est proche de l'utilisateur, plus le besoin de résolution se fait ressentir : un affichage à 32 dpi placé à 2m de distance produit une image perçue de même qualité qu'un affichage à 128 dpi placé à 50 cm. Différentes approches peuvent être privilégiées en fonction d'un besoin de résolution ou d'un besoin de large champ de vision. En termes de latence et de précision de capture de position, les approches de RA projecteur-caméras et de RA mobile sont en retard par rapport aux autres approches.

	Distance affichage (m)	Résol. affichage (dpi)	Champ de vision (°)	Fréqu. affichage (Hz/œil)	Précision capture (mm)	Latence (ms)
CAVE	2	32	>180	60	0.02	30
Aquarium de RV	0.5	160-320	34-68	60	0.02	30
Casque de RV	1	25	110	90	0.13	30
projecteur-caméras	1	58-134	40-80	60	11	100
RA mobile	0.5	460	16	60	20-60	100-200
Casque de RA	1	67	40	120	0.2	20
APCD	0.5	90	35	60	0.02	27

TABLE 1.1: Résumé des performances techniques des différentes approches HF existantes en 2020.

	Stéréo	Contrôle	Mobilité	isolement	Confort
CAVE	oui	contrôleurs 6D	$\simeq 10m^2$	espace dédié	oui
Aquarium de RV	possible	lancé de rayon	non	non	oui
Casque de RV	oui	contrôleurs 6D	$\simeq 10m^2$	total	non
projecteur-caméras	possible	main levée	non	non	oui
RA mobile	non	tactile ou main levée	oui	non	non
Casque de RA	oui	main levée	oui	non	non
APCD	oui	contrôle 6D	non	non	oui

TABLE 1.2: Résumé de caractéristiques générales des différentes approches HF existantes en 2020.

Le tableau 1.2 résume l'état de l'art concernant certaines caractéristiques centrées utilisateurs des différentes approches HF existant en 2020. La colonne confort fait référence à l'encombrement des casques et des appareils de RA mobile, qui ont tendance à provoquer de la fatigue musculaire, ainsi que de la fatigue mentale dans le cas des casques (nausée, maux de tête, mal aux yeux). Ce tableau met en avant certaines différences fondamentales qu'il est nécessaire de prendre en compte en fonction du contexte d'utilisation ciblé, telles que la qualité de contrôle recherché, le besoin de mobilité ou encore la nécessité d'être immergé ou au contraire de communiquer avec son environnement physique.

Nous entrons donc dans une ère où les systèmes d'interaction HF deviennent progressivement une réalité, et maintenant qu'il est possible de concevoir de telles technologies, il nous reste à identifier à quelles fins l'utilisation de techniques HF est bénéfique, et comment optimiser ces systèmes pour s'adapter au mieux à différents contextes d'utilisation.

Chapitre 2

Les bénéfices attendus de la haute fidélité

Introduction

De nombreuses tâches nécessitent que nous interagissions avec des entités 3D pour pouvoir être réalisées. En fonction des processus cognitifs mis en jeu lors de la réalisation de la tâche, l'expérience interactive peut prendre des formes particulièrement variées. Par exemple, utiliser un marteau pour enfoncer un clou, ou un livre pour lire une histoire, sont deux expériences très différentes, malgré l'exploitation des mêmes modalités perceptives (la vue et le toucher). L'utilisation d'un système d'interaction haute-fidélité (HF) pour interagir avec une entité 3D virtuelle pourrait s'avérer plus ou moins bénéfique selon les enjeux cognitifs de la tâche considérée. Nous avons voulu rendre compte de cette variété d'enjeux dans notre étude des bénéfices des systèmes HF. Ne pouvant pas nous attarder sur un panel exhaustif de tâches, nous nous sommes concentrés sur trois enjeux hétéroclites qui font appel à des processus cognitifs variés : la manipulation, l'apprentissage, et la présence. Ce chapitre regroupe l'état de l'art autour de ces trois enjeux. Les contributions correspondantes seront présentées dans les chapitres 3 à 5.

Dans la section 2.1, nous présentons l'état de l'art concernant notre premier enjeu : la manipulation d'objets dans l'espace, une tâche considérée comme fondamentale dans le domaine de l'interaction Homme-machine. Contrairement à des interfaces plus traditionnelles telles que les ordinateurs de bureau avec clavier et souris ou les surfaces tactiles, plus adaptés à des interactions en 2D, les dispositifs HF semblent permettre une manipulation intuitive et efficace d'objets dans l'espace 3D. Nous commencerons donc par nous demander si l'utilisation de dispositifs HF pourrait permettre **un contrôle d'objets plus performant**.

Après l'étude de performances directement liées à des capacités motrices, nous nous intéresserons dans la section 2.2 à une tâche cognitivement plus exigeante : l'apprentissage. Nous nous sommes naturellement dirigés vers cette tâche, car notre travail s'intégrait dans un projet de recherche orienté sur l'apprentissage de l'anatomie. Après avoir étudié quel rôle joue le choix d'une technique d'interaction dans un processus d'apprentissage, nous nous concentrerons sur le cas de l'*apprentissage d'objets 3D* : une tâche particulière dont les performances d'apprentissage semblent dépendre de la qualité des modèles utilisés, et des possibilités d'interactions disponibles pour l'apprenant. Ainsi, l'utilisation de la HF pourrait permettre **un apprentissage plus efficace**.

Enfin, nous nous attarderons sur le concept de présence dans la section 2.3 : un enjeu historiquement lié à la recherche sur la haute fidélité et dont la littérature foisonne. Contrairement aux deux premiers axes étudiés, il a été montré que la HF offre des bénéfices en lien avec la présence qui sont intéressants pour différents domaines de recherche.

Dans ce cas, après avoir fourni la réponse à la question “pourquoi”, nous questionnerons “comment” offrir **un meilleur sentiment de présence**.

Sommaire

Introduction	54
2.1 Un contrôle d’objets plus performant	56
2.1.1 Placer des objets dans l’espace	56
2.1.2 Périphériques permettant un contrôle 6D d’objets virtuels . . .	56
2.1.3 Bénéfices d’un contrôle isomorphe	61
2.1.4 Bilan	62
2.2 Un Apprentissage plus efficace	63
2.2.1 Adapter l’interaction aux besoins de la tâche d’apprentissage .	63
2.2.2 Apprendre un objet 3D	64
2.2.3 Des techniques d’interaction plus fidèles pour mieux apprendre un objet 3D	67
2.2.4 Bilan	70
2.3 Un meilleur sentiment de présence	71
2.3.1 Les définitions de la présence dans la littérature	71
2.3.2 Les applications de la présence	72
2.3.3 Comment produire un haut niveau de présence	75
2.3.4 Mesurer la présence	77
2.3.5 Bilan	79
Conclusion	80

2.1 Un contrôle d'objets plus performant

2.1.1 Placer des objets dans l'espace

Afin de permettre une exploration active de scènes virtuelles 3D, il est important que l'utilisateur ait la possibilité d'interagir avec les divers éléments de la scène. Même pour examiner un objet statique tel qu'une tasse, il semble plus efficace de tourner la tasse entre ses mains plutôt que de tourner soi-même autour de la tasse. Dans leur livre sur les interfaces utilisateurs 3D, Bowman et al. [22] expliquent que pouvoir déplacer un objet n'importe où dans une scène et être capable de l'orienter représentent l'une des tâches les plus fondamentales pour interagir avec des environnements virtuels.

Placer un objet dans une scène 3D est une tâche complexe à 6 degrés de liberté (6D), qui consiste à positionner l'objet au bon endroit dans l'espace 3D *et* à l'orienter dans la bonne direction dans cet espace. La plupart des logiciels de conception 3D sont aujourd'hui utilisés via une interface de bureau classique avec clavier et souris 2D. Réaliser une tâche 6D avec un périphérique de contrôle 2D est difficile en un seul geste. Pour permettre de placer des objets dans l'espace virtuel malgré tout, ces logiciels font appel à différents widgets qui permettent une décomposition de la tâche 6D en plusieurs tâches 2D solvables avec une souris d'ordinateur, comme l'illustre la figure 2.1.

Pourtant, placer un objet dans notre réalité physique peut sembler une tâche triviale, qui n'a pas besoin d'être décomposée en plusieurs sous-tâches pour être exécutée. Lorsque nous mettons une tasse dans le lave-vaisselle par exemple, nous effectuons un seul geste continu, qui a pour but de déplacer la tasse tout en la retournant (simultanément) afin qu'elle soit correctement orientée dans le lave-vaisselle. Une telle tâche demande à la fois un mouvement ample et précis, que nous arrivons à planifier inconsciemment. Si ces tâches sont simples à réaliser dans notre environnement physique, alors permettre d'interagir de manière naturelle dans l'environnement virtuel grâce à un contrôle d'objets haute-fidélité semble être une bonne piste pour améliorer les performances de placement 6D.

2.1.2 Périphériques permettant un contrôle 6D d'objets virtuels

Un premier effort de recherche a été de proposer de nouveaux périphériques de contrôle avec un nombre accru de degrés de liberté par rapport à la souris d'ordinateur. Ces appareils prennent en charge la résolution des 6D de la tâche *simultanément*. Certains périphériques cherchent à reproduire les mouvements faits par l'utilisateur tels quels sur l'objet

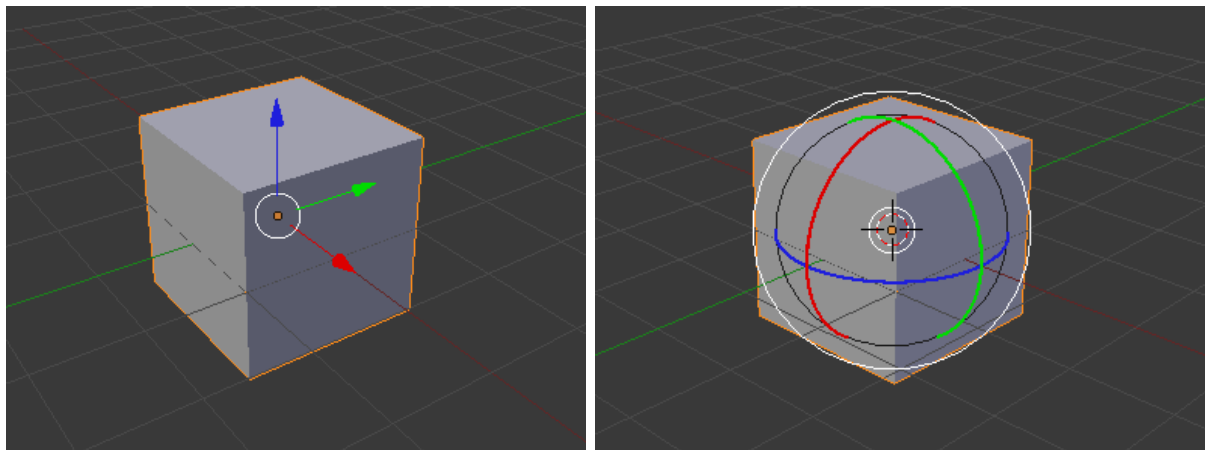


FIGURE 2.1: Widgets utilisés dans le logiciel blender pour traduire un objet (à gauche) ou pour orienter un objet (à droite). Sélectionner avec la souris l'un des axes ou des arcs de couleur permet de contraindre le mouvement suivant le seul degré de liberté sélectionné.

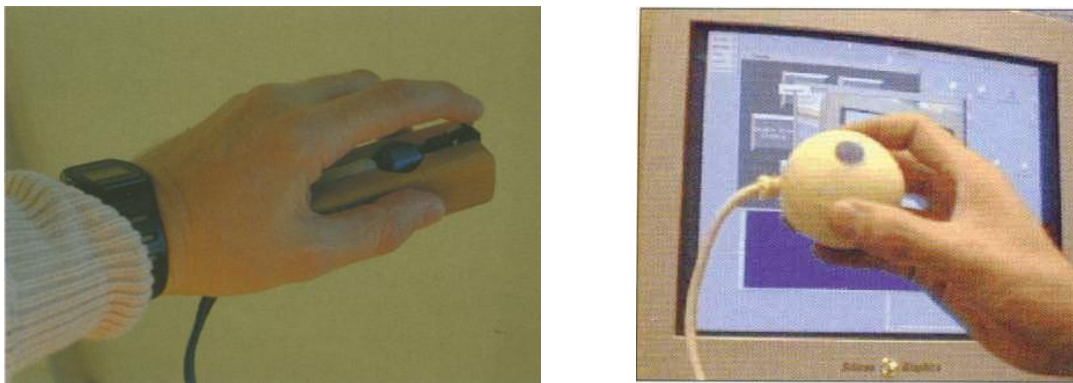


FIGURE 2.2: Périphériques de contrôle 6D électromagnétiques. À gauche : le Bat de Ware (source : [160]). À droite : le Finger Ball de Zhai et al. (source : [168]).

virtuel manipulé, de manière *isomorphe*. D'autres techniques, à l'inverse, cherchent à bénéficier au mieux de la stabilité et du confort possible avec un périphérique fixe. Un autre effort de recherche a été consacré à trouver les fonctions de transfert les plus compréhensibles entre les mouvements physiques effectués par l'utilisateur et les mouvements virtuels retournés par le système ; pour permettre un contrôle 6D intuitif et performant avec des périphériques d'entrée offrant moins de degrés de liberté de contrôle.

Des périphériques 6D électromagnétiques

Les premiers prototypes de contrôle 6D sont apparus vers la fin des années 80. Ware a implémenté un périphérique permettant un contrôle isomorphe *relatif* à main levée, et présentant un bouton pour activer des commandes [160]. Ce dispositif baptisé *Bat* est illustré sur la figure 2.2, à gauche. La position 6D du Bat est traquée par un système électromagnétique léger qui offre une précision de position de l'ordre de 2 millimètres, et une précision d'orientation de l'ordre de 2° . Cette technologie souffre d'un manque de stabilité conséquent, rendant impossible son utilisation pour un contrôle absolu. Ware n'a pas comparé les performances de ce dispositif avec d'autres dispositifs, mais il a montré qu'avec le Bat, les utilisateurs parvenaient à positionner et à tourner des objets virtuels simultanément. Il a mesuré que des tâches de placement 6D pouvaient être effectuées en moyenne en 14 secondes en simultané, au lieu de $13 + 9$ secondes respectivement en séparant la composante de rotation et de position. Ware a également montré que l'utilisation d'un affichage stéréoscopique permettait un large gain de performance dans des tâches de placement 6D, et tout particulièrement pour évaluer la profondeur de la cible.

Zhai et al. ont testé deux techniques permettant un contrôle isomorphe absolu à main levée : un gant et une balle manipulable avec les doigts (la Finger Ball illustrée dans la figure 2.2, à droite) [168]. Ces deux techniques utilisent la même technologie de capture de mouvements électromagnétiques Ascension Bird, possédant les mêmes limitations de précision que le Bat. Les auteurs ont démontré l'avantage d'utiliser les doigts pour manipuler le périphérique d'entrée plus efficacement. Ils concluent que les muscles des doigts permettent un contrôle plus fin que les muscles du bras et de l'épaule mis en jeu avec le gant. Avec cette étude, Zhai et al. ont également répandu l'utilisation d'un tétraèdre marqué pour l'évaluation des performances de placement 6D.

Des périphériques fixes

Des périphériques dits *élastiques* ont vu le jour. Il s'agit de dispositifs qui contrôlent la vitesse des objets manipulés à l'aide de joysticks qui retournent automatiquement à leur

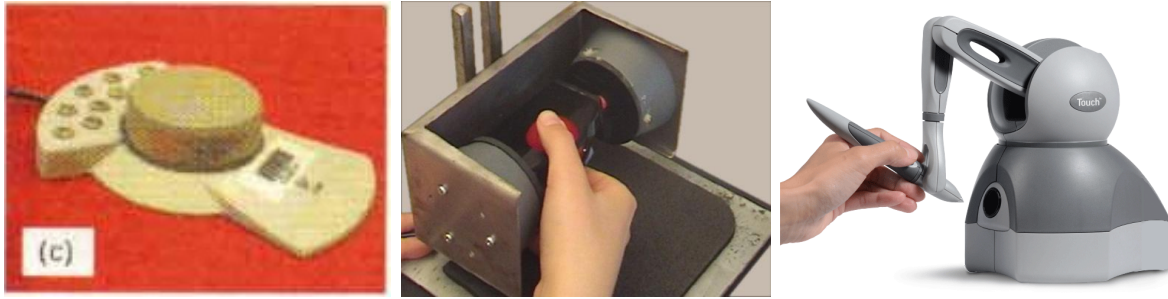


FIGURE 2.3: Des périphériques de contrôle 6D fixe. A Gauche : le Space Mouse de DLR (source : [167]). Au milieu : le GlobeFish de Froehlich et al. (source : [54]). A droite : le Phantom Omni de Geomatic.

position de repos dès qu'ils sont lâchés. Plus le joystick est loin de sa position de repos, plus l'objet se déplace vite dans cette direction. Ces périphériques ont l'avantage de rester fixes sur le bureau, permettant ainsi d'être manipulés en permanence dans une position confortable. Un exemple de périphérique 6D fixe (la Space Mouse) est illustré dans la figure 2.3, à gauche. Zhai a mesuré des performances similaires entre une Finger Ball et une Space Mouse dans une tâche de placement 6D [167]. D'après lui, l'un des avantages majeurs de ces périphériques est qu'ils permettent un contrôle stable et peu fatigant des objets, rendant possible une utilisation lors de longues sessions en contexte écologique telle que de la conception assistée par ordinateur. En revanche ces dispositifs semblent plus difficiles à utiliser et imposent une plus longue phase d'apprentissage.

Maslah et al. ont étudié la coordination de la résolution des différentes dimensions d'une tâche de placement. Ils ont constaté que résoudre toutes les dimensions en même temps peut avoir une demande en contrôle moteur trop élevée pour les utilisateurs ; et que ceux-ci ont tendance à décomposer la tâche en sous-tâches de dimensions inférieures quand cela se produit [100]. Ils ont en particulier observé ce phénomène avec l'utilisation de périphériques élastiques, qui demandent à l'utilisateur un contrôle moins naturel et donc cognitivement plus exigeant. Ils constatent que les utilisateurs décomposent alors la tâche en deux sous-tâches : une tâche de position 3D et une tâche d'orientation 3D.

Froehlich et al. ont capitalisé sur cette observation en proposant deux dispositifs hybrides basés sur un contrôle élastique de la position, combiné à un contrôle isomorphe de la rotation avec une trackball [54]. En effet un contrôle isomorphe de la rotation est possible, même avec un dispositif qui reste stable à une position confortable sur le bureau. L'un de leurs dispositifs (le GlobeMouse) exigeait un changement de prise pour passer du contrôle de rotation au contrôle de position, contrairement au second dispositif : le GlobeFish illustré sur la figure 2.3, au milieu. Ils ont mesuré les temps d'exécution d'une tâche de placement 6D avec une Space Mouse et avec ces deux périphériques, en utilisant les mêmes tétraèdres que Zhai. Ils ont montré un gain de 22% et 30% avec le GlobeMouse et le Globefish respectivement, comparé aux résultats du Space Mouse.

Des périphériques fixes non élastiques, les bras articulés tels que le Phantom, permettent une manipulation 6D isomorphe dans un espace de contrôle restreint, de l'ordre de 10 cm de côté. Le Phantom est illustré dans la figure 2.3, à droite. Il se manipule comme un stylo, permettant de bénéficier de la haute précision motrice des doigts. En appliquant un gain fort, voire dynamique, ce dispositif permet de manipuler des objets dans une large scène 3D sans lever le poignet de sa position de confort, comme avec une souris. La position de la main autour du stylo, avec poignet posé permet également une forte stabilité. Glessner et al. [57] ont mesuré un gain de performance de plus de 40% par rapport à l'utilisation d'une souris dans une tâche de placement 6D. Pour permettre une manipulation 6D avec la souris, une vue en quatre panneaux était affichée (avec trois

projections orthographiques et une vue en perspective), et le clic droit était utilisé pour les rotations.

Des périphériques tactiles multi-touch

Plus récemment, des travaux de recherche se sont intéressés à la manipulation 6D d'objets à l'aide de surfaces tactiles telles que l'iPad illustré dans la figure 2.4, à gauche. Ces surfaces permettent à l'utilisateur un contrôle 2D avec la position d'un doigt. Les technologies utilisées pour ces surfaces sont capables de détecter plusieurs positions en même temps (multi-touch), chaque doigt en plus offrant un nouveau contrôle 2D. Ainsi, un contrôle 4D est possible avec deux doigts, et un contrôle 6D avec trois doigts.

Hancock et al. ont introduit diverses techniques d'interaction multi-touch basées sur des interactions à 1, 2 ou 3 doigts pour de la manipulation 6D [65]. La fonction de transfert proposée pour permettre un contrôle de la position 3D et de l'orientation 3D de l'objet à partir de trois positions 2D est assez simple : le premier doigt contrôle les deux premiers degrés de liberté de la position ; le second doigt contrôle le dernier degré de liberté de la position, et le premier de l'orientation ; le troisième doigt contrôle les deux derniers degrés de liberté de l'orientation. Pour permettre un contrôle 6D à partir d'un seul doigt, les auteurs utilisent une fonction de transfert bien plus complexe basée sur la technique du doigt collant qui impose notamment une rotation de l'objet autour de son centre à l'instant où l'utilisateur touche la surface. Cette fonction de transfert nécessite plusieurs touchers successifs pour pouvoir résoudre une tâche 6D. Ils ont évalué les trois techniques proposées dans une tâche de placement 6D. Ils ont mesuré un gain de performance de 29% avec la technique à trois doigts par rapport à la technique à un doigt (la moins performante). En analysant le temps passé à toucher la surface, ils se sont rendu compte qu'une grande partie du temps passé à résoudre la tâche avec la technique à un doigt n'était pas consacrée à interagir avec la surface, mais à réfléchir à l'interaction qui devait être effectuée pour atteindre le résultat voulu. Les auteurs concluent que la fonction de transfert complexe de la technique à un doigt engendre une charge cognitive d'utilisation importante, incluant des tâches de rotations mentales. La technique à trois doigts semble être plus facile à utiliser, avec pour résultat de meilleures performances.

Kin et al. ont proposé des techniques d'interaction tactiles basées sur l'utilisation de plusieurs doigts, avec les deux mains, pour construire des scènes 3D organiques, telles que celles qui sont utilisées pour des films d'animation [82]. Ils proposent de permettre des contrôles différents en fonction du nombre de doigts utilisés dans différentes zones de l'espace moteur. Une zone est ainsi réservée à la main non dominante, permettant de dif-



FIGURE 2.4: Techniques d'interaction tactile multi-touch. À gauche : un I-pad Air de Apple. À droite : des rotations 2D mappées sur les déplacements 2D d'un doigt lorsqu'un autre doigt reste fixe, technique introduite par Liu et al. (source : [91]).

férencier une interaction à deux doigts de la main dominante, d'une interaction à un doigt par main. De cette manière, il est possible d'alterner de manière particulièrement fluide et intuitive entre différents modes de manipulation 2D. Après avoir interrogé des professionnels de l'animation 3D, les auteurs conseillent de séparer les tâches qui n'ont pas un besoin capital d'être effectuées simultanément, plutôt que de permettre des manipulations plus complexes mais cognitivement plus lourdes.

Liu et al. ont proposé une technique de manipulation 6D à deux doigts grâce à deux modes permettant de contrôler différemment l'objet si l'utilisateur bouge les deux doigts en même temps (permettant un contrôle 4D), ou si un seul doigt est déplacé pendant que le second reste fixe (permettant le contrôle des deux derniers degrés de liberté, comme cela est illustré dans la figure 2.4 à droite) [91]. Ils obtiennent des performances comparables avec les meilleures techniques à trois doigts de la littérature dans une tâche de placement 6D. Malgré l'utilisation d'un espace de contrôle 4D, cette technique bénéficie d'une fonction de transfert simple à comprendre et demandant une charge cognitive moins importante pour être utilisée car ne nécessitant de coordonner que deux mouvements au lieu de trois.

Si les périphériques fixes présentés ci-dessus exploitent la finesse du contrôle moteur des doigts pour bénéficier au mieux des capacités des utilisateurs, il semblerait qu'exploiter un contrôle facile à comprendre et à réaliser permette également de bénéficier au mieux des capacités mentales des utilisateurs, pour un contrôle 6D d'objets plus performant.

De meilleures performances à main levée avec des systèmes de capture optique

L'apparition des technologies de suivi de mouvement optique a permis de développer des techniques d'interaction à main levée nettement plus précises et stables que les technologies électromagnétiques. Ces technologies permettent également de suivre la position de nombreux objets physiques ou parties du corps de l'utilisateur sans nécessiter de porter d'équipement, grâce à des techniques logicielles d'analyse d'images, telle celle que proposent Wang et al. [159]. Leur solution tourne à une fréquence de 20 Hz et permet de détecter des gestes de pincement avec une précision de 99%. Les technologies les plus performantes, quant à elles, rendent possible la détection de marqueurs légers et peu invasifs, en atteignant des précisions jusqu'à 0.02 mm à des fréquences supérieures à 240 Hz. Nous avons également mesuré une précision de 0.2 mm à une fréquence de 90 Hz pour le HTC Vive, qui intègre des contrôleurs 6D pour chaque main. Un contrôleur est illustré sur la figure 2.5, à gauche.

Dans une étude de placement 6D, Vuibert et al. ont comparé trois techniques d'interaction à main levée avec un bras articulé [157]. Les trois techniques à main levée différaient

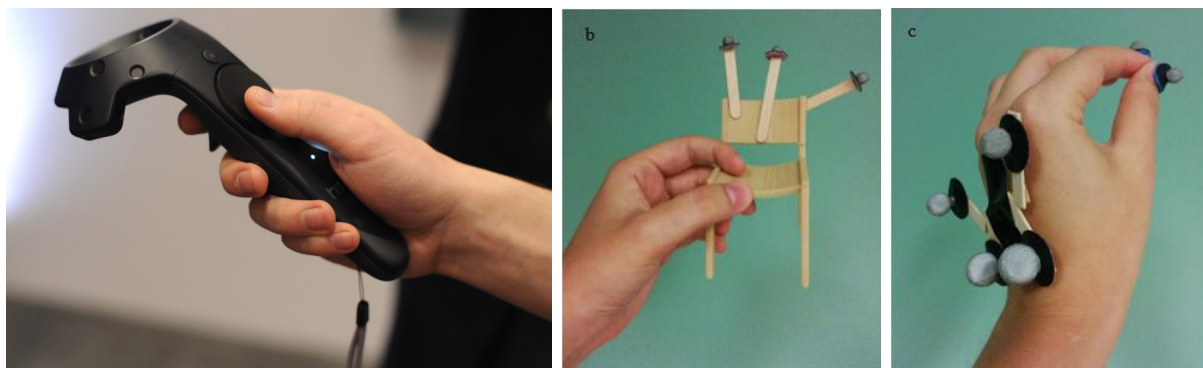


FIGURE 2.5: Techniques d'interaction à main levée avec systèmes de capture optique. À gauche : le contrôleur du Vive de HTC, avec plusieurs boutons. Au milieu : interagir en manipulant une représentation physique de l'objet virtuel (source : [157]). À droite : manipuler en bougeant sa main, sans périphérique d'entrée. (source : [157]).

de l'objet physique manipulé pour contrôler la position de l'objet virtuel : la première technique demandait de manipuler un stylo ; la seconde, une reproduction de l'objet virtuel ; la troisième se basait directement sur le mouvement du pouce et de l'index de l'utilisateur. Ces deux dernières techniques sont illustrées sur la figure 2.5, au milieu et à droite respectivement. Au lieu d'utiliser le tétraèdre de Zhai, Vuibert et al. ont constaté que l'utilisation d'un modèle 3D plus courant tel qu'une chaise réduisait la complexité perceptuelle d'une tâche de placement 6D. Elles ont montré que les deux techniques utilisant le stylo et le modèle de chaise permettaient de résoudre les tâches de placement jusqu'à 23% plus rapidement que le Phantom. En contrepartie, le Phantom s'est montré 24% et 15% plus précis que la deuxième technique la plus précise, la mini chaise, pour la position et la rotation respectivement. Les techniques à main levée n'ont pas été jugées plus fatigantes que l'utilisation du phantom par les participants. Ces résultats témoignent d'un compromis précision-vitesse qui nécessite d'être pris en compte dans ce genre de tâches.

Besançon et al. ont comparé trois paradigmes d'interaction totalement différents dans une tâche de placement 6D : une technique souris + clavier, une interaction multi-touch et une interaction tangible [15]. Le contrôle 6D a été implémenté à la souris avec plusieurs modes activables grâce à des raccourcis clavier et aux deux boutons de la souris. L'interaction tactile était également implémentée via deux modes : un contrôle 2D avec un doigt et un contrôle 4D avec deux doigts. L'interaction tangible permettait un contrôle isomorphe de l'objet virtuel en manipulant un dodécaèdre passif, traqué avec un système de suivi optique. Ils ont trouvé que l'interaction isomorphe était la plus rapide avec de très larges marges de 45% et 57% par rapport à l'interaction tactile et à la souris respectivement, tandis que toutes les conditions offraient le même niveau de précision. De plus, ils ont constaté que l'interaction isomorphe permettait d'atteindre les performances optimales quasiment instantanément (dès le premier essai), tandis que les performances des deux autres techniques, lors des premiers essais, étaient largement moins bonnes que par la suite, témoignant d'une nécessité de prendre en main ces interactions.

2.1.3 Bénéfices d'un contrôle isomorphe

Les techniques à main levée présentées ci-dessus permettent un déplacement isomorphe et naturel des objets : le mouvement 6D appliqué par l'utilisateur au périphérique de contrôle est reproduit "tel quel" sur l'objet numérique sélectionné. De cette façon, les tâches de placement peuvent être exécutées en effectuant les mouvements que nous ferions naturellement pour déplacer et orienter des objets physiques, si tant est que les retours visuels produits par le système permettent une coordination visuo-motrice efficace. On peut même s'attendre à ce que nos vastes compétences de manipulation, acquises dans le monde physique au cours de notre vie, se transfèrent ainsi automatiquement à l'environnement virtuel, permettant une interaction particulièrement intuitive.

Il est vrai que certaines techniques d'interaction non isomorphe peuvent se montrer plus efficaces dans certains contextes, comme la souris utilisée avec un gain dynamique pour interagir dans un espace 2D. Poupyrev et al. ont également observé qu'un contrôle non isomorphe de rotations 3D était 13% plus efficace qu'un contrôle isomorphe [115]. Cependant, ils ont démontré que des fonctions de transfert non isomorphes peuvent briser des propriétés d'utilisabilité importantes de l'interaction : la conformité directionnelle et la conformité nulle.

Un contrôle isomorphe permet également d'activer l'utilisation de la proprioception. Mine et al. définissent la proprioception comme "le sens qu'a une personne de la position et de l'orientation de son corps et de ses membres" et ils plaident fortement en faveur de l'utilisation de la proprioception dans des environnements virtuels [104]. La proprioception nous aide à attraper nos clés dans notre poche à l'aveugle, et fait preuve de toute son

utilité pour des mouvements ciblés.

Les techniques d'interaction haute-fidélité, tels les casques de réalité virtuelle ou augmentée, suppriment encore un niveau d'indirection dans l'interaction en affichant les objets virtuels au même endroit que le périphérique de contrôle utilisé. Cette superposition d'informations devrait permettre une interaction plus intuitive et plus performante comme il a déjà été observé par Ware et al. dans le cas d'une tâche de rotation [161].

2.1.4 Bilan

Plusieurs approches ont été testées pour permettre un contrôle performant dans des tâches de placement 6D. Les plus communes tendent à séparer la tâche en sous-tâches 2D ou 3D, exécutables à partir de plusieurs modes. Des périphériques 6D fixes permettent un contrôle précis et confortable en exploitant la finesse du contrôle moteur des doigts, tandis que des techniques tangibles ou à main levée permettent d'effectuer des manipulations moins précises mais plus intuitives et plus faciles à réaliser. Cette approche basée sur des techniques toujours plus haute fidélité semble particulièrement prometteuse malgré des performances souvent limitées par les capacités technologiques.

Une limite importante de ces études est que le temps de réalisation de la tâche n'a jamais été lié à son niveau de difficulté. Pour des tâches 2D, l'exploitation de la loi de Fitts [53] (qui lie le temps d'acquisition d'une tâche à son amplitude et la largeur de sa cible) a permis de comparer les performances de différentes techniques dans des contextes expérimentaux différents. En 6D, seuls Zhai et al. ont fait varier l'amplitude de la tâche (et pas la taille de la cible), mais ils n'ont pas rapporté comment cela avait influencé les temps d'acquisition [168]. En conséquence, il est difficile de comparer les différentes interactions testées dans ces études. Nous n'avons pas trouvé dans la littérature d'indice de difficulté similaire à celui qui est utilisé en 2D pour des tâches de placement 6D. Une norme serait indispensable pour pouvoir généraliser et comparer les différents résultats de la littérature, ou pour estimer un indice de performance des interactions testées. Nous proposons en section 3.1 un indice de difficulté 6D allant dans ce sens.

Nous n'avons pas non plus trouvé d'étude montrant la plus-value de la proprioception dans une tâche de placement 6D, bien que certains indices laissent penser qu'elle pourrait être bénéfique pour des interactions naturelles à échelle humaine : des interactions proches de celles qui profitent de la proprioception dans la réalité physique. Enfin, bien que l'utilisation d'un APCD pour du placement dans l'espace ait été envisagée par Spindler et al. [141], son efficacité n'a jamais été formellement testée dans la littérature. Nous présentons en section 3.2 une étude utilisateur qui montre leur supériorité dans une tâche de placement 6D désignée pour mettre en jeu les bénéfices de la proprioception.

2.2 Un Apprentissage plus efficace

Certaines théories de l'apprentissage indiquent que toute tâche d'apprentissage induit une charge cognitive à l'apprenant [146]. Pour que l'apprentissage soit un succès, cette charge cognitive ne doit pas surcharger la mémoire de travail de l'apprenant. La charge cognitive totale d'un apprentissage est la somme de deux composantes : la charge cognitive intrinsèque de la tâche et la charge cognitive extrinsèque imposée par l'utilisation du support nécessaire à l'apprentissage. Si la charge cognitive intrinsèque est directement liée aux notions qui doivent être apprises, la charge cognitive extrinsèque dépend de la technique utilisée par l'apprenant pour consulter et interagir avec les notions à apprendre. L'utilisation de différentes techniques d'interaction, lors d'un apprentissage, peut faire varier la proportion de mémoire de travail nécessaire pour interagir avec le contenu d'une leçon, et ainsi modifier la proportion de mémoire de travail disponible pour la tâche d'apprentissage en elle-même. L'utilisation d'une technique d'interaction particulièrement adaptée à la leçon permet ainsi d'apprendre des notions plus complexes, plus facilement. Les techniques HF, permettant d'interagir naturellement pourraient ainsi permettre d'apprendre des notions 3D avec une charge cognitive extrinsèque moindre.

Dans la première partie de cette section, nous allons voir que la technique d'interaction la plus adaptée à une tâche d'apprentissage n'est pas forcément la plus réaliste ni la plus efficace en terme d'interaction pure, et dépend beaucoup de la notion à apprendre. Nous présentons ensuite une tâche particulière : l'*apprentissage d'objets 3D*, qui semble pouvoir bénéficier de l'utilisation d'une technique d'interaction 3D, compte tenu de sa forte demande en capacité de *rotation mentale* (RM) chez les apprenants, soit leur capacité à imaginer des rotations d'un objet sans avoir celui-ci dans les mains. Enfin, nous présentons l'état des recherches qui tendent vers l'utilisation de techniques d'interaction de plus en plus fidèles pour réduire la charge cognitive extrinsèque dans des tâches d'apprentissage d'objets 3D.

2.2.1 Adapter l'interaction aux besoins de la tâche d'apprentissage

Comme nous l'avons vu en section 2.1, un dispositif permettant de manipuler des objets 3D efficacement peut induire une charge cognitive extrinsèque élevée, alors que visualiser passivement des images 2D n'induit à fortiori aucune charge cognitive de manipulation. À l'inverse, comprendre une animation 3D portant sur des mouvements musculaires à partir de schéma en 2D et de flèches demande aux apprenants un travail de reconstruction mentale à partir d'informations abstraites. Ce travail mental peut induire une charge cognitive extrinsèque importante qui pourrait être évitée en présentant une visualisation 3D animée du mouvement en question.

Dans une étude récente sur les effets de la réalité virtuelle (RV) immersive dans l'apprentissage d'un cours de science biochimique, Makranskya et al. mentionnent que "la réalité virtuelle devrait créer un changement de paradigme dans l'éducation et la formation, mais il y a peu de preuves empiriques de sa valeur éducative" ("Virtual reality is predicted to create a paradigm shift in education and training, but there is little empirical evidence of its educational value") [99]. En effet, ils ont comparé une simulation sur ordinateur standard par rapport à une simulation utilisant un casque de RV dans une tâche d'apprentissage de biochimie portant sur les mammifères. Les deux techniques étudiées permettaient de visualiser un laboratoire virtuel et de parcourir des procédures expérimentales virtuelles telles que la culture cellulaire en actionnant les équipements correspondants. L'activation des différentes étapes d'une procédure s'effectuait en cliquant

sur les éléments concernés, soit via une souris classique, soit avec un pointeur fixe au centre du champ de vision de l'utilisateur et en cliquant sur un touchpad présent sur le côté du casque de RV. Les auteurs ont mesuré un apprentissage moins efficace des étudiants avec la condition RV (mesuré par un test de connaissance en fin de session), ainsi qu'une charge cognitive significativement plus élevée. La charge cognitive a été mesurée par l'intermédiaire d'un électroencéphalographe (EEG). Avec les deux techniques, l'interaction avec les éléments de la simulation se faisait via un pointage 2D, et la qualité de l'apprentissage dépendait de la compréhension du fonctionnement des outils virtuels, plus que de la qualité de leur représentation. Il semblerait que dans ce cas, la représentation stéréoscopique n'ait pas simplifié la compréhension des outils, tandis que le pointage par mouvement de la tête s'est avéré plus lourd cognitivement à utiliser qu'un pointage à la souris.

Dans une autre étude, Parong et al. ont comparé l'efficacité pédagogique de la réalité virtuelle immersive par rapport à l'utilisation d'un diaporama comme support pour enseigner une leçon de biologie sur le fonctionnement du système circulatoire du corps humain [108]. Les résultats ont montré que les élèves qui ont visionné le diaporama ont obtenu de bien meilleurs résultats au post-test que le groupe RV. Dans cette étude, la condition RV était une condition HF mettant en jeu un casque de RV (HTC Vive) permettant de se déplacer librement dans la pièce expérimentale et de manipuler intuitivement des objets d'intérêt tels que des cellules. Les notions à apprendre, en revanche demandaient une compréhension fonctionnelle du système circulatoire et non une bonne représentation 3D des différents éléments. Les auteurs supposent que la perte de performance avec la condition RV est due à une distraction de la leçon orale causée par les différents éléments visuels perpétuellement animés. Ces éléments visuels animés, tel l'écoulement de cellules dans un vaisseau sanguin, étaient pensés pour favoriser l'immersion, mais n'apportaient que peu d'informations sur les notions à apprendre. La surcharge d'informations visuelles intrusives et d'informations auditives est ici la source de l'augmentation de la charge cognitive extrinsèque lors de l'apprentissage.

Dans ces deux études, les notions apprises n'étaient pas spécifiquement liées à l'amélioration de la perception 3D offerte par l'interaction RV. Pourtant, les environnements de RV et de réalité augmentée (RA) offrent un affichage stéréoscopique couplé à la position de l'utilisateur, qui s'est avéré offrir une meilleure perception de profondeur. Cela pourrait favoriser l'apprentissage lorsque la perception 3D est essentielle à la tâche d'apprentissage. En outre, l'exploration active s'est avérée bénéfique pour des tâches de reconnaissance visuelle par rapport à une observation plus guidée (comme l'utilisation de vidéos ou d'images statiques) [66]. En cas d'exploration active, Hinckley et al. ont montré que la charge cognitive extrinsèque des environnements de RV et de RA est réduite si les objets virtuels sont manipulés avec un contrôle isomorphe, qui permet un contrôle intuitif et efficace des rotations [70]. L'utilisation de systèmes HF pourrait donc se montrer bénéfique pour une tâche d'apprentissage qui met en jeu la reconnaissance visuelle d'objets 3D.

2.2.2 Apprendre un objet 3D

Dans ce manuscrit, nous désignons par **apprentissage d'un objet 3D**, le processus cognitif de création d'une représentation mentale de la forme d'un objet initialement inconnu. Cette représentation mentale peut ensuite être utilisée pour reconnaître l'objet dans une scène, quelle que soit son orientation et malgré des occlusions partielles.

Les étudiants en anatomie doivent apprendre des objets 3D (tels que des os et des organes). C'est également le cas pour beaucoup d'autres étudiants en sciences, technologie,

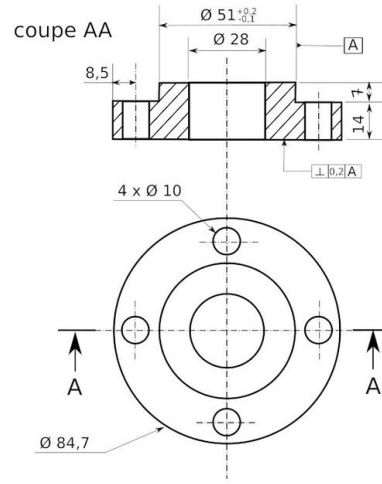


FIGURE 2.6: À gauche : une image d'IRM de genou. À droite : un dessin technique au trait.

ingénierie et mathématiques (STIM). Traditionnellement, les objets 3D doivent être appris à partir d'un ensemble de représentations statiques telles que des images ou des dessins, comme l'illustre la figure 2.6. Cette tâche est connue pour sa forte demande en capacités de rotation mentale.

Lee et al. [88] ont prédit il y a 10 ans que la réalité virtuelle serait un outil prometteur pour améliorer les capacités de RM des étudiants et leurs applications dans divers cours. Ils ont fait l'hypothèse que le rendu plus réaliste et la manipulation plus intuitive fournis par les techniques HF devraient faciliter la construction d'une représentation mentale de l'objet en réduisant la charge cognitive extrinsèque de la tâche. Cette hypothèse n'a pas encore pu être formellement testée.

Une demande en capacités de rotation mentale

Une exigence importante pour les étudiants en médecine et en STIM est d'être capable de se représenter des objets intrinsèquement 3D : des organes ou des pièces mécaniques 3D, à partir de points de vue 2D statiques comme des radiographies ou des dessins au trait. Apprendre à identifier des objets dépend fortement de la capacité de rotation mentale des élèves. Il a été démontré que les capacités de RM améliorent l'apprentissage de diverses connaissances, incluant l'anatomie fonctionnelle, la géométrie spatiale, la chimie, la chirurgie, l'architecture et la conception technique [72, 158]. Une exigence courante de ces domaines est de pouvoir représenter et manipuler mentalement des objets 3D.

Guillot et al. [64] ont suggéré que le score aux tests de RM tels que le test de rotation mentale de Vandenberg (VMRT) [154] prédit le succès de l'apprentissage de l'anatomie.

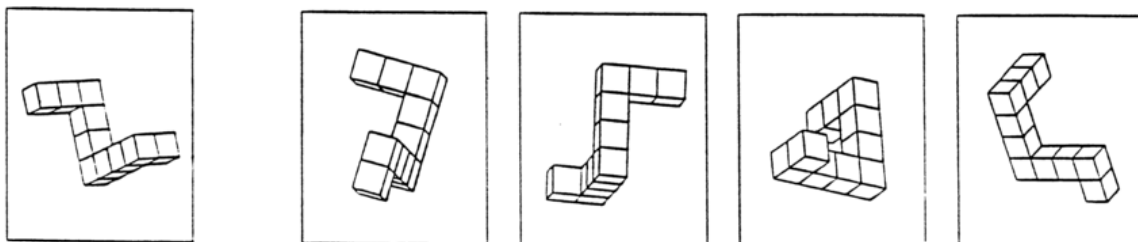


FIGURE 2.7: Un élément du test de rotation mentale de Vandenberg (VMRT). Le participant doit trouver deux images parmi les quatre de droite qui représentent le même objet que celui qui est affiché sur l'image de gauche, vu à partir de différents points de vue.

Un élément du VMRT est illustré dans la figure 2.7. Il a été observé que certains étudiants échouent à leurs examens en raison d'une faible capacité de RM et plus généralement de faibles capacités spatiales [123, 158]. Comme la RM varie considérablement d'une personne à l'autre, les étudiants à faible capacité de RM sont pénalisés dans les cours de STIM.

Améliorer les capacités de rotation mentale

En conséquence, les chercheurs ont étudié les moyens d'améliorer les capacités de RM des étudiants et l'effet de cette amélioration sur les scores des étudiants dans les tests de STIM. Hoyek et al. [72] ont démontré que l'entraînement des capacités de RM pouvait améliorer les performances des étudiants dans les tests d'anatomie. Passig et al. [110] ont testé les différences d'amélioration de la RM des enfants lors de la manipulation d'objets 3D dans un environnement en RV ou d'objets 2D avec un ordinateur de bureau. Ils ont constaté que l'utilisation de la RV améliorerait considérablement les capacités de RM des participants. Plus récemment, Contero et al. [39] ont montré que les leçons basées sur des exercices de RA sur ordinateur améliorent significativement les capacités de RM des utilisateurs après un programme de formation de 8 heures. Ils ont utilisé un système de RA basse fidélité qui affiche sur un écran plat classique une vue de webcam augmentée. Différents exercices qu'ils ont mis en jeu sont illustrés dans la figure 2.8.

Bien que l'entraînement des capacités de RM puisse améliorer l'apprentissage des objets 3D, ce processus peut être long, fastidieux et peu motivant. Une autre approche pour réduire les inégalités entre les apprenants consiste à réduire l'effet des capacités de RM sur l'apprentissage d'objets 3D en exploitant des systèmes interactifs plus fidèles.

Réduire la demande en capacités de rotation mentale

Guillot et al. [64] ont identifié le fait que la RM joue un rôle majeur dans la construction et la mémorisation de la représentation mentale d'un objet 3D, tout particulièrement lorsque seules des vues en perspective statique sont utilisées (comme des images dans un livre technique). Ils suggèrent que l'utilisation de représentations plus fidèles de l'objet comme des vidéos animées, des modèles physiques ou des données virtuelles 3D, peut réduire l'impact des capacités de RM lors de l'apprentissage d'un objet 3D.

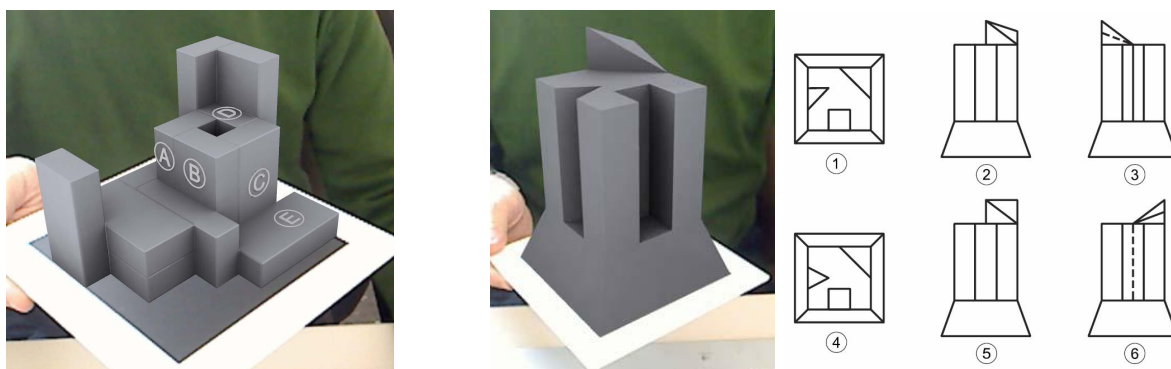


FIGURE 2.8: Des exercices proposés par Contero et al. pour entraîner les capacités de RM en utilisant un système de RA sur ordinateur. L'utilisateur manipule une feuille sur laquelle le modèle 3D est ajouté virtuellement. À gauche, les prismes marqués des lettres A à E doivent être identifiés sur une vue 2D au trait. À droite, les vues 2D correspondant au modèle 3D manipulé doivent être identifiées parmi un ensemble de vues disponibles (source : [39]).

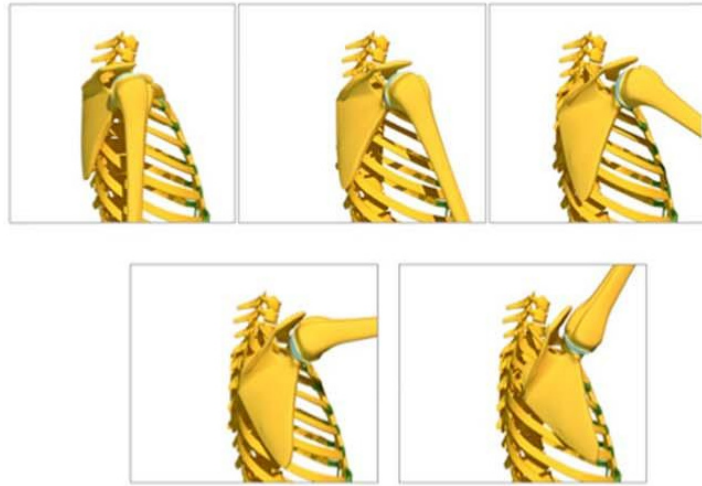


FIGURE 2.9: Une visualisation statique proposée par Berney et al., affichant cinq images du mouvement de l’omoplate. Ces images sont extraites d’une visualisation dynamique affichant la totalité du mouvement dans une vidéo de 118 secondes (source : [13]).

Hoyek et al. ont observé des améliorations dans leur enseignement en anatomie avec l’utilisation de visualisations animées sous forme de vidéos, à la place de visualisations statiques sous forme de diaporamas. [71]. Cependant, l’avantage observé peut être lié aux capacités de rotation mentale de l’apprenant : Berney et al. ont étudié les avantages d’une visualisation dynamique par rapport à une visualisation statique pour apprendre l’anatomie fonctionnelle de l’omoplate et de ses mouvements [13]. La visualisation statique consistait en l’affichage de plusieurs images clefs extraites de la visualisation dynamique, dont un exemple est illustré dans la figure 2.9. Ils ont constaté qu’une visualisation dynamique aidait les apprenants ayant de faibles capacités de RM à créer une représentation mentale de l’omoplate, tandis que les élèves ayant des capacités de RM plus élevées apprenaient mieux avec la visualisation statique. Meijer et al. ont fait des observations similaires en étudiant l’impact de l’exploration active pour apprendre des objets composés de plusieurs geons de Biederman [102]. Les objets étaient visualisés sur un écran plat standard et devaient être tournés de 180° , soit automatiquement (exploration passive), soit manuellement à l’aide d’une souris (exploration active). Ils ont montré que les personnes ayant de faibles capacités de RM ont grandement bénéficié de l’exploration active tandis que les personnes ayant les meilleures capacités de RM apprenaient l’objet sans avoir besoin d’une manipulation active.

Les études mentionnées dans ce paragraphe n’ont utilisé que des représentations basse fidélité d’objets physiques, mais les auteurs ont tout de même pu observer des améliorations de l’apprentissage par l’utilisation d’animations ou d’interactions avec l’objet.

2.2.3 Des techniques d’interaction plus fidèles pour mieux apprendre un objet 3D

Exploiter les bénéfices de la réalité virtuelle

Les systèmes interactifs 3D haute-fidélité offrent à l’apprenant la possibilité d’interagir avec une représentation plus fidèle de l’objet appris. Schnabel et al. [126] ont exploré comment les environnements immersifs affectent la façon dont les gens créent un modèle mental d’un volume spatial. Les volumes étudiés étaient des formes cubiques composées d’un assemblage de huit cuboïdes, de formes et de couleurs différentes. Ils ont testé un environnement virtuel stéréoscopique immersif utilisant un casque de RV à côté d’un environnement de bureau non immersif affichant des représentations 2D. Le système de



FIGURE 2.10: À gauche : le dispositif de RV utilisé par Schnabel et al. et un volume cubique à analyser (source : [126]). Au milieu : L'environnement de RV de dissection de grenouille utilisé par Lee et al. (source : [87]). À droite : le dispositif de RV (Dextroscope) utilisé par Jang et al. (source : [77]).

RV ainsi que les volumes utilisés sont illustrés dans la figure 2.10, à gauche. Les auteurs ont constaté que les participants utilisant des systèmes de RV avaient une meilleure compréhension du volume 3D et de ses composants que les participants utilisant des représentations 2D.

Lee et Wong [87] ont étudié l'impact d'un environnement d'apprentissage en RV pour enseigner l'anatomie de la grenouille, en comparaison avec une leçon sous forme de diaporama. L'environnement virtuel, illustré sur la figure 2.10, au milieu, était accessible sur un ordinateur de bureau classique. Les élèves pouvaient déplacer la caméra, utiliser un outil de scalpel virtuel pour découper les différents tissus et obtenir des informations correspondant aux parties sélectionnées à l'aide d'un outil d'information. Ils avaient également accès à un outil d'endoscopie. Les mêmes images et les mêmes textes étaient utilisés dans le diaporama pour le groupe contrôle. Les élèves ayant de faibles capacités de RM ont grandement bénéficié de l'aspect intuitif et implicatif de l'environnement de RV. Les auteurs affirment que recréer mentalement un objet 3D à partir de représentations 2D produit une charge cognitive extrinsèque qui peut surcharger la mémoire de travail des élèves, en particulier pour ceux qui ont du mal à effectuer des tâches mentales de transformation spatiale.

Jang et al. ont comparé l'exploration active et passive dans une tâche d'apprentissage de la structure anatomique de l'oreille interne, avec une configuration en RV [77]. Le système de RV mis en jeu est un Dextroscope (illustré dans la figure 2.10, à droite), qui offre une visualisation stéréoscopique et une manipulation 6D isomorphe du modèle virtuel grâce à un contrôleur à main levée. Des binômes d'étudiants ont été séparés : le premier étudiant pouvait manipuler le modèle 3D librement pour l'étudier, tandis que le second voyait une rediffusion stéréoscopique exacte de ce qu'avait vu le premier. Les tests utilisés pour mesurer la qualité de l'apprentissage consistaient en une série de dessins en 2D dont une partie de la structure était manquante et devait être dessinée par les étudiants. Les résultats obtenus sont similaires à ceux de Berney et al. [13] : les élèves ayant de faibles capacités de RM ont grandement profité de l'exploration active, contrairement aux étudiants ayant des capacités de RM élevées.

Les études en réalité augmentée jusqu'à aujourd'hui

Les systèmes de réalité virtuelle séparent les utilisateurs de leur environnement, ce qui peut être préjudiciable dans un contexte d'enseignement où les étudiants ont besoin

d'échanger avec leur enseignant ou de collaborer avec d'autres étudiants. Cet argument justifie l'utilisation de systèmes de réalité augmentée qui permettent d'afficher des représentations haute-fidélité sans couper l'étudiant de son environnement physique d'apprentissage. À ce jour, peu d'études ont réussi à démontrer un avantage de la RA dans un environnement d'enseignement. La mise en œuvre d'un système de RA robuste et convaincant est indispensable pour pouvoir être utilisé en contexte écologique. Un tel système doit être à la fois suffisamment précis, avoir une haute résolution d'affichage, une fréquence de rafraîchissement élevée et une faible latence. À l'heure actuelle, implémenter un système de RA qui réponde à ces critères reste un défi par rapport à l'implémentation d'un système de RV similaire. Ces limitations technologiques peuvent expliquer ce manque de littérature.

En 2011, Chen et al. ont testé l'utilisation de modèles tangibles et de modèles en RA dans un cours d'ingénierie graphique [34]. L'environnement de RA de bureau affichait la vue d'une webcam sur l'écran en y incrustant des modèles 3D au niveau de QRcodes imprimés que les étudiants pouvaient manipuler librement devant la webcam, comme cela est illustré sur la figure 2.11, à gauche. Alors que les étudiants ont accueilli avec enthousiasme l'utilisation des modèles en RA, Chen et al. n'ont observé qu'une mince amélioration de la capacité des élèves à reconnaître des objets 3D sur des images 2D, alors que l'utilisation d'un modèle tangible a considérablement augmenté leurs performances. Cependant, le système de RA ne rend pas toutes les rotations du modèle possible, et les auteurs ont signalé qu'il était parfois utilisé de manière inappropriée. L'utilisation d'un système de RA plus fidèle, permettant un contrôle du modèle virtuel plus proche de la façon dont nous manipulons les objets physiques, aurait pu aboutir à des performances plus proches de celles du modèle tangible.

Shin et al. [131] ont étudié comment un observateur perçoit un objet 3D complexe dans une scène en RA en fonction de la méthode utilisée pour changer son point de vue. Pour simuler un casque de RA, ils ont attaché une caméra faible latence à l'avant d'un casque opaque, comme l'illustre la figure 2.11, à droite. Le casque affichait l'image de la caméra pour chaque œil en y incrustant les informations virtuelles. Pour changer son point de vue, l'observateur pouvait soit se déplacer autour de l'objet, soit tourner l'objet lui-même. Les objets observés représentaient des réseaux de tuyauterie plus ou moins complexes. Ils ont montré que le déplacement autour de l'objet était plus efficace pour le percevoir que la rotation de l'objet lui-même. Ces résultats mettent en évidence les avantages du couplage

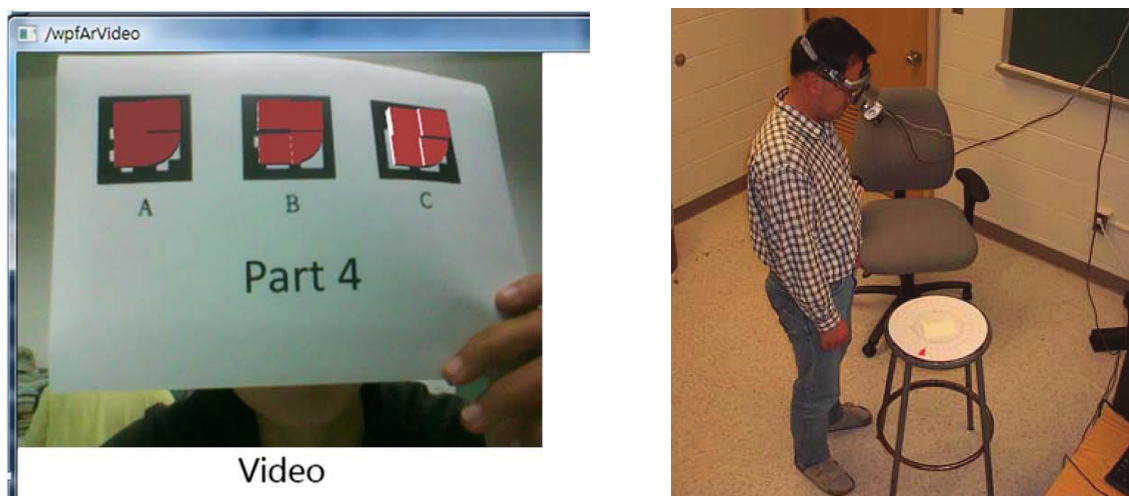


FIGURE 2.11: À gauche : La visualisation de RA sur ordinateur utilisée par Chen et al. (source : [34]). À droite : Le dispositif de RA utilisé par Shin et al., implémenté avec un casque opaque et une caméra faible latence (source : [131]).

de l’affichage en fonction de la position de l’utilisateur lors de l’interaction avec des scènes 3D.

Exploiter la haute fidélité pour apprendre en manipulant

L’étude des technologies de RV et de RA haute-fidélité est une piste intéressante pour permettre d’apprendre efficacement des objets 3D en les explorant aussi naturellement que possible. De plus, la pratique de nombreuses manipulations d’objets 3D dans l’espace pourrait être une méthode intuitive et facile pour entraîner la RM, car elle nécessite que l’utilisateur explore et fasse pivoter les objets de manière active. En effet, la théorie de la *cognition ancrée* [8] suggère que des réactivations futures de l’expérience sensorimotrice avec l’objet aideraient à créer des stratégies de rotation mentale incarnées qui pourraient se généraliser à des objets similaires. De cette façon, l’exploitation des systèmes de RV et de RA haute-fidélité récents dans des cours de STIM pourrait améliorer à la fois la rotation mentale des élèves et l’apprentissage des notions concrètes enseignées.

2.2.4 Bilan

Alors que les environnements de RV et de RA semblent pouvoir renforcer la motivation et l’intérêt des étudiants, les preuves empiriques de meilleurs résultats d’apprentissage sont difficiles à trouver dans la littérature. En effet, proposer une visualisation 3D n’est pas forcément bénéfique pour toute tâche d’apprentissage, d’autant plus si l’interface est surchargée d’informations et d’animations. De plus, les limites technologiques ne permettent pas toujours de proposer des solutions interactives fluides et intuitives. Malgré tout, certaines études récentes ont pu révéler un avantage concernant des cours en RV ou en RA comportant certaines tâches d’apprentissage d’objets 3D spécifiques. L’apprentissage d’objets 3D est une tâche importante dans de nombreux cours de STIM, et pour laquelle les étudiants souffrent d’inégalités de par sa forte demande en rotation mentale. S’il a été montré que les représentations en 3D et l’exploration active peuvent réduire la demande en RM et la charge cognitive extrinsèque d’une tâche d’apprentissage d’objets 3D, il n’y a toujours pas de preuve formelle de l’impact du niveau de fidélité d’un système interactif sur la reconnaissance des objets. De plus, nous n’avons connaissance d’aucune étude incluant une comparaison d’une condition de RV et d’une condition de RA dans ce contexte d’apprentissage.

2.3 Un meilleur sentiment de présence

Le concept de présence est d'abord apparu dans le domaine de la réalité virtuelle comme un sentiment éprouvé en réponse à un système immersif. On retrouve des concepts similaires dans différents domaines. Nous commençons donc cette section d'état de l'art en introduisant les différentes définitions de la présence de la littérature. Nous présentons ensuite différents domaines applicatifs qui tirent profit d'un bon sentiment de présence, et qui justifient de s'intéresser à ce sentiment. Il semblerait que la présence ne puisse être ressentie que par l'intermédiaire de systèmes HF, mais la question de savoir comment optimiser ce sentiment reste ouverte, en partie parce que les moyens de mesurer la présence ne font pas l'unanimité dans la communauté scientifique. Dans la suite de cette section, nous présentons les pistes de la littérature concernant l'optimisation de la présence ainsi que les limites des moyens existants pour mesurer le sentiment de présence.

2.3.1 Les définitions de la présence dans la littérature

Réalité virtuelle

La présence a d'abord été introduite dans le domaine de la réalité virtuelle comme le sentiment psychologique "d'être là" ("being there") dans l'environnement virtuel (Slater et al. [138]). Elle décrit l'état psychologique humain lorsqu'il est exposé à un système de RV immersif. Les systèmes de RV immersifs visent à remplacer entièrement l'environnement physique des utilisateurs par un environnement virtuel. Pour que l'utilisateur puisse se sentir présent dans une réalité qui lui est complètement étrangère, le système doit, pour le plus grand nombre possible de ses récepteurs sensoriels, soit fournir un stimuli haute résolution qui sature le récepteur, soit bloquer le récepteur.

Télé-coopération

En téléconférence et télé-coopération, le terme télé-présence fait référence au sentiment d'être physiquement là dans un espace de travail distant [130]. Dans ce cas, les utilisateurs veulent se sentir comme s'ils étaient présents dans une réalité physique à distance. L'attention étant essentiellement portée sur la collaboration, l'environnement devient superflu : seule la coprésence des utilisateurs avec leurs collègues distants est essentielle [127]. Ici, l'accent est mis sur un sentiment de présence envers des individus sociaux, et non envers une réalité entière. Il a été démontré que le réalisme des avatars des collègues a un impact positif sur le sentiment de coprésence ainsi que sur la compréhension de la communication non verbale entre les utilisateurs. Bailenson et al. ont étudié la qualité d'une communication par téléconférence entre un participant et son interlocuteur en fonction du réalisme de l'avatar de l'interlocuteur. Ils ont mesuré le sentiment de coprésence des participants, la qualité de la communication verbale et non verbale perçue, et leurs capacités à reconnaître les émotions de leur interlocuteur [6]. Ils ont testé trois conditions : pas d'avatar, un avatar polygonal impersonnel qui reproduit les expressions faciales de l'interlocuteur, et une vidéoconférence (c'est-à-dire un avatar photo-réaliste de haute qualité). Ils ont signalé un impact négatif de l'avatar polygonal sur la coprésence et une amélioration de la communication non verbale avec les deux types d'avatars.

Réalité augmentée

La réalité augmentée permet d'incruster des objets virtuels dans la réalité physique de l'utilisateur, contrairement à la RV qui vise à sortir l'utilisateur de sa réalité physique. Un avantage important de la RA est que les applications n'ont pas à reproduire un environnement complet. Cependant, le réalisme de l'objet virtuel doit être très élevé pour que

l'illusion qu'il existe réellement dans l'environnement physique fonctionne. À ce jour, le réalisme est limité par plusieurs problèmes techniques répertoriés et classés par Kruijff et al. [84], incluant la gestion de l'occultation, la stabilité de l'ancrage de l'objet, l'éclairage et les jeux d'ombres, l'accommodation et le champ de vision limité. Chen et al. élargissent la notion de présence définie en RV pour l'appliquer en RA. Ils s'attachent à la présence des entités virtuelles dans la réalité physique de l'utilisateur au lieu de la présence de l'utilisateur dans la réalité virtuelle [35]. Ils mesurent la présence en RA en adaptant le questionnaire commun de présence de Witmer et Singer [163] au contexte de la réalité augmentée. Ici, le concept de présence renvoie au sentiment, pour les utilisateurs, que l'objet virtuel est avec eux dans leur réalité physique.

Absence de définition universelle

Chaque domaine possède sa propre littérature et sa propre définition de la présence. Pourtant, il semblerait que toutes ces définitions soient liées à un même sentiment, ressenti inconsciemment par les utilisateurs dans leur réalité physique, et que les chercheurs essaient de reproduire avec des environnement virtuels. L'absence d'une définition universelle regroupant tous ces concepts est un frein qui a tendance à isoler les différents domaines de recherche. Utiliser une définition moins spécifique, plus englobante, telle que la définition que nous avons proposée dans la section 1.1.4, permettrait de supprimer ce frein.

2.3.2 Les applications de la présence

Intuitivité

Avec un haut niveau de présence, une entité virtuelle est perçue comme une entité physique, favorisant un transfert des compétences acquises dans la réalité physique vers l'environnement virtuel. En effet, Bérard et Rochet-Capellan proposent une définition du caractère *naturel* d'une interaction qui s'appuie sur la possibilité d'un transfert de compétences entre les réalités physique et virtuelle [27]. Ce transfert de compétences rend l'interaction plus intuitive (en réduisant le temps d'apprentissage) et plus efficace. Jacob et al. ont remarqué que de nombreux systèmes d'interaction récents cherchent à tirer parti des habitudes de l'utilisateur pour réduire l'écart entre l'intention d'un utilisateur et

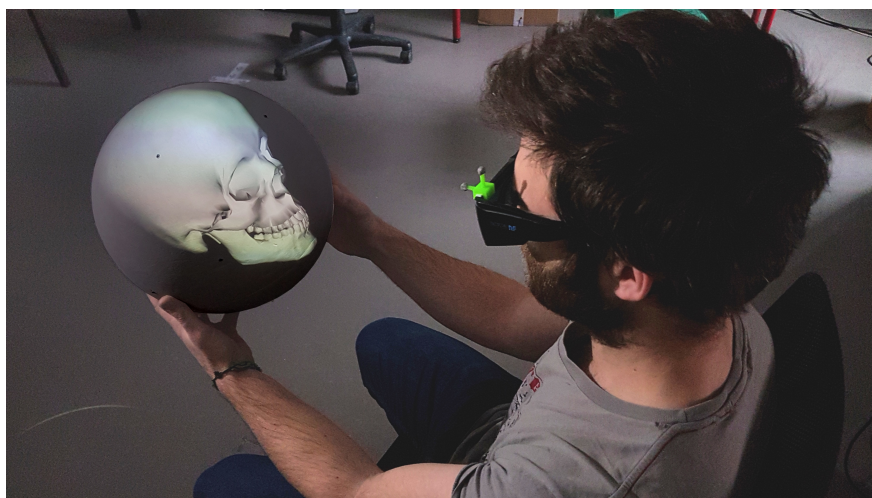


FIGURE 2.12: Un utilisateur manipulant intuitivement un crâne virtuel, qu'il perçoit avec un fort sentiment de présence dans la sphère physique. Il peut retourner la sphère pour voir l'arrière du crâne.

les moyens d'exécuter l'interaction correspondante [76]. Ils regroupent ces systèmes en les qualifiant d'interaction basée sur la réalité (Reality-Based Interaction) et affirment que ces systèmes émergents puisent leur inspiration dans la réalité physique. La figure 2.12 illustre comment un utilisateur peut interagir naturellement avec un objet virtuel en interagissant de la même manière que s'il s'agissait d'un objet physique. Cette intuition d'interaction vient naturellement s'il ressent la présence de l'objet virtuel dans sa réalité physique.

Réponses émotionnelles

L'amélioration de la présence favorise l'implication des utilisateurs dans une expérience virtuelle. Une forte présence peut également conduire à des réactions émotionnelles similaires à celles que les utilisateurs auraient dans un contexte physique. Dans une revue de littérature faite en 2001, Schuemie et al. ont recensé plusieurs études en RV qui montrent une corrélation entre la présence mesurée par questionnaire et des réponses émotionnelles telles que l'acrophobie (phobie de l'altitude) et la glossophobie (phobie de parler en public) [128]. Plus récemment, Diemer et al. ont mis en évidence une relation entre la présence et des émotions à haute activation physiologique telles que la peur et l'excitation : ils ont trouvé plus de 30 études qui rapportent des corrélations entre la présence et l'expérience émotionnelle sur les vingt dernières années [47]. Cette littérature montre qu'une forte présence dans un environnement de RV immersif tend à augmenter les réponses émotionnelles.

Générer des émotions à partir d'une expérience en RV a une application importante dans le secteur de la santé : la thérapie par l'exposition pour traiter les phobies. Parsons et al. ont fait une méta-revue de nombreuses études sur la thérapie par l'exposition en RV [109]. Ils reportent que la thérapie par l'exposition est plus efficace lorsque l'exposition se fait in vivo plutôt que par l'imagination (par exemple via l'hypnose) dans le traitement de phobies spécifiques. Cette méta-analyse suggère que la thérapie par l'exposition en RV serait efficace pour réduire l'anxiété des patients. Price et al. ont analysé le comportement de trente-six patients souffrant d'agoraphobie (phobie de la foule) ou d'acrophobie (phobie de l'altitude) au cours d'une thérapie par l'exposition clinique en RV [118]. Une exposition en RV à l'acrophobie est représentée dans la figure 2.13. Les auteurs ont trouvé des corrélations entre la présence signalée et l'anxiété phobique. Ils concluent que se sentir présent dans l'environnement de RV semble nécessaire au succès de la thérapie par l'exposition mais peut ne pas être suffisant. Ces résultats suggèrent que les patients qui atteignent un niveau élevé de présence pendant la thérapie en RV sont capables de se sentir physique-



FIGURE 2.13: Un utilisateur est virtuellement exposé à son acrophobie.

ment exposés à leur phobie, ce qui est un facteur crucial pour le succès de la thérapie. Une thérapie par l'exposition a également été mise en œuvre en RA dans des cas de phobie des petits animaux. Selon Wrzesien et al., la proximité du patient, et surtout de ses mains, avec les stimuli est primordiale dans la thérapie [164]. Créer un avatar qui reproduit les mouvements des mains de l'utilisateur suffisamment réaliste pour permettre une grande présence est un défi très difficile en RV. L'approche de la RA élimine cette difficulté, car les utilisateurs perçoivent leurs bras et mains réels. Dans leur étude, Wrzesien et al. ont testé la réponse de douze patients souffrant de phobie des petits animaux à une thérapie de l'exposition via la RA ou in vivo. Les résultats indiquent une réduction comparable de l'anxiété phobique chez les patients des deux groupes après le traitement, soutenant l'efficacité de la réalité augmentée en tant que traitement.

Environnements virtuels d'apprentissage

Chen et al. ont étudié l'effet de la présence dans un contexte d'apprentissage en réalité augmentée. Ils ont testé deux systèmes de RA dans une expérience inter-sujets : un casque semi-transparent et un système écran+webcam. [35]. Lors d'une leçon présentant le concept de la structure des protéines, ils ont testé une classe de 58 étudiants universitaires inscrits en cours de chimie organique. Les résultats suggèrent que les élèves qui ont connu un niveau élevé de présence ont subi une charge cognitive plus faible lors de l'interaction avec le système de RA. De plus, les auteurs ont signalé une corrélation entre la présence et l'implication dans le cours.

Dans des domaines tels que l'aéronautique, mettre des professionnels en situation réelle afin de les entraîner peut s'avérer très coûteux et dangereux, comme par exemple entraîner un pilote d'avion à réagir en cas de panne de réacteur. L'utilisation de simulateur virtuel tel que celui qui est illustré dans la figure 2.14 est alors indispensable dans la formation. Pour faciliter le transfert de connaissances entre le simulateur virtuel et la situation réelle, il est important que la simulation soit très fidèle à la réalité. Vora et al. ont étudié l'utilisation d'un simulateur virtuel pour entraîner des techniciens à la maintenance des avions [156]. Elles ont comparé une simulation sur ordinateur traditionnel et une simulation immersive avec un casque de réalité virtuelle, dans une tâche d'inspection visuelle de défauts de l'avion. Elles ont constaté que le simulateur haute-fidélité permettait aux techniciens de repérer une plus grande fraction des défauts, plus rapidement. Dans une autre expérience, elles ont montré une corrélation entre les performances d'inspection et la présence ressentie dans l'environnement de RV. Elles ont relevé que les techniciens qui s'en sortaient le mieux



FIGURE 2.14: Des pilotes sont entraînés par l'intermédiaire d'un simulateur de vol.



FIGURE 2.15: Interface pour personnifier son personnage dans le jeu vidéo Monster Hunter World développé par Capcom (2018).

ressentaient très peu de différence entre l’entraînement virtuel et l’entraînement physique.

Jeux vidéo

Les mesures de performance les plus utilisées en IHM ne s’appliquent pas aux jeux vidéo, car les utilisateurs ont des intérêts différents. Traditionnellement, les chercheurs en IHM visent à améliorer les performances des utilisateurs lors de la réalisation de différentes tâches, tandis que les utilisateurs de jeux vidéo cherchent à vivre des émotions et relever des défis intéressants. Le modèle flux-canal (flow-channel model) est fréquemment utilisé pour évaluer la qualité des jeux vidéo. Ce concept a été introduit pour la première fois par Csikszentmihalyi en 1975 [42]. Un flux optimal se produit dans une situation où les compétences requises et les défis proposés sont équilibrés. La théorie fournit un questionnaire de soixante-dix éléments pour analyser les retours subjectifs des joueurs. Takatalo et al. ont appliqué le modèle pour évaluer deux jeux vidéo de tir à la première personne populaires, en analysant les réponses de plus d’une centaine de joueurs experts. Ils ont soutenu l’importance de la présence comme condition préalable qui favorise un état optimal de flux [147]. De nombreux jeux vidéo permettent aux joueurs de personnaliser leur avatar ou le héros incarné, afin de favoriser leur identification à leur personnage, ou pour augmenter leur empathie envers le héros incarné. Cela pourrait, in fine, augmenter leur sentiment de présence dans l’univers virtuel. Une interface de personnalisation d’avatar est illustrée dans la figure 2.15.

2.3.3 Comment produire un haut niveau de présence

Des retours sensoriels haute définition

Dans le domaine de la recherche en RV, la présence est la réponse humaine à un environnement virtuel immersif : plus le système est immersif, plus ses utilisateurs ressentent un fort sentiment de présence. Produire une expérience perceptive de haute qualité offre une meilleure immersion qui, à son tour, est susceptible d’améliorer la présence. Cependant, la qualité de l’expérience perceptive dépend de nombreux facteurs qui ne sont pas comparables entre eux. Il est difficile de prédire quel facteur doit être amélioré en priorité pour aboutir à une meilleure immersion, afin d’améliorer la présence. La qualité de la réponse visuelle, par exemple, dépend de nombreux facteurs répertoriés par Slater [133] : résolution d’affichage, champ de vision, fréquence d’affichage, stéréoscopie, résolution du

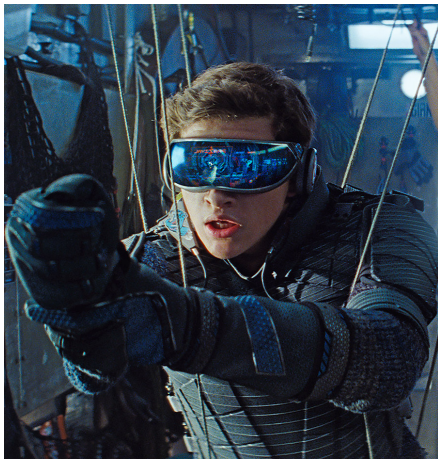


FIGURE 2.16: À gauche : un système très haute fidélité incluant un casque opaque couvrant la totalité du champ de vision et produisant des retours auditifs 3D, une combinaison complète à retour haptique, et une plateforme permettant de simuler un déplacement naturel, imaginé dans le film *Ready Player One* (2018). À droite : l’effet du ventriloque qui produit l’illusion que la voix provient de la marionnette dont les lèvres bougent.

maillage de l’entité affichée, latence totale du système, précision et stabilité de la capture de mouvement. De plus, Slater mentionne que la cohérence entre les diverses réponses sensorielles est obligatoire pour permettre la présence. Par exemple, la présence est faible si un lourd objet métallique a une représentation visuelle parfaite mais que les utilisateurs ne ressentent aucun poids lorsqu’ils le prennent dans leurs mains. La présence dépend donc forcément de la tâche : si ce même objet n’était observé qu’à distance, sa présence ne serait pas réduite par son manque de poids. Atteindre une présence parfaite avec des scènes complexes peut nécessiter des réponses visuelles de très haute qualité, une latence imperceptible, une interaction complète avec des retours haptiques généralisés, des retours auditifs en 3D, la possibilité de se déplacer naturellement dans l’environnement virtuel, etc. Il n’est pas impossible qu’un tel niveau de fidélité soit disponible un jour grâce à des périphériques tels que ceux qu’illustre le film *Ready Player One* (figure 2.16, à gauche), mais tous ces défis ont un coût et sont limités par la technologie disponible. En pratique, il est important de connaître l’effet sur la présence de chacun de ces paramètres afin de choisir le meilleur compromis. Ceci fait l’objet de notre contribution détaillée dans le chapitre 5.

Prééminence du sens visuel

Cummings et al. ont fait une méta-analyse de plus de quatre-vingts études qui ont entrepris d’évaluer l’effet de divers paramètres sur la présence [43]. Pour effectuer l’analyse, ils ont normalisé la taille de l’effet pour tous les paramètres testés dans les études examinées. Les principaux paramètres étaient la fréquence de rafraîchissement de l’affichage, le “niveau de capture” défini comme le nombre de degrés de liberté capturés, le champ de vision, la résolution d’image, la stéréoscopie et l’utilisation du son. Ils ont signalé un fort effet sur la présence de la fréquence de rafraîchissement et du niveau de capture. Le manque de son et de qualité de l’image s’est révélé avoir un impact moindre sur la présence. Bien que cette enquête soit assez vaste, ses résultats sont basés sur des études disparates impliquant des tâches et des contextes expérimentaux différant d’une étude à l’autre. Ces résultats n’offrent pas la robustesse qu’aurait une étude utilisateur contrôlée complète, regroupant des comparaisons de l’effet de chaque paramètre. Le faible impact de la qualité du retour auditif sur la présence est compatible avec l’effet de capture visuelle : la tendance qu’a le cerveau humain à préférer faire confiance aux informations visuelles

lorsque les informations multimodales sont contradictoires [101]. Un exemple bien connu de capture visuelle est l'effet du ventriloque (illustré figure 2.16, à droite), qui produit l'illusion d'entendre la voix du mannequin parce que sa bouche bouge, bien que le son provienne du ventriloque. L'effet de capture visuelle souligne l'importance de produire une réponse visuelle de haute qualité.

Plausibilité

À un niveau d'abstraction plus élevé, il faut considérer la plausibilité du scénario présenté à l'utilisateur. Slater a défini l'illusion de plausibilité comme "l'illusion que ce qui se passe apparemment se produit réellement" [135]. En RV, l'environnement dans lequel les utilisateurs sont immergés doit être crédible pour qu'ils se sentent totalement présents. Immerger les participants dans une pièce qui ressemble à la pièce physique dans laquelle ils se trouvent est plus susceptible de favoriser une grande présence que de les immerger dans une scène de jungle sauvage. En RA, un dragon virtuel est moins susceptible d'être perçu comme présent qu'un papillon virtuel. Lors de l'interaction avec des êtres humains virtuels, la façon dont l'entité virtuelle regardera l'utilisateur et répondra aux questions est cruciale pour éviter de briser la présence. De plus, selon Diemer et al., les scénarios de RV qui engagent des émotions ont tendance à augmenter la présence [47].

APCD : une approche prometteuse

Les APCD disposent de plusieurs avantages pour favoriser la présence : comme avec la RA, les utilisateurs ne sont pas coupés de leur environnement physique, et celui-ci ne doit donc pas être reproduit numériquement. L'appareil lui-même a un sens crédible dans la réalité physique en tant que récipient transparent. En tant qu'objet physique, il fournit un retour haptique passif naturel qui fait comprendre aux utilisateurs qu'ils ne peuvent pas atteindre les entités virtuelles à l'intérieur de l'appareil. Enfin, tant que le poids de l'écran reste comparable à celui de l'entité virtuelle, le sens du toucher est rendu de manière crédible sans aucun effort de mise en œuvre.

En résumé, l'approche APCD résout d'importants problèmes de cohérence entre les réponses sensorielles sans effort de développement, tout en évitant les problèmes que rencontre la RV pour créer un avatar fidèle du corps des utilisateurs. L'affichage d'entités virtuelles avec un haut niveau de présence semble faisable, et dépend principalement de la qualité des stimuli visuels créés par l'appareil.

2.3.4 Mesurer la présence

Mesures subjectives par questionnaires

Dans une étude de la littérature en RV liée à la présence, Schuemie et al. ont analysé les différentes méthodes utilisées pour mesurer la présence [128]. Dans une écrasante majorité des études, la présence a été évaluée à l'aide de questionnaires subjectifs. La présence étant par définition une notion subjective, une auto-évaluation fait sens. Faisant écho à la diversité des définitions, de nombreux questionnaires ont été utilisés pour mesurer la présence. Les plus courants sont : le questionnaire Slater–Usoh–Steed [137], le questionnaire de présence de Witmer et Singer [163] et l'inventaire ITC du sens de présence [90].

Mesures objectives

Les méthodes objectives de mesure de la présence n'ont été explorées que dans quelques études. Slater et al. ont proposé de mesurer la façon dont les utilisateurs hiérarchisent les signaux lorsque les signaux virtuels et réels sont en conflit [136]. Ils ont mis en œuvre cette

méthode en immergeant les participants dans une réplique virtuelle de la salle expérimentale où une radio était présente. Une fois les participants immergés, la radio physique était allumée et déplacée secrètement tandis que la radio virtuelle restait immobile. Les participants devaient ensuite pointer la radio. Les résultats suggèrent une corrélation entre la présence auto-évaluée à l'aide du questionnaire Slater–Usoh–Steed et cette mesure objective. Une autre approche consiste à enregistrer la conductance cutanée des utilisateurs : plusieurs études ont remarqué une corrélation entre la présence et la conductance cutanée lorsqu'il s'agit d'émotions à haute activation physiologique comme les phobies [47]. Cependant, la conductance cutanée est connue pour être une mesure directe du stress et non de la présence. Par conséquent, elle ne peut être appliquée qu'en tant que métrique indirecte pour la présence dans des contextes spécifiques où un stress est induit en cas de niveau élevé de présence, rendant celle-ci intrusive.

Aucune mesure ne fait l'unanimité

Compte tenu de la nature subjective de la présence et de l'absence d'une mesure objective fiable, le consensus est d'utiliser les questionnaires auto-évalués. Une revue récente a fourni la preuve de ce consensus : Cummings et al. ont examiné plus de quatre-vingts études qui ont analysé l'effet des technologies immersives sur la présence [43]. Dans un souci de comparaison des résultats, ils ont choisi de rejeter les études qui n'évaluaient pas la présence en utilisant des questionnaires.

Cependant, la présence auto-évaluée par questionnaires ne fait pas l'unanimité. Slater a avancé un argument important contre la validité des questionnaires sur la présence. Dans la vie de tous les jours, la présence est un fait. C'est là où nous sommes, et nous ne nous posons pas la question de savoir si l'environnement qui nous entoure répond de manière cohérente à nos actions, parce que c'est indubitablement le cas. Comme nous n'avons jamais à évaluer la présence, Slater avance qu'il n'est pas possible d'évaluer le niveau de présence de n'importe quelle situation virtuelle dans une échelle absolue mais non définie [134]. Dans leur récente enquête sur la présence, Skarbez et al. [132] ont étudié les méthodes pour mesurer la présence qui n'impliquent pas de questionnaires subjectifs car ils "ajoutent de la complexité à la recherche sur la présence". Ils ont présenté des "méthodes comportementales" qui analysent le comportement des participants face à un scénario donné. Étant des mesures objectives, ces méthodes comblent les principales lacunes des questionnaires. Cependant, elles nécessitent généralement des éléments spécifiques supplémentaires pour déclencher le comportement à évaluer en fonction du scénario. En tant que telles, elles ne sont pas adaptées à la généralisation.

Les méthodes psychophysiques

Skarbez et al. présentent également les "méthodes psychophysiques" telles que des méthodes destinées à évaluer l'importance relative des différents aspects d'une expérience virtuelle sur la présence. Ces nouvelles méthodes ont été utilisées en RV pour mesurer l'impact de certains paramètres d'immersion et de cohérence de scénarios sur la présence auto-évaluée des participants. Bergstrom et al. ont mis en œuvre une telle méthode pour évaluer la contribution de quatre paramètres sur la plausibilité lors d'une performance virtuelle d'un quatuor à cordes [12]. Après avoir expérimenté la performance avec des paramètres optimaux, les participants sont exposés à un faible niveau de fidélité et sont invités à modifier les paramètres pour maximiser la plausibilité tout en minimisant le nombre de changements de paramètres. Les participants ont tendance à commencer par améliorer le rendu du regard des artistes virtuels et la crédibilité environnementale du rendu du son, suggérant l'importance de ces deux caractéristiques pour améliorer la plausibilité de la performance. Les méthodes psychophysiques semblent prometteuses pour

étudier différents facteurs de l'expérience virtuelle et leurs effets relatifs sur la présence.

2.3.5 Bilan

De nombreux travaux de la littérature ont montré que le sentiment de présence ressentie par les utilisateurs lorsqu'ils utilisent des systèmes d'interaction de RV ou de RA haute-fidélité a des effets bénéfiques pour de nombreux domaines d'application tels que les thérapies par l'exposition, les simulateurs d'entraînement, ou encore les jeux vidéo. Cependant, peu de travaux ont été capables d'identifier les caractéristiques techniques qui impactent le plus la présence, notamment par manque de mesure assez précise et assez fiable. Cette connaissance permettrait d'orienter les recherches vers l'amélioration des paramètres les plus à même d'optimiser le sentiment de présence.

Conclusion

L'étude de la littérature nous a permis d'identifier trois axes de recherche prometteurs dans l'analyse des bénéfices des systèmes interactifs haute-fidélité : la manipulation 6D d'objets dans l'espace, l'apprentissage d'objets 3D, et la présence ressentie par les utilisateurs.

Les techniques d'interaction HF les plus récentes, incluant les APCD, n'ont pas encore été testées dans une tâche de placement 6D, et le manque de normes pour caractériser ces tâches limite la comparaison entre différentes techniques HF. Notre contribution vis-à-vis de cet axe de recherche est d'apporter des preuves quantitatives des bénéfices de techniques d'interaction haute-fidélité récentes dans une tâche de placement 6D, et de proposer une approche générale pour quantifier cette performance. Cette contribution est exposée dans le chapitre 3.

La littérature suggère que l'utilisation de techniques d'interaction particulièrement réalistes et intuitives pourrait améliorer l'apprentissage d'objets 3D, notamment en réduisant le besoin de rotation mentale qui ajoute une surcharge cognitive extrinsèque à la tâche d'apprentissage. Cependant aucune preuve empirique d'un bénéfice de la HF pour de l'apprentissage d'objets 3D n'a encore pu être apportée. Cette recherche de matériel empirique fait l'objet de notre seconde contribution, présentée dans le chapitre 4.

La présence est quant à elle une expression de la haute fidélité par définition, et ses nombreux avantages ont été identifiés dans la littérature. Cependant, ce sentiment reste encore mal compris, difficile à mesurer, et donc difficile à optimiser. Notre contribution en rapport avec la présence est de mesurer comment différents paramètres techniques impactent le niveau de présence ressentie par les utilisateurs. Cette contribution est exposée dans le chapitre 5.

Deuxième partie

Pourquoi

Chapitre 3

Supériorité de l'APCD dans une tâche de placement 6D

Introduction

Dans ce chapitre, nous nous focalisons sur l'une des tâches les plus fréquentes et les plus importantes pour interagir avec des scènes virtuelles 3D : le *placement* d'objets à une position et une orientation cible. Dans l'espace, il s'agit d'une tâche à 6 degrés de liberté (6D) : une combinaison d'une translation 3D et d'une orientation 3D de l'objet.

À l'heure actuelle, la plupart des logiciels d'édition 3D professionnels implémentent le placement à partir d'une séquence de mouvements 2D de la souris, en décomposant la tâche de placement 6D en plusieurs sous-tâches. Pourtant, dans notre réalité physique, placer un objet semble être une tâche triviale que nous sommes capables d'effectuer sans avoir besoin de décomposer le mouvement, comme pour mettre une tasse dans le lave-vaisselle en la retournant. Faire appel à un système d'interaction haute-fidélité (HF) pour manipuler des objets virtuels le plus naturellement possible devrait donc permettre d'effectuer efficacement des tâches de placement 6D.

Malgré tout, nous avons trouvé peu d'études qui aient quantifié empiriquement les performances de placement 6D de systèmes d'interaction HF lors de notre analyse de la littérature (c.f. section 2.1). À notre connaissance, aucune étude n'a encore mis en œuvre un appareil à perspective corrigée déplaçable (APCD) dans une telle tâche. Ce manque de résultats s'explique en partie par la difficulté technique que représentait encore l'implémentation de systèmes d'interaction HF crédibles il y a seulement quelques années, mais également par le manque de norme pour modéliser les performances de placement 6D des systèmes d'interaction.

La première contribution de cette thèse consiste en une étude utilisateur contrôlée visant à quantifier l'efficacité des utilisateurs dans une tâche de placement 6D effectuée avec deux techniques d'interaction HF : notre prototype d'APCD sphérique, et la même sphère utilisée comme périphérique d'entrée avec un casque de réalité virtuelle (RV). Nous comparons ces performances avec celles obtenues avec deux techniques d'interaction basse fidélité. Ces deux techniques emploient un écran plat standard pour l'affichage, l'une d'entre elles utilise également la sphère comme périphérique d'entrée indirect, tandis que l'autre utilise un bras articulé 6D (Phantom Omni). Cette expérimentation montre une amélioration importante (17%) des performances des utilisateurs dans une tâche de placement lorsqu'ils utilisent un APCD sphérique par rapport à l'utilisation de la deuxième interaction la plus efficace : le casque de RV. Elle montre également une amélioration de 43% par rapport au bras articulé, qui était auparavant considéré comme l'une des techniques d'interaction les plus efficaces pour le placement 6D d'objets.

Nous introduisons également une nouvelle approche pour quantifier la difficulté d'une tâche de placement 6D. Cette approche combine les tolérances en translation et en orientation de la tâche et permet le calcul d'une métrique unique, l'*indice de performance* (IP), que nous estimons pour les 4 techniques d'interaction testées.

Dans ce chapitre, nous commençons par présenter notre nouvelle approche pour estimer la difficulté d'une tâche de placement 6D (section 3.1). Nous présentons ensuite l'étude utilisateur que nous avons menée pour mesurer les performances de placement des quatre techniques que nous avons testées, ainsi que les résultats obtenus (section 3.2). Enfin, nous discutons de la validité de l'indice de difficulté que nous avons introduit, et des principaux facteurs pouvant expliquer les importantes différences de performances mesurées (section 3.3).

Sommaire

Introduction	82
3.1 Un indice de difficulté pour les tâches de placement 6D . . .	84
3.1.1 Intérêt de la loi de Fitts	84
3.1.2 Adaptation à une tâche 6D	84
3.2 Étude utilisateur	88
3.2.1 Détails technique	88
3.2.2 Conditions expérimentales	89
3.2.3 Protocole expérimental	91
3.2.4 Résultats	93
3.3 Discussion	97
3.3.1 Validité de l'indice de difficulté	97
3.3.2 Supériorité des techniques haute-fidélité	97
3.3.3 Supériorité de l'APCD sur le casque de RV	99
Conclusion	102

3.1 Un indice de difficulté pour les tâches de placement 6D

Dans ce chapitre, nous nous concentrons sur des tâches de placement dans un espace libre où les déplacements ne sont ni contraints ni perturbés par la présence d'obstacles.

Le problème que nous traitons ici est de pouvoir quantifier le *débit d'information* qu'une technique d'interaction permet à un utilisateur de faire passer vers le système pour résoudre une tâche donnée. Il est de coutume dans le domaine de l'interaction Homme-machine d'utiliser des modèles simplifiés qui permettent de calculer efficacement une approximation assez fidèle du débit d'information. Ces modèles ne prétendent pas pour autant décrire finement le contrôle moteur de l'utilisateur. De nombreuses études ont montré l'efficacité d'un tel modèle, la loi de Fitts, pour évaluer les performances de différents systèmes interactifs dans des tâches de placement 2D. Dans cette section, nous proposons un modèle basé sur la loi de Fitts et étendu aux tâches de placement 6D.

3.1.1 Intérêt de la loi de Fitts

Les résultats de nombreuses expériences montrent que quelle que soit la technique d'interaction utilisée, le temps nécessaire pour effectuer une tâche de placement peut être fidèlement prédit en connaissant uniquement 1) l'amplitude A du mouvement à parcourir pour atteindre la cible et 2) la largeur W de la cible. Pour des tâches de placement 1D, la loi de Fitts modélise une relation linéaire entre le temps moyen TM nécessaire pour atteindre une cible et l'*indice de difficulté* (ID) de la tâche, exprimé en fonction de A et W , en *bit* d'information. L'équation 3.1 exprime la formulation de Shannon [140] de cette relation.

$$TM = a + b.ID, \quad ID = \log_2 \left(\frac{A}{W} + 1 \right) \quad (3.1)$$

De très nombreuses études utilisateurs ont mesuré des comportements cohérents avec les prédictions de cette modélisation, faisant de la loi de Fitts une norme [140] pour mesurer l'efficacité des techniques d'interaction 1D, 2D (en utilisant des cibles circulaires) et 3D (avec des cibles sphériques). Elle a également fait ses preuves pour modéliser des tâches de placement par rotation dans un plan. Enfin, cette modélisation a pour avantage de permettre le calcul d'un *indice de performance* (IP), qui représente la bande passante d'information des techniques d'interaction en bit/s, et qui ne dépend pas de la précision demandée. L'IP permet une comparaison robuste des performances des techniques d'interaction évaluées à travers les nombreux résultats de la littérature.

La loi de Fitts ne permet néanmoins pas de modéliser le comportement d'une tâche de placement 6D qui combine une translation 3D et une rotation 3D. Nous avons cherché comment adapter au mieux cette relation à une tâche 6D, afin de supporter le calcul d'un IP pour les techniques d'interaction que nous avons testées, et ainsi faciliter la comparaison de nos résultats avec les études futures.

3.1.2 Adaptation à une tâche 6D

Stoelen et Akin ont posé la première pierre pour combiner les composantes de translation et d'orientation dans un indice de difficulté unique d'une tâche de placement [144]. Cependant, ils n'ont traité que le cas de la combinaison d'une translation 1D et d'une rotation 1D pour une tâche de placement 2D théorique. De plus, la combinaison ne concernait que le calcul de l'ID et non pas la tâche elle-même : ils utilisaient deux cibles distinctes pour les composantes translationnelle et rotationnelle de la tâche. Nous nous appuyons sur

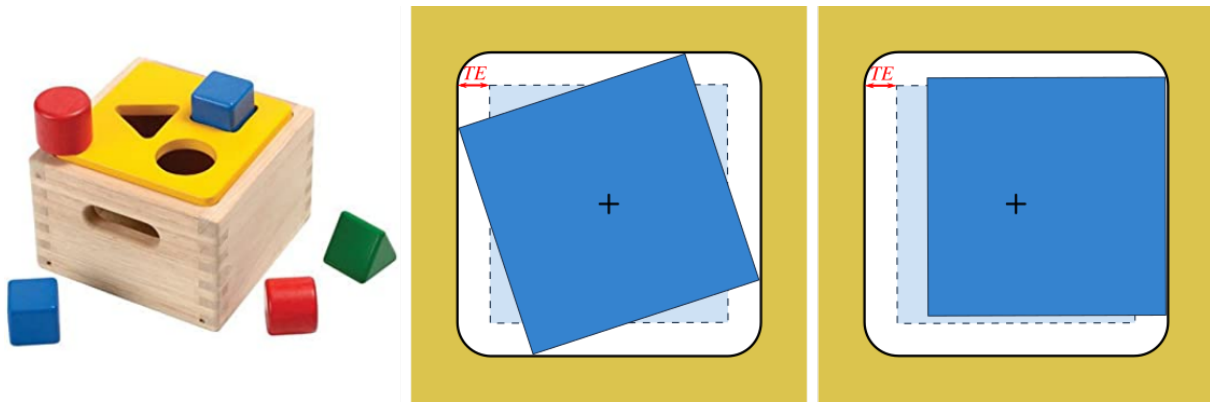


FIGURE 3.1: Illustration de la tolérance d'erreur avec l'exemple de la boîte à formes (illustrée à gauche). Au milieu : cas où toute la tolérance d'erreur (TE) est utilisée pour une erreur de rotation. À droite : cas où toute la tolérance d'erreur est utilisée pour une erreur de translation.

cette première pierre pour proposer une nouvelle formulation de l'indice de difficulté qui modélise un placement 6D complet en combinant les erreurs de translation et d'orientation effectuées lors d'une tâche.

Tolérance d'erreur

Nous utilisons une seule *tolérance spatiale* pour quantifier l'espace libre autour d'une cible : une tâche de placement 6D est réussie si la combinaison des erreurs de translation et de rotation effectuées est inférieure à sa tolérance spatiale. Nous supposons que cette modélisation est plus fidèle à la tâche écologique que l'utilisation de deux tolérances indépendantes. Pour illustrer cette supposition, imaginez une boîte à formes pour enfants (figure 3.1, à gauche) où un bloc en forme de cube doit être inséré dans un trou carré. La différence de taille entre le trou et le cube représente une seule tolérance spatiale pour cette tâche. La tâche est réussie, c'est-à-dire que le cube peut être inséré, uniquement si la projection du cube dans le plan du trou est entièrement contenue dans la limite du trou carré *quelle que soit l'erreur de rotation et/ou de position faite*. En particulier, si le cube est parfaitement centré sur le trou, alors la tolérance entière peut être utilisée pour une erreur de rotation (voir figure 3.1, au milieu), et inversement (même figure, à droite). Dans le cas plus général, les erreurs de rotation et de translation se combinent et l'insertion est possible si leur somme est plus petite que la tolérance d'erreur spatiale totale, que nous appelons TE .

Pour généraliser TE à n'importe quelle forme d'objet, nous définissons le *volume cible* d'une tâche de placement comme une expansion du volume de l'objet lorsque ce dernier est placé au niveau de la position et de l'orientation cible de la tâche. Cette expansion est calculée de sorte à inclure une marge de largeur TE tout autour du volume de l'objet. Une tâche de placement est réussie lorsque l'objet à positionner est entièrement contenu dans le volume cible, quelles que soient son erreur de translation et son erreur d'orientation par rapport aux positions et orientations cibles. Ceci est illustré dans la figure 3.2.

Indice de difficulté

Pour exprimer l'ID de la tâche, nous partons du modèle de Stoelen et Akin. Il calcule l'ID d'une tâche de placement comme la somme d'un ID de translation et d'un ID de rotation (equation 3.2, de [144]), où les IDs sont exprimés comme dans la formulation de Shannon de la loi de Fitts (equation 3.1).

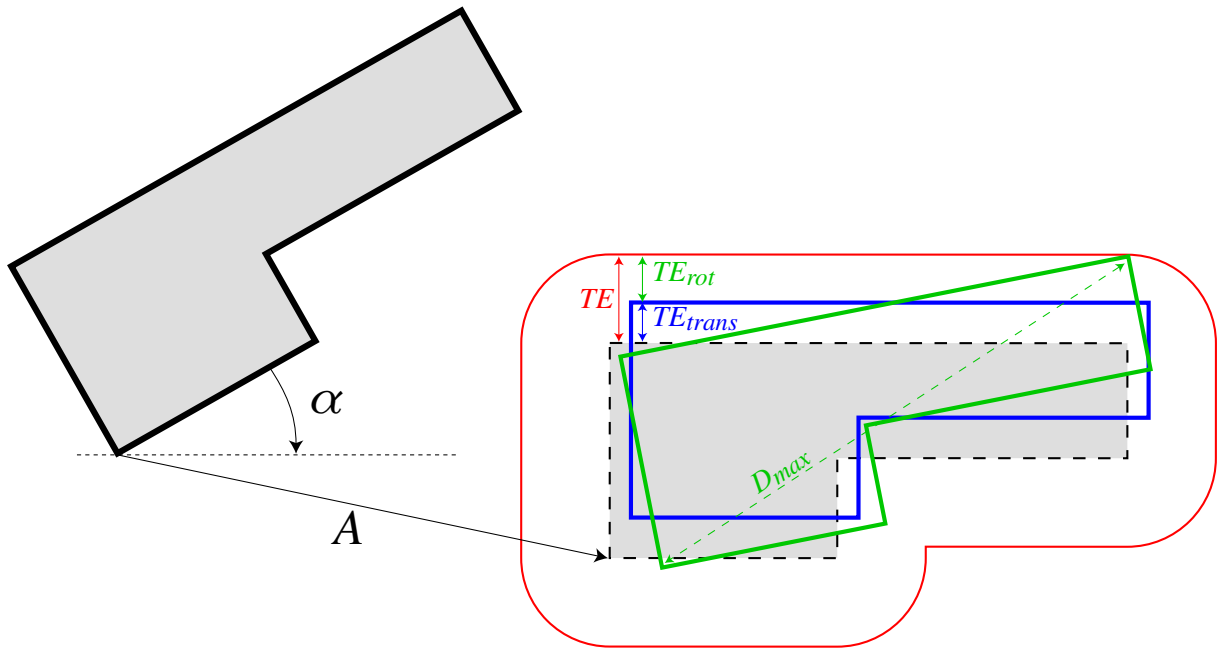


FIGURE 3.2: Paramètres d'une tâche de placement 6D d'amplitude de translation A et d'amplitude de rotation α . L'objet (forme grise avec un contour plein épais) doit être placé dans une cible (expansion rouge englobant le volume de l'objet d'une marge TE). Dans cet exemple, les composantes d'erreurs de translation (en bleu) et de rotation (en vert) sont explicitées. La tolérance d'erreur globale TE peut être approximée comme la somme de tolérances d'erreurs flottantes de translation TE_{trans} et d'orientation TE_{rot} . D_{max} est la plus grande distance entre deux points de l'objet, elle permet la conversion d'une erreur d'orientation en une erreur spatiale (détails dans le texte).

$$ID_{combine} = ID_{translation} + ID_{rotation} \quad (3.2)$$

L'équation 3.2 peut être développée en utilisant des paramètres d'amplitude et de largeur de cible pour la composante de translation (A , W) et des paramètres pour la composante de rotation (α , ω) :

$$ID_{combine} = \log_2 \left(\frac{A}{W} + 1 \right) + \log_2 \left(\frac{\alpha}{\omega} + 1 \right) \quad (3.3)$$

Dans cette équation 3.3, les amplitudes A et $alpha$ sont simplement définies comme les distances de translation et de rotation entre la position initiale de l'objet et la position cible. Les largeurs de cible W et $omega$ sont plus difficiles à déterminer. Il s'agit de les exprimer en fonction de TE . Pour cela, nous exprimons TE comme la somme d'une tolérance d'erreur de translation TE_{trans} et d'une tolérance d'erreur de rotation TE_{rot} . Cette relation est exprimée dans l'équation 3.4 :

$$TE_{trans} + TE_{rot} = TE \quad (3.4)$$

À noter que la largeur de cible W considérée dans la loi de Fitts est le double de la tolérance d'erreur TE_{trans} telle que nous l'avons définie ci-dessus. En effet, la tâche d'acquisition est réussie si le centre de l'objet n'est pas éloigné du centre de la cible de plus de la moitié de la largeur de cible W . On peut le voir aussi comme le fait que l'on dispose d'une tolérance d'erreur dans les deux sens et dans toutes les directions. En incluant cette relation entre tolérance d'erreur et largeur de cibles pour chaque composante dans l'équation 3.4, on obtient l'équation suivante :

$$\frac{W}{2} + \frac{f(\omega)}{2} = TE \quad (3.5)$$

Ou encore :

$$W + f(\omega) = 2 \times TE \quad (3.6)$$

où $f(\omega)$ convertit la largeur angulaire de cible de la composante de rotation en une tolérance d'erreur spatiale. Pour évaluer cette fonction, nous faisons l'approximation qui suppose que les utilisateurs résoudre la composante de rotation de la tâche de placement en tournant autour de l'axe optimal : celui qui permet de minimiser la quantité de rotation à effectuer. Nous utilisons un *plan de rotation optimal*, orthogonal à cet axe, pour projeter les points de l'objet à placer. La distance maximale D_{max} entre deux points projetés est utilisée pour calculer la plus grande *erreur spatiale* résultant d'une *erreur angulaire* ω . En d'autres termes, nous calculons $f(\omega)$ comme suit :

$$f(\omega) = \sin(\omega) \times D_{max} \quad (3.7)$$

Notons que l'équation 3.7 n'est valide que pour des largeurs angulaires de cible inférieures à 90° , mais ce n'est pas une contrainte pour des tâches de placement concrètes. En résumé, pour une tâche de placement particulière, il existe une infinité d'ID possibles où W et ω peuvent varier de sorte que :

$$W + \sin(\omega) \times D_{max} = 2 \times TE \quad (3.8)$$

Ici, nous résolvons numériquement la valeur minimale de $ID_{combine}$ selon les équations 3.3 et 3.8 en faisant l'approximation qui suppose que le participant visera inconsciemment la meilleure combinaison de largeurs de cible de translation et de rotation : celles qui minimisent la difficulté de la tâche. Ce système d'équations final est exprimé ci-dessous :

$$\left\{ \begin{array}{l} ID = \min(ID_{combine}(W, \omega)) \\ ID_{combine}(W, \omega) = \log_2 \left(\frac{A}{W} + 1 \right) + \log_2 \left(\frac{\alpha}{\omega} + 1 \right) \\ W + \sin(\omega) \times D_{max} = 2 \times TE \end{array} \right. \quad (3.9)$$

Les paramètres inhérents à la tâche sont indiqués en gras dans le système d'équation 3.9. Au final, nous calculons l'ID d'une tâche de placement 6D en fonction des amplitudes en translation A et en rotation α de la tâche, de la tolérance d'erreur TE de la cible et de la taille D_{max} de la projection de l'objet dans le plan de rotation optimal de la tâche. La validité de cet indice de difficulté sera discutée au regard des résultats de l'étude utilisateur.

3.2 Étude utilisateur

Nous avons mené une étude utilisateur pour mesurer les performances des participants dans des tâches de placement 6D en fonction de la technique d'interaction utilisée. Nous avons testé quatre techniques permettant un déplacement isomorphe des objets manipulés. Elles sont illustrées dans la figure 3.3. Elles consistent en : un bras articulé, une sphère tangible utilisée avec un écran plat classique, la même sphère tangible utilisée avec un casque de RV et un APCD sphérique. Nous nommons ces quatre conditions PHANTOM, PROP, RV et APCD, respectivement.

3.2.1 Détails technique

Les différences de performances techniques entre les systèmes peuvent largement impacter les performances des utilisateurs, en particulier les différences en termes de précision du périphérique d'entrée et de résolution d'affichage. Nous avons donc essayé de réduire au maximum ces différences en utilisant autant que possible les mêmes dispositifs pour la capture de mouvement et l'affichage entre les différentes conditions. Nous n'avons notamment pas utilisé le système de capture natif du casque de RV ni celui du bras articulé. La capture a été mise en œuvre dans toutes les conditions à l'aide du même système de suivi optique Optitrack, particulièrement précis, avec 10 caméras fonctionnant à 240 Hz. Ce sys-

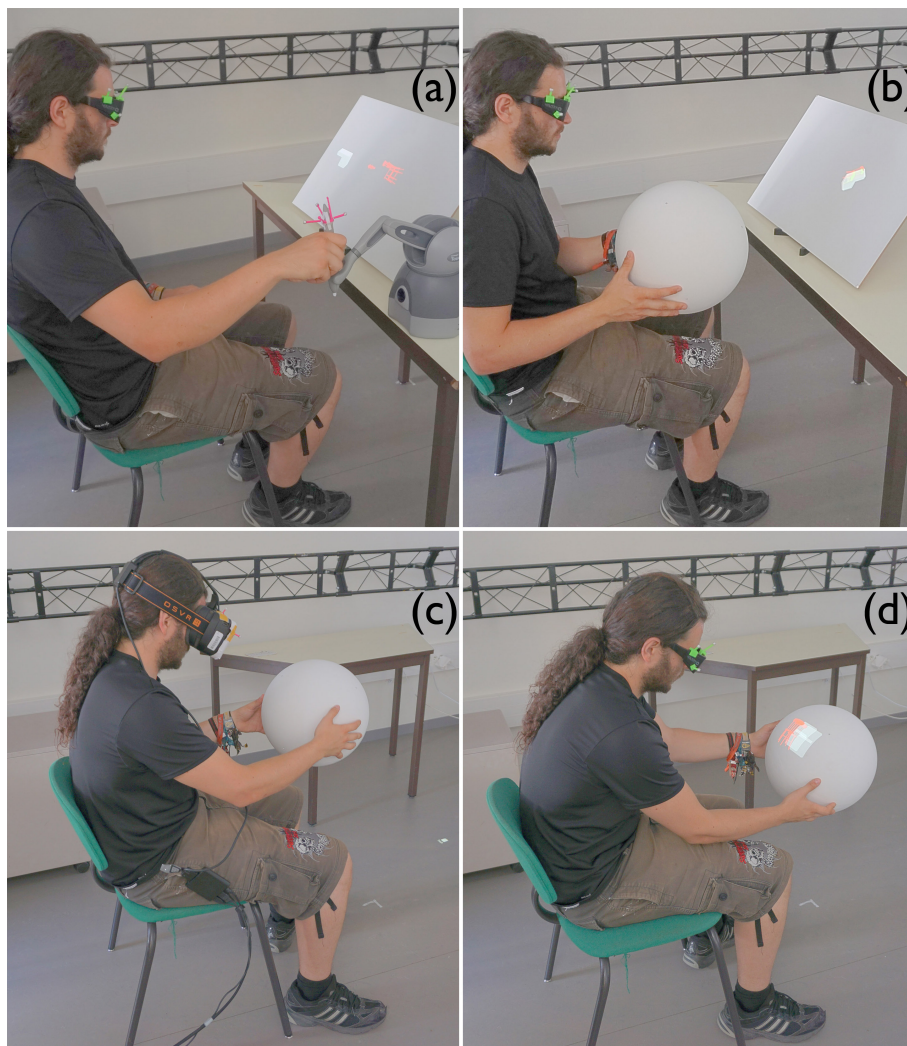


FIGURE 3.3: Les conditions expérimentales. a) le bras articulé (PHANTOM). b) l'appareil tangible (PROP). c) le casque de RV (RV). d) l'APCD (APCD).

tème a été utilisé pour suivre à la fois la position de la tête des participants (c'est-à-dire soit le casque de RV, soit les lunettes à obturateur stéréo) et la position du périphérique d'entrée. Des constellations de marqueurs ont été attachées au casque, aux lunettes à obturateur stéréo, au stylo du bras articulé et à la sphère, afin de récupérer leurs positions 6D dans l'espace.

Le retour visuel a été créé pour toutes les conditions par un projecteur de résolution 2560×1600 pixels à 120 Hz, sauf pour le casque de RV. Des lunettes à obturateur LCD, synchronisées avec le projecteur ont été utilisées pour fournir un rendu stéréoscopique à 60 Hz par œil. Dans les conditions PHANTOM et PROP, les images étaient affichées sur un écran plat fabriqué à partir d'une planche blanche de 600×375 mm (28", format 16 :10, 82 dpi). Dans la condition APCD, les images étaient affichées sur une sphère blanche en polystyrène de 30 cm de diamètre. La résolution d'image était d'environ 82 dpi, en fonction de la distance avec le projecteur. La condition RV utilisait un casque opaque OSVR HDK 2 fournissant une résolution d'affichage de 1080×1200 pixels par œil à 90 Hz, et un champ de vision de 110° . Les performances graphiques de la condition RV étaient donc différentes des autres conditions, avec une fréquence de rafraîchissement $1,5 \times$ supérieure mais une densité de pixels, d'environ 32 dpi, nettement inférieure à celle du projecteur. En effet, avec le casque, les pixels sont répartis sur un large champ de vision, si bien qu'ils étaient discernables au niveau du centre de l'attention visuelle.

La même sphère blanche jouait le rôle de périphérique d'entrée pour les trois conditions PROP, RV et APCD. La sphère était tenue à deux mains en raison de son grand diamètre. Sans fil et légère, elle n'imposait elle-même aucune contrainte en termes d'espace de travail, sauf pour la condition APCD, où l'espace de travail était limité à la zone de projection. Pour la condition PHANTOM, nous avons utilisé un bras articulé à retour de force Geomagic Touch. Le bras n'était pas alimenté et ne générait aucun retour de force, sauf du fait de sa résistance passive, qui contribuait de plus à la stabilisation du stylo. Le bras articulé offrait un espace de travail de $160 \times 120 \times 70$ mm.

Pour toutes les conditions, l'affichage était stéréoscopique et couplé à la position de la tête, afin de produire un rendu 3D fidèle.

3.2.2 Conditions expérimentales

Dans cette section, nous nous référons à la *taille de la scène* comme la distance entre les deux cibles les plus éloignées de l'expérience (c'est-à-dire le diamètre de la sphère englobant l'ensemble des cibles).

PHANTOM

Nous avons choisi d'inclure cette technique car plusieurs études ont montré son efficacité [159] et sa précision [157] pour des tâches de placement 6D. Les participants tenaient un stylo qui était attaché au bras articulé, comme illustré sur la figure 3.3, a. Le mouvement 6D de la pointe du stylo était reproduit tel quel sur l'objet contrôlé. Le bras était placé directement à droite de l'écran (ou à sa gauche pour les participants gauchers). Lors d'essais pilotes, nous avons observé que fournir un soutien physique au poignet ou à l'avant-bras améliorerait la stabilité de la manipulation mais gênait certains mouvements au cours de la séquence d'acquisition des cibles. Nous avons donc choisi de laisser un maximum d'espace libre devant l'appareil afin de ne pas entraver les mouvements de la main des participants. La taille de la scène dans l'*espace moteur* (c.à.d. le référentiel utilisé pour la manipulation) était de 97 mm, ce qui représente un sixième de la taille de la scène dans les conditions RV et APCD. Nous avons utilisé cette mise à l'échelle pour nous adapter à l'espace de travail limité de ce périphérique d'entrée, et afin que toutes les cibles puissent être acquises sans nécessiter de débrayage (l'action de repositionner la souris en la levant

pour pouvoir continuer un mouvement trop ample pour l'espace moteur disponible par exemple). L'affichage a été calculé de sorte que le centre de la scène (le barycentre de toutes les cibles) apparaisse au centre de l'écran plat. La scène dans l'*espace visuel* a été zoomée d'un facteur trois afin de maximiser l'utilisation de la surface d'affichage.

PROP

Nous avons testé cette condition afin d'isoler l'effet du périphérique d'entrée (la sphère), par rapport à l'effet des différents périphériques d'affichage : les conditions RV et APCD utilisent le même périphérique d'entrée avec des périphériques d'affichage différents. Les participants tiennent la sphère entre leurs mains pour la manipuler en translation et en rotation. Le mouvement 6D de la sphère est reproduit tel quel sur l'objet contrôlé, et la scène est affichée sur l'écran plat, comme cela est illustré dans la figure 3.3, b. La taille de la scène dans l'*espace moteur* était de 290 mm, soit la moitié de la taille par rapport aux conditions RV et APCD. Utiliser cette mise à l'échelle était nécessaire pour éviter d'imposer aux participants des postures trop inconfortables, où ils auraient dû déplacer la sphère loin de l'écran sans lâcher celui-ci des yeux, et subir ainsi une torsion fatigante de la nuque. La scène affichée dans l'*espace visuel* était la même que pour la condition PHANTOM, mais cette fois-ci, la taille de la scène dans l'espace moteur était identique à sa taille dans l'espace visuel, les deux espaces ayant une taille comparable pour cette condition.

RV

Les participants sont équipés du casque de RV et ils tiennent la sphère comme cela est illustré dans la figure 3.3, c. Afin de fournir un repère contextuel dans la scène virtuelle, nous avons ajouté un plan texturé dans la scène au niveau du sol, dont la texture était similaire à la texture du sol physique de la salle expérimentale, comme l'illustre la figure 3.4. L'espace visuel était exactement superposé à l'espace moteur, si bien que les mouvements de la sphère étaient exactement répliqués sur l'objet contrôlé, sans facteur d'échelle. La taille de la scène était de 580 mm. Toutes les cibles étaient à portée de main des participants assis. Certaines cibles nécessitaient néanmoins une rotation importante du torse des participants pour être vues et atteintes.

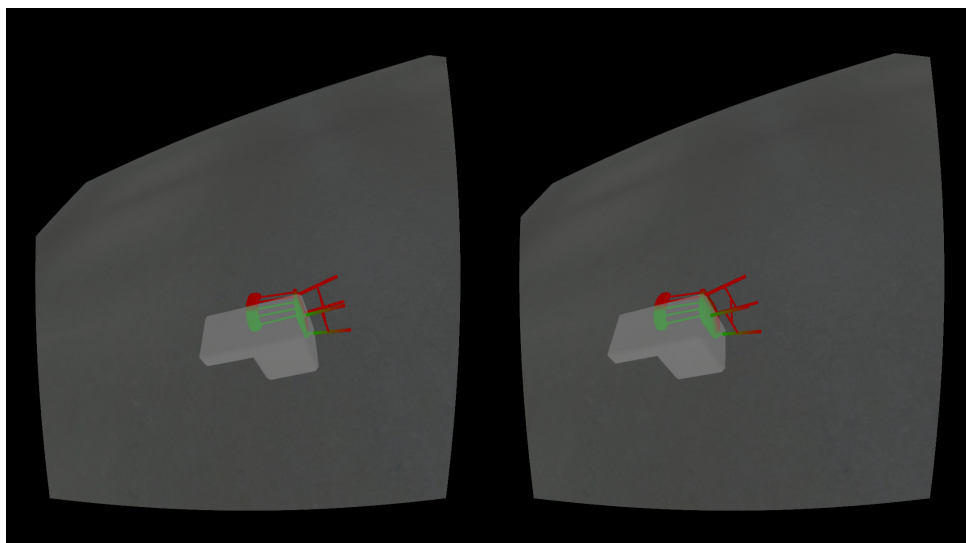


FIGURE 3.4: La scène vue au travers du casque de RV. Seul un plan texturé au niveau du sol est ajouté en plus de la chaise et de la cible semi-transparente.

APCD

Les participants visualisaient la scène virtuelle “à travers” la sphère comme illustré dans la figure 3.3, d. La scène virtuelle pour cette condition avait la même taille, et était placée de la même manière dans l’espace moteur du participant que pour la condition RV. Ainsi, les participants devaient effectuer exactement les mêmes mouvements pour atteindre les cibles dans les conditions APCD et RV.

À propos des variations d’échelle de la scène

Comme cela est décrit ci-dessus, nous avons mis à l’échelle la scène dans l’espace moteur de chaque condition afin d’optimiser l’espace utilisé. Pour les conditions RV et APCD, la scène est agrandie au maximum de la portée atteignable confortablement par les participants, car la manipulation de la sphère permet de faire des mouvements amples très rapidement. Agrandir la scène n’est donc pas un problème, et cela permet d’être moins exigeant en termes de précision, car les cibles sont plus grosses. Cette grande scène ne convient pas à la condition PHANTOM, car elle nécessiterait beaucoup de débrayage pour exécuter de grandes translations, ce qui est très peu efficace. En variant l’échelle de la scène selon les conditions, on peut donner l’impression de privilégier la validité écologique de l’expérience (les conditions sont testées à une échelle réaliste) par rapport à sa validité interne (qui suggérerait d’utiliser la même échelle avec toutes les conditions). Cependant, il est important de noter que la mise à l’échelle de la scène change à la fois l’amplitude *et* la taille des cibles dans les mêmes proportions, et donc ne change pas l’indice de difficulté de la tâche, ni le temps d’accomplissement de la tâche tel que le modélise l’équation 3.1. Par ailleurs, cette mise à l’échelle n’a aucune incidence sur la composante de rotation de la tâche de placement.

3.2.3 Protocole expérimental

Tâche

La tâche expérimentale a été implémentée de façon à pouvoir calculer l’indice de performance des conditions testées en utilisant la méthode présentée dans ce chapitre (section 3.1). Nous avons notamment mis en jeu la tolérance d’erreur TE décrite pour créer les différentes cibles. Notre étude s’est concentrée sur les qualités de *contrôle moteur* des quatre techniques d’interaction testées et non sur la difficulté de percevoir l’erreur de positionnement à corriger pour atteindre la cible. C’est pourquoi nous avons tenté de minimiser les problèmes de perception visuelle. L’objet contrôlé est un modèle 3D d’une chaise comme dans Vuibert et al. [157]. Les cibles sont représentées par des surfaces blanches semi-transparentes résultant des expansions de la chaise par diverses tolérances spatiales. La chaise est affichée en rouge lorsqu’elle est à l’extérieur de la cible. Les parties de la chaise qui se trouvent à l’intérieur de la cible deviennent vertes : l’objectif des participants est donc de rendre la chaise entièrement verte. Ce code couleur aide les participants à identifier les parties de la chaise qui sont en-dehors de la cible et à planifier un mouvement correctif. À chaque instant, la scène virtuelle contient la chaise contrôlée et le volume cible. Lorsque l’objet est loin de la cible, une flèche rouge pointant vers la cible s’affiche près de l’objet, comme l’illustre la figure 3.5.

Nous voulions observer comment chaque technique permet de tirer profit de la proprioception des participants pour améliorer les performances de placement. Par conséquent, nous avons sélectionné un nombre limité de cibles (5), et les cibles étaient toujours présentées suivant la même séquence. De cette façon, les participants ont rapidement pu

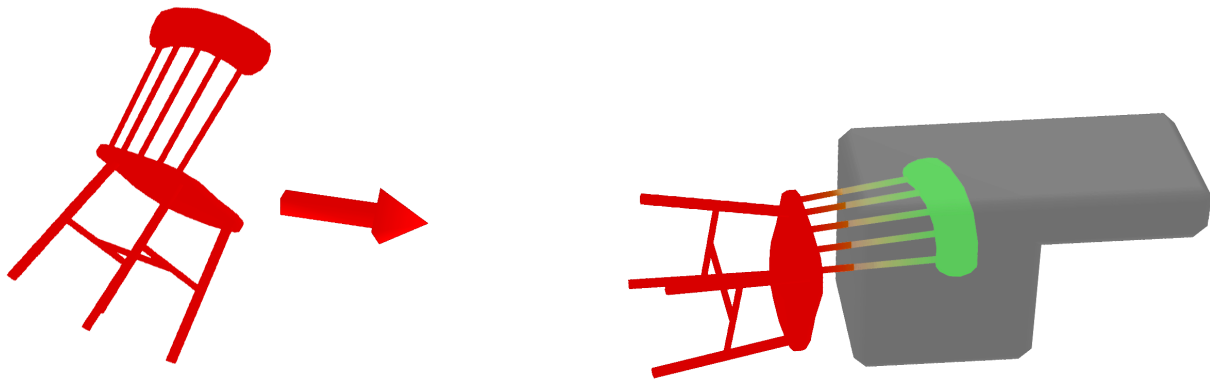


FIGURE 3.5: Visualisation de la scène lors d'une tâche de placement. Cette capture d'écran est éditée à des fins d'illustration : l'arrière-plan était noir et non pas blanc pendant l'expérience, la cible était blanche en semi-transparence et non pas grise, et une seule chaise était affichée et contrôlée en temps réel. Les parties de la chaise à l'intérieur de la cible deviennent vertes. La flèche rouge pointant vers la cible n'est affichée que lorsque la cible est éloignée.

apprendre la séquence de cibles et leurs positions 6D dans l'espace. Une fois la séquence apprise, les participants étaient en mesure d'effectuer une grande partie de la tâche suivante avec les yeux fermés en utilisant uniquement sa proprioception (même s'ils gardaient les yeux ouverts pendant l'expérience). De plus, l'impact du temps de réaction sur le temps total de résolution d'une tâche était limité car le participant pouvait anticiper le déplacement à faire pour atteindre la prochaine cible.

Aucune validation d'acquisition n'était demandée aux participants : une tâche de placement était validée dès que l'objet était entièrement dans les limites de la cible. Ce choix a pour avantage d'empêcher des mouvements indésirables causés par toute forme de validation, et d'implicitement uniformiser le compromis vitesse/précision entre les participants (par exemple, les participants ne peuvent pas perdre de temps à essayer d'être trop précis avant de valider).

Un objectif secondaire de cette expérience était de tester la validité de l'ID 6D introduit dans ce chapitre. Nous avons choisi l'ensemble des 5 cibles de sorte à générer un panel d'ID régulier, comme c'est indiqué dans la table 3.1

Méthode

Nous avons accueilli 20 participants bénévoles (seize hommes et quatre femmes) , dont deux gauchers (âge moyen 27,4 [20..38]). Nous avons utilisé le test de vision stéréoscopique RANDDOT pour vérifier que tous les participants avaient une vision stéréoscopique correcte. Aucun d'entre eux n'avait d'expérience ni avec un casque de RV ni avec le bras articulé. Un participant était habitué à la manipulation d'objets virtuels et un autre à la

ID (bit)	6.0	8.0	10.0	12.0	13.5
TE (mm)	7.2	9.1	5.8	4.2	4.3
A (mm)	54	242	203	343	580
alpha (°)	23	40	80	108	180

TABLE 3.1: Indices de difficulté utilisés dans l'expérience. Ils sont calculés en résolvant numériquement le système d'équation 3.9. A et TE sont exprimés pour les conditions RV et APCD. Ces grandeurs sont toutes deux réduites d'un facteur six pour la condition PHANTOM, et d'un facteur deux pour la condition PROP, sans que cela n'impacte l'ID.

visualisation stéréoscopique de scènes 3D. Nous avons vérifié que les résultats de ces deux participants étaient similaires à ceux des autres. Après avoir accueilli les participants, nous leur fournissions une feuille d'instructions expliquant le but de l'expérience et le protocole.

Nous avons mis en place un protocole intra-sujet. L'ordre de présentation des conditions était équilibré entre les participants. Nous avons demandé aux participants d'être aussi rapides que possible pour accomplir les tâches de placement de la chaise virtuelle. Au début d'une nouvelle condition, les participants s'entraînaient sur un bloc de 5 répétitions de la séquence de cinq cibles (25 placements), afin de se familiariser avec le nouveau dispositif. Après l'entraînement, ils effectuaient deux blocs successifs de onze répétitions de la séquence de cinq cibles. Entre chaque bloc, les participants pouvaient faire une pause s'ils le souhaitaient. La première répétition de chaque bloc évalué était considérée comme un échauffement et a été ignorée dans l'analyse. En résumé, nous avons enregistré $20 \text{ participants} \times 4 \text{ conditions} \times 2 \text{ blocs} \times 10 \text{ répétitions} \times 5 \text{ cibles} = 8000$ tâches de placement.

Nous avons évalué deux variables indépendantes : la TECHNIQUE et l'ID. Nous avons enregistré le *temps d'acquisition* comme le temps entre deux tâches réussies. Nous avons également enregistré toutes les données de capture de mouvements. À la fin de la session, les participants remplissaient un questionnaire pour fournir leur jugement subjectif sur les techniques d'interaction. Dans l'ensemble, les sessions expérimentales ont duré environ 45 minutes.

En accord avec les nouvelles préoccupations concernant l'utilisation de catégories de significativité statistique dans divers domaines de recherche [2, 14], nous rapportons, dans toutes les contributions de cette thèse, les valeurs exactes de la probabilité de l'hypothèse nulle et nous évitons d'utiliser le vocabulaire de significativité statistique.

3.2.4 Résultats

Nous avons effectué une ANOVA à mesures répétées sur le *temps d'acquisition* avec la TECHNIQUE et l'ID comme facteurs principaux. Elle a révélé un fort effet de la

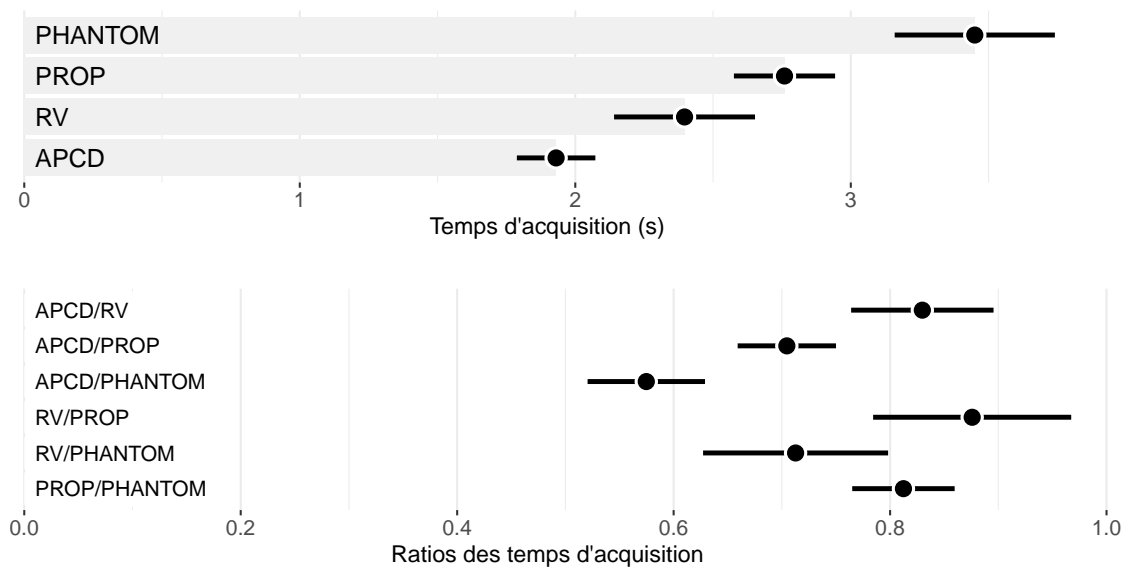


FIGURE 3.6: En haut : Le *temps d'acquisition* dans chaque condition, moyenné sur les participants et avec intervalles de confiance à 95%. En bas : ratio des *temps d'acquisition* des participants pour chaque paire de conditions, moyenné sur les participants et avec intervalle de confiance à 95%.

TECHNIQUE ($F(3, 57) = 47,2, p = 1,9e-15$), un fort effet de l'ID ($F(4, 76) = 88,6, p < 2e-16$), et une forte interaction entre les deux ($F(12, 228) = 3,85, p = 2,3e-05$).

Temps d'acquisition

Le temps d'acquisition moyen par condition, toutes cibles et participants confondus, est de 3,45 s, 2,76 s, 2,40 s et 1,93 s pour les conditions PHANTOM, PROP, RV et APCD, respectivement. Les temps d'acquisition sont représentés sur la figure 3.6, en haut. Nous avons calculé tous les ratios de temps d'acquisition des participants pour chaque paire de conditions. Ils sont représentés sur la figure 3.6, en bas. On observe des différences très nettes dans les performances de chaque technique : en moyenne, les performances des participants avec l'APCD sont 17,0% meilleures que leurs performances avec le casque de RV (RV), qui sont elles-mêmes 12,4% meilleures que leurs performances avec la condition PROP, elles-mêmes 18,8% meilleures que pour la condition PHANTOM. Tous les t-tests par paire avec correction de Holm ont révélé des différences robustes ($p \leq 0,003$ dans tous les cas, sauf pour PROP vs RV où $p = 0,013$).

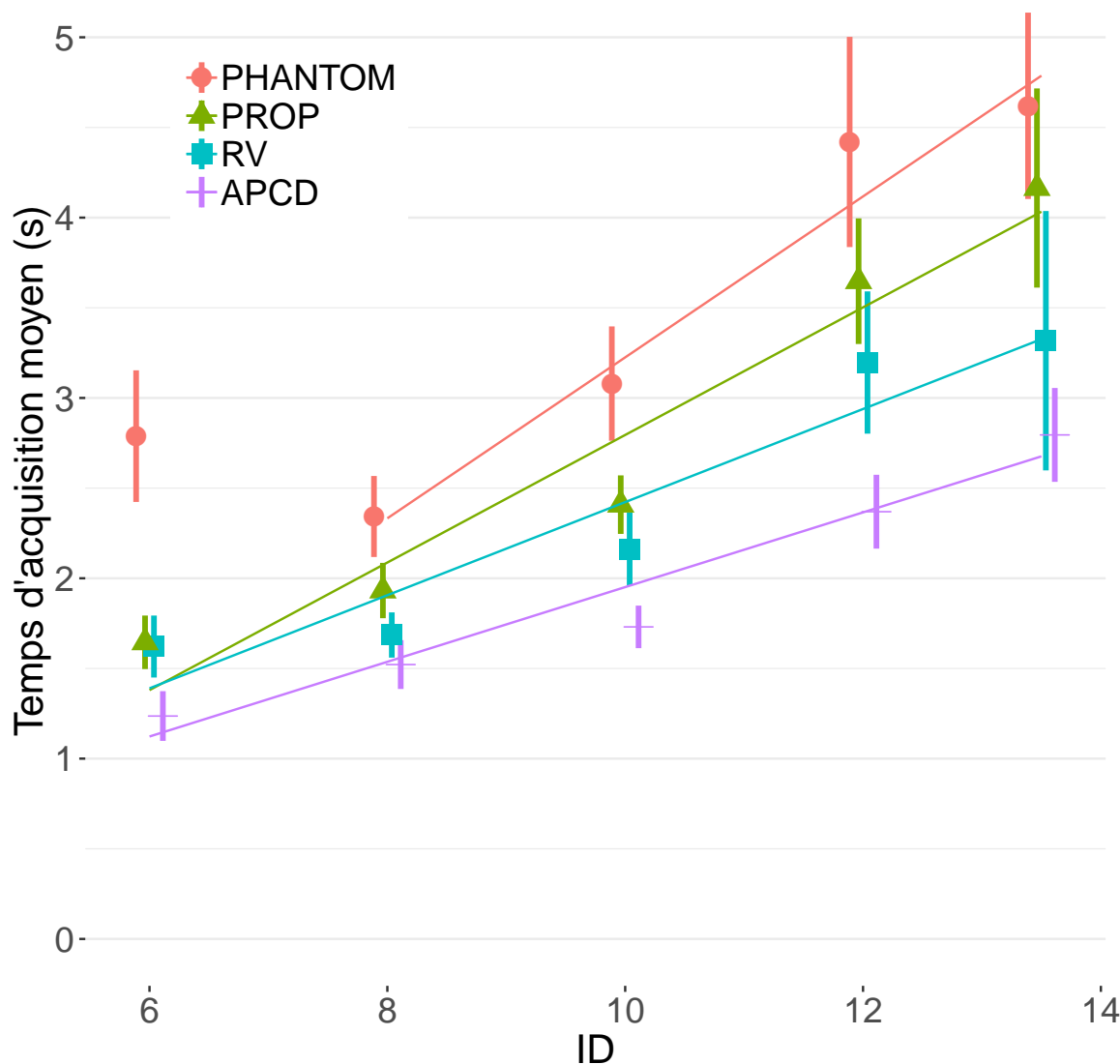


FIGURE 3.7: Le *temps d'acquisition* en fonction de l'ID des cibles pour chaque condition, avec intervalle de confiance à 95%. Les régressions linéaires sont représentées avec l'ID de 6 ignoré pour la condition PHANTOM (détails dans le texte).

Condition	Pente	p-value	R-squared	Indice de performance bit/s
PHANTOM (incluant l'ID de 6)	0.301	0.041	0.798	3.1
PHANTOM (excluant l'ID de 6)	0.446	0.020	0.961	3.3
PROP	0.354	0.006	0.941	3.8
RV	0.258	0.012	0.911	4.4
APCD	0.207	0.004	0.953	5.4

TABLE 3.2: Paramètres des régressions linéaires prédisant le *temps d'acquisition* en fonction des ID pour chaque condition.

Régressions de Fitts

La forte interaction entre TECHNIQUE et ID est une première indication que l'ID peut être utilisé pour calculer un indice de performance par technique, et ainsi comparer l'efficacité générale des techniques. Pour chaque technique, nous avons tracé le *temps d'acquisition* en fonction de l'ID des tâches effectuées, et calculé une régression linéaire. Ceci est affiché dans la figure 3.7. Le *temps d'acquisition* de la tâche de l'ID le plus petit est clairement mal aligné pour la condition PHANTOM. Cette cible nécessitait une translation de 54 mm vers la droite et une rotation de 23° autour d'un axe horizontal parallèle à l'écran. La réalisation de cette rotation avec la pointe du stylo s'est avérée difficile, alors qu'il n'y a pas eu de difficulté particulière pour l'effectuer avec la sphère. Nous rapportons

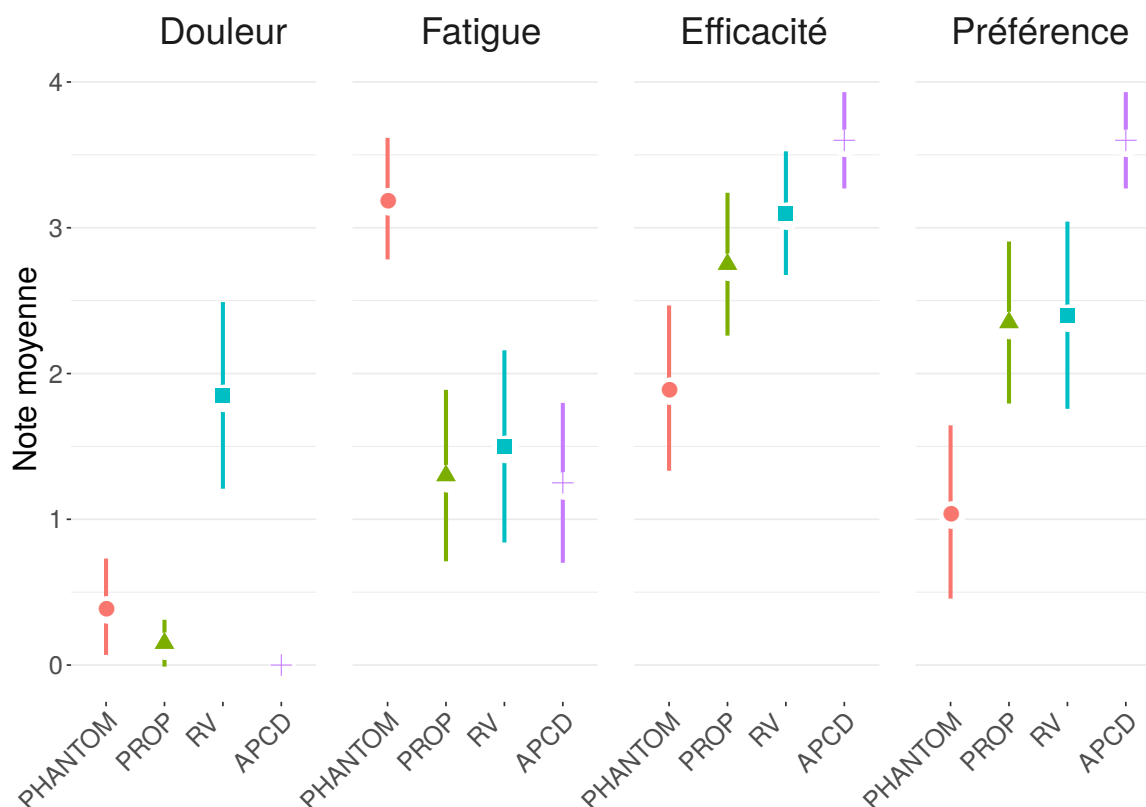


FIGURE 3.8: Évaluations subjectives des participants, avec intervalle de confiance à 95%.

dans le tableau 3.2 les paramètres des régressions linéaires, avec deux ensembles pour la condition PHANTOM : en incluant ou en excluant la tâche d'ID de 6. L'IP a été estimé par condition comme la moyenne du ratio ID sur *temps d'acquisition* calculé séparément pour chaque ID et chaque participant, comme cela est préconisé dans [140].

Résultats subjectifs

Le questionnaire post-expérience nous a permis de mesurer l'évaluation subjective de chaque technique par les participants selon quatre critères. Les participants devaient évaluer quatre phrases par technique avec un score allant de 0 ("Je ne suis pas du tout d'accord") à 4 ("Je suis totalement d'accord"). Les phrases étaient : "Cette technique me donne mal aux yeux, mal à la tête ou provoque des nausées." (Douleur), "Cette technique provoque de la fatigue musculaire." (Fatigue), "Cette technique me permet de positionner la chaise efficacement (rapidement et avec précision)." (Efficacité), "J'aime utiliser cette technique." (Préférence). Les résultats sont résumés dans la figure 3.8.

3.3 Discussion

3.3.1 Validité de l'indice de difficulté

Nous avons introduit une nouvelle méthode de calcul d'un indice de difficulté pour une tâche de placement 6D, qui modélise une tolérance d'erreur spatiale partagée entre l'erreur de translation et l'erreur de rotation. Un premier indice de la validité de cet ID est que le temps d'acquisition augmente systématiquement avec l'ID pour chaque condition. Un objectif plus ambitieux est d'utiliser cet ID pour permettre la mesure d'un indice de performance sous la forme d'une bande passante d'information en bit/s pour chaque technique. Cela nécessite que le temps d'exécution des tâches varie linéairement avec l'ID. Bien que l'on puisse voir sur la figure 3.7 que les alignements ne sont pas parfaits (avec une tendance à surestimer pour l'ID de 10), les paramètres des régressions linéaires rapportés dans le tableau 3.2 indiquent une relation proche du linéaire. La qualité de ces valeurs soutient la validité des IPs rapportés dans le même tableau.

Dans leur revue des études de Fitts concernant des tâches de pointage 2D, Soukoreff et Mackenzie rapportent que les ID ont été testés dans la gamme 1 à 6 (tableau 4 dans [140]). Nous avons testé ici des tâches plus difficiles, dans une gamme de 6 à 13,5. Il fallait s'y attendre, car nos tâches de placement à 6 degrés de liberté combinent la difficulté de deux tâches de placement 3D, et nécessitent par conséquent beaucoup plus d'informations qu'une seule tâche de placement 2D. Néanmoins, les valeurs d'IP représentent la bande passante d'information que les utilisateurs sont capables de traiter avec la technique d'interaction en question. Ils doivent être du même ordre de grandeur, quelle que soit la difficulté de la tâche, en supposant que les techniques ne sont pas d'une inefficacité critique. Soukoreff et Mackenzie ont trouvé des IPs compris entre 3,7 bit/s et 4,9 bit/s pour les études mettant en jeu des souris 2D standard. Les IPs estimés dans notre expérience de placement 6D sont compris entre 3,1 bit/s à 5,4 bit/s, et donc remarquablement similaires. Avec une valeur de 5,4 bit/s, l'IP estimé pour l'APCD est similaire aux meilleures techniques de souris 2D rapportées dans [140]. Cela indique que l'APCD fait un très bon usage de la capacité de contrôle de ses utilisateurs, bien que qu'il convienne de rester prudent lorsque nous comparons des valeurs d'IP estimées de différentes manières et sur des tâches très différentes. Pourtant, les résultats de cette première étude d'estimation d'IP dans une tâche de placement à 6D sont encourageants et devraient motiver de nouvelles études et répliques mettant en jeu l'ID que nous avons proposé.

3.3.2 Supériorité des techniques haute-fidélité

Dans cette expérience, nous avons fait attention à minimiser les différences techniques entre les conditions. En particulier, nous avons veillé à ce que les caractéristiques essentielles soient partagées entre les conditions : PHANTOM et PROP mettaient en jeu le même affichage indirect, PROP, RV et APCD le même périphérique d'entrée. Malgré tout, les performances sont réparties presque régulièrement entre les conditions, comme l'illustre la figure 3.6, en haut. Cela indique que les facteurs que nous avons testés, soit les périphériques d'affichage et de contrôle et leurs combinaisons, ont eu une influence importante sur l'efficacité des utilisateurs dans les tâches de placement.

Deux points importants différencient la condition basse fidélité qu'est PHANTOM des deux conditions HF que sont APCD et RV. Le premier point est l'utilisation d'une large sphère permettant des manipulations confortables adaptées à l'échelle naturelle des utilisateurs. Le bras articulé, à l'instar de la souris d'ordinateur, est un dispositif moins naturel mais qui tire avantage de la grande précision de notre poignet et de nos doigts. Le second point est la superposition de l'objet manipulé et de l'objet affiché en un unique

objet qui réagit comme un objet physique réagirait. Analysons ces deux points plus en détail.

Efficacité de la large sphère pour le contrôle 6D

Dans deux études récentes, la technique PHANTOM s'est révélée être la technique de placement 6D la plus efficace [159] et la plus précise [157]. Cependant, ces études n'incluaient pas une sphère tenue à deux mains comme périphérique d'entrée. Dans notre expérience, les trois conditions qui ont utilisé la large sphère ont donné des performances supérieures à celles du PHANTOM, indiquant un contrôle spatial plus performant que le statu quo.

Pour aller plus loin dans l'analyse de ce résultat, nous nous sommes concentrés sur les conditions PROP et PHANTOM, qui ne diffèrent que par le périphérique d'entrée. Globalement, les performances d'acquisition avec la condition PROP étaient 18,8% meilleures que les performances avec la condition PHANTOM. Pour rentrer dans le détail, nous avons analysé la trajectoire de l'objet contrôlé pendant les tâches. La figure 3.9 illustre une trajectoire typique de correction d'erreur de translation et de rotation d'une tâche de placement. Nous avons segmenté chaque essai en une phase balistique suivie d'une phase corrective. Pour segmenter, nous avons calculé la vitesse de correction instantanée globale (à deux dimensions : translation et rotation). Nous avons placé le seuil de segmentation à l'instant où la vitesse de correction instantanée devient inférieure à 10% de son maximum. La figure 3.10 permet de comparer les durées moyennes passées dans chaque phase avec les différentes conditions.

Il semble que la condition PROP soit plus efficace que la condition PHANTOM dans les deux phases, bien que l'avantage soit plus marquant dans la phase corrective. La scène était plus petite dans l'espace moteur pour la condition PHANTOM, ce qui aurait dû favoriser les performances durant la phase balistique, car de plus petits mouvements

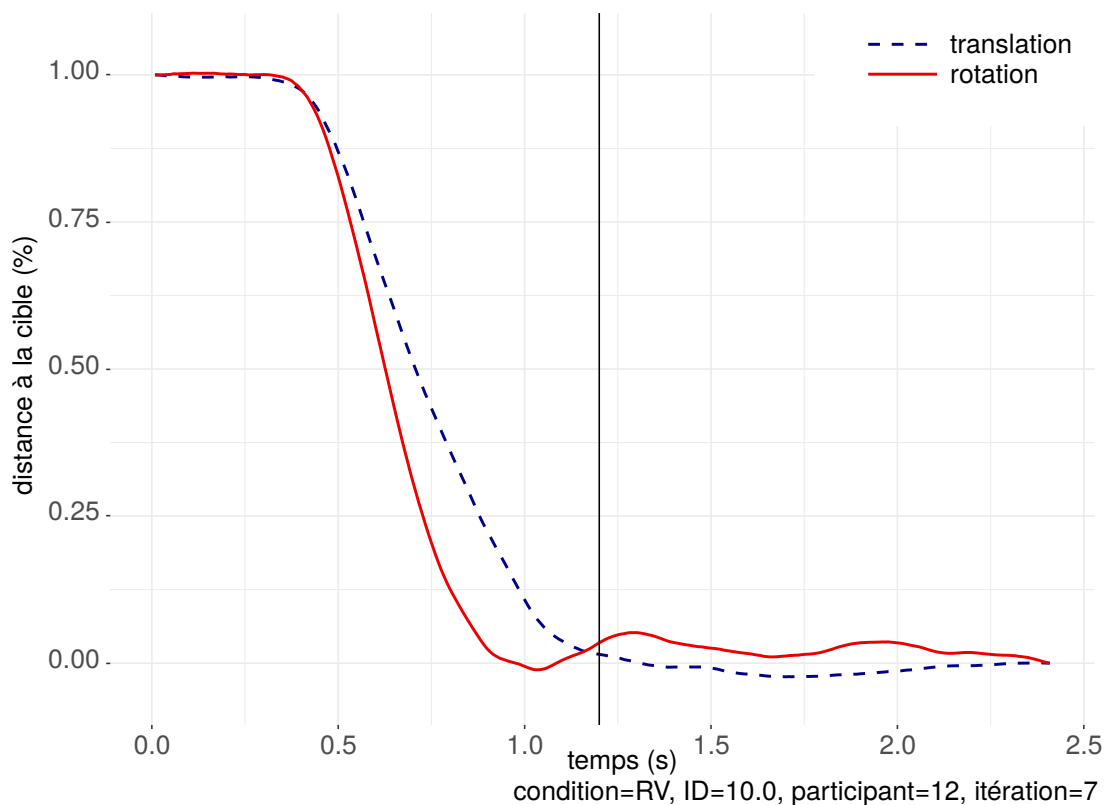


FIGURE 3.9: Correction d'erreurs en fonction du temps écoulé. Les essais sont segmentés en deux phases (ligne verticale) : une phase balistique suivie d'une phase corrective.

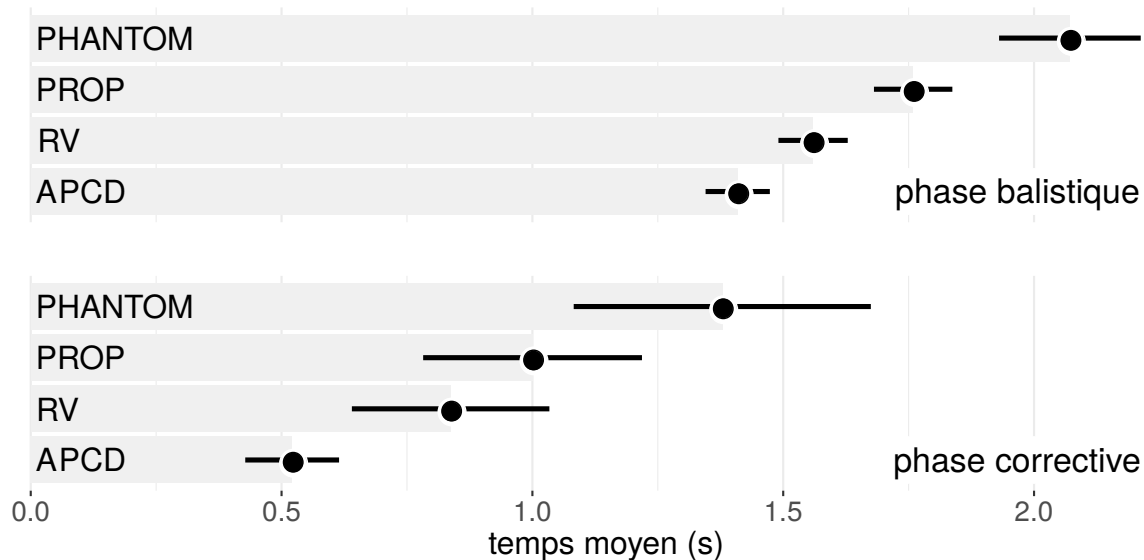


FIGURE 3.10: Temps passé dans la phase balistique (en haut) par rapport au temps passé dans la phase corrective (en bas), au cours des tâches de placement, moyenné sur les participants et les ID, avec intervalles de confiance à 95%.

devaient être réalisés. Cependant, la durée de la phase balistique pour la condition PROP était plus petite, ce qui indique que couvrir une plus grande distance avec la grande sphère tenue à deux mains restait plus efficace que de parcourir de petites distances avec le stylo. Le gain de performance notable de la condition PROP dans la phase corrective est quant à lui attendu, grâce à la stabilité offerte par la tenue de la sphère à deux mains, à la précision de rotation rendue possible par le grand rayon de la sphère, et à l’absence de contraintes ergonomiques pour effectuer les rotations. De plus, nous avons observé que quatre des vingt participants ont eu besoin d’aide lors de leur première manipulation du stylo du PHANTOM, tandis que tous les utilisateurs ont rapidement réussi à acquérir la première cible avec la sphère sans aide. Cela indique un contrôle plus intuitif de la sphère que du stylo.

Bénéfice d’une superposition des espaces moteur et visuel

Bien que les conditions PROP, RV et APCD utilisent le même périphérique d’entrée, les participants étaient en moyenne 12,4% et 29,5% plus efficaces dans les conditions RV et APCD respectivement, en comparaison avec la condition PROP. RV et APCD créent tous deux l’illusion pour l’utilisateur que l’objet contrôlé est entre ses mains. Cela confirme un résultat précédent de Ware et Rose qui ont observé qu’“avoir la main physiquement au même endroit que l’objet virtuel manipulé” était un facteur important pour une rotation efficace des objets virtuels avec des contrôleurs tangibles [161].

3.3.3 Supériorité de l’APCD sur le casque de RV

Les conditions APCD et RV utilisaient le même périphérique d’entrée et superposaient toutes deux la scène dans les espaces moteurs visuels. En d’autres termes, les participants devaient effectuer exactement la même séquence de gestes dans les deux conditions. Pourtant, les performances de placement sont très différentes : les performances de l’APCD sont 17,0% meilleures que les performances du casque de RV. Pour interpréter ce résultat, analysons plus en détails les différences d’informations visuelles fournies avec les deux techniques, qui représentent la seule différence entre ces deux conditions. Les informations

visuelles disponibles diffèrent en deux points entre les techniques : la qualité d’affichage et la quantité d’informations contextuelles disponibles.

Différences d’affichage

L’APCD présente un certain nombre d’inconvénients par rapport au casque de RV, mais les performances supérieures obtenues avec ce système indiquent que ces inconvénients n’empêchent pas d’effectuer des tâches de placement 6D efficacement :

- Les participants pouvaient bloquer la projection avec leur tête, mais ils ont tous rapidement su adopter une posture où ils ne faisaient pas d’ombre.
- La limite de la zone de projection aurait pu être un autre problème, mais nous avons choisi un ensemble de positions cibles incluses dans la zone de projection pour cette expérience.
- Le champ de vision de l’APCD est limité et dépend de la position de la sphère. Cependant, comme nous nous sommes intéressés aux avantages potentiels de la proprioception, nous avons utilisé un nombre limité de cibles qui étaient toujours au même endroit, et nous avons utilisé une flèche (dans toutes les conditions) pour aider les participants à trouver rapidement les cibles en-dehors du champ de vision en début de session. Ce faisant, nous avons supprimé le problème de *découverte de la scène* : un problème différent que nous n’avons pas souhaité traiter, et qui dépend certainement des capacités de perception spatiale de l’utilisateur.

Un facteur expliquant les performances inférieures du casque pourrait être la densité de pixels inférieure dans la zone de focus visuel des participants. Cependant, cela n’a pas empêché le casque de produire des performances supérieures à la condition PROP, pour laquelle la densité de pixels était similaire à celle de la condition APCD. En conséquence, nous ne pensons pas que la densité de pixels soit un facteur majeur, bien qu’il soit intéressant de tester de nouveau cet effet lorsque la résolution sera améliorée sur les futurs casques. Le port du casque était inconfortable pour certains participants, comme l’indique la figure 3.8 (douleur). C’est un problème bien connu avec les casques opaques : ils sont très bien tolérés par certains utilisateurs tandis que d’autres ressentent rapidement une gêne. À l’inverse, aucun participant n’a indiqué une gêne avec l’APCD.

Différences d’informations contextuelles

Outre les différences dans les performances d’affichage, il est important de noter que l’APCD et le casque de RV diffèrent radicalement dans les informations contextuelles qui sont visuellement présentées aux utilisateurs. Avec l’APCD, la scène virtuelle contenant la chaise manipulable et les cibles semble incluse dans une scène physique plus grande qui comprend également la sphère, les mains et les avant-bras de l’utilisateur, mais également la pièce et les meubles qui l’entourent. Avec le casque opaque, la totalité du contexte visuel doit être recréée, et elle est donc souvent beaucoup plus pauvre. Dans notre expérience, nous avons simplement ajouté un plan texturé au niveau du sol. Les participants ont exprimé qu’ils se sentaient coupés du monde réel. Cette différence ne devrait pas influencer la phase balistique du mouvement, qui bénéficie de la mémoire proprioceptive des positions cibles et pourrait être exécutée à l’aveugle. En effet, la figure 3.10 révèle que la plupart du gain de performance de la condition APCD sur la condition RV se fait sur la phase corrective des tâches de placement. Nous pensons que les informations contextuelles riches offertes par l’APCD, en particulier la vision de la sphère et des mains, peuvent avoir fourni un retour visuel de meilleure qualité, permettant un contrôle des rotations plus précis, plus exact. En effet, les rotations ont un effet bien plus perceptible sur les objets les plus éloignés du point de pivot, c’est-à-dire la surface de la sphère et les mains, que sur la petite

chaise centrée sur le point de pivot dans la sphère. Cet avantage fourni par la perception de l'environnement physique est cohérent avec une étude antérieure de Krichenbauer et al., qui a comparé les performances de deux techniques d'interaction proches, l'une en RV et l'autre en RA, dans une tâche de placement 9D, et a trouvé une amélioration de 22% des performances avec la RA. [83].

Nous avons calculé la coordination du contrôle de l'objet effectué par les participants en calculant la m-métrique sur les trajectoires enregistrées. La m-métrique mesure à la fois la simultanéité d'exécution des différents degrés de liberté du mouvement effectué et l'efficacité de la coordination [100]. Elle a des valeurs comprises entre 0 (pas de coordination ou inefficacité totale) à 1 (coordination totale avec une efficacité optimale). Alors que la coordination est similaire dans la phase balistique (0,69 pour les conditions RV et APCD), le contrôle est bien plus coordonné avec l'APCD dans la phase corrective (0,41 pour RV, 0,72 pour APCD). Cette mesure confirme que la plupart des avantages des performances de placement de l'APCD par rapport au casque de RV se situent dans la phase corrective.

Acceptabilité de l'APCD

Il existe une correspondance remarquable entre l'efficacité objective mesurée dans l'expérience (c'est-à-dire la figure 3.6, en haut) et l'efficacité subjective rapportée par les participants (figure 3.8, efficacité). Cela indique que non seulement les participants étaient plus efficaces avec l'APCD, mais aussi qu'ils *sentaient* l'avantage en terme de performances. C'est un bon signe pour l'acceptabilité de cette nouvelle forme de technique d'interaction. Cela peut également expliquer pourquoi les participants ont préféré cette condition (figure 3.8, préférence).

Cette grande différence de performances entre l'APCD et le casque de RV est un résultat intrigant qui nécessitera des études supplémentaires pour comprendre pleinement ses causes. Une piste d'étude consisterait à enrichir les informations contextuelles fournies dans le casque de RV et à évaluer leur effet sur les performances.

Dans tous les cas, cette étude indique qu'un APCD peut fournir une technique d'interaction très efficace pour du placement dans l'espace, en particulier lorsque les utilisateurs peuvent s'appuyer sur une mémoire proprioceptive des positions cibles. D'autres études sont nécessaires pour vérifier si l'excellente performance persiste lorsque les cibles ne sont pas bien connues.

Conclusion

Dans ce chapitre, nous avons présenté une étude utilisateur contrôlée qui nous a permis de quantifier les différences de performances de quatre techniques d'interaction dans une tâche de placement 6D, tâche fondamentale pour interagir avec des environnements virtuels 3D. Nous avons testé deux techniques haute-fidélité : un APCD et un casque de RV, qui se sont montrées respectivement 43% et 29% plus efficaces que le statu quo : une technique basse fidélité mettant en jeu un bras articulé et un écran plat. Cette étude apporte la première preuve empirique de la supériorité d'un APCD dans une tâche de placement 6D.

Nous avons également introduit une méthode pour évaluer un indice de difficulté pour décrire les tâches de placement 6D. Nous avons observé une forte relation linéaire entre le temps d'exécution d'une tâche et son ID, ce qui fournit un premier indice sur la validité de cet ID. Son utilisation permet le calcul d'un indice de performance pour chacune des techniques testées. Les indices de performances que nous avons calculés pour les deux techniques HF sont de 4,4 et 5,4 bit/s pour le casque de RV et l'APCD respectivement. Ces indices de performances sont comparables à ceux des meilleures souris d'ordinateur, ce qui témoigne de la capacité de ces systèmes à exploiter la capacité de traitement de l'information des utilisateurs. Notre méthode devrait faciliter la comparaison des résultats entre les études de placement 6D à venir.

Les gains de performances observés entre l'APCD et le casque de RV, malgré l'utilisation du même périphérique d'entrée avec les deux techniques, suggèrent que les signaux visuels riches offerts par le monde physique et visibles avec un APCD peuvent être d'une importance particulière pour un contrôle précis d'objets dans l'espace. Ce résultat appelle à de nouveaux travaux pour étudier cette hypothèse, par exemple en contrôlant l'ajout de plusieurs indices visuels avec le casque opaque, y compris une représentation des bras, des mains et des doigts de l'utilisateur. Cette approche pourrait fournir des directives pour améliorer les performances de placement 6D avec les casques de RV.

Chapitre 4

Supériorité de la haute fidélité dans une tâche d'apprentissage d'objets 3D

Introduction

Les étudiants en anatomie doivent acquérir des connaissances à propos d'objets 3D tels que des os et des organes, notamment leur apparence et leur forme. D'autres étudiants en sciences, technologie, ingénierie et mathématiques (STIM) ont des besoins d'apprentissage similaires. Dans cette thèse, *l'apprentissage d'objets 3D* se réfère au processus cognitif de création d'une représentation mentale de la forme d'un objet initialement inconnu. Cette représentation mentale peut ensuite être utilisée pour reconnaître l'objet dans une scène, quelle que soit son orientation. Dans les cours de STIM traditionnels, les objets 3D doivent être appris à partir d'un ensemble de représentations statiques telles que des images ou des dessins. Cette tâche est connue pour requérir de bonnes capacités de rotation mentale (RM) : la capacité d'imaginer des rotations d'un objet sans avoir l'objet dans les mains. Il a été démontré que ces capacités sont extrêmement variables d'un élève à l'autre, ce qui entraîne des inégalités dans le processus d'apprentissage, et dans les résultats aux examens [123, 158].

Nous avons vu dans la section 2.2 que la littérature a étudié si l'utilisation d'interfaces interactives plus fidèles permettait de réduire l'impact de la RM sur le processus d'apprentissage. Des études ont montré les bénéfices de l'utilisation de systèmes interactifs basse fidélité, permettant un rendu dynamique et une exploration active d'objets, par rapport à l'utilisation de livres ou de diaporamas. L'utilisation de systèmes plus fidèles mettant en jeu des environnements de réalité virtuelle (RV) et de réalité augmentée (RA) a également commencé à être étudiée. Ces systèmes haute-fidélité (HF) semblent augmenter la motivation et l'intérêt des étudiants, mais des preuves empiriques de leurs bénéfices pour de l'apprentissage, par rapport à des systèmes basse fidélité, sont difficiles à trouver dans la littérature. Cet axe de recherche doit être approfondi, car cette connaissance est importante pour les institutions qui envisagent d'investir dans des systèmes de RV et de RA. En outre, les tâches d'apprentissage d'objets 3D pourraient particulièrement bénéficier de l'amélioration de la perception de profondeur offerte par les systèmes d'interaction HF.

Nous avons mis au point une étude utilisateur afin de mesurer si l'utilisation de systèmes d'interaction HF permet d'améliorer la qualité de l'apprentissage d'un objet 3D. Nous avons évalué trois techniques d'interaction. Comme *statu quo*, nous avons utilisé une tablette qui permet l'exploration active d'un objet 3D avec une interaction multi-touch et un rendu en perspective simple de l'objet. Nous avons comparé cette technique avec deux techniques HF qui offrent un contrôle isomorphe de l'objet et un rendu stéréoscopique couplé à la position de l'utilisateur : une technique de RV emploie un casque

opaque, et un appareil à perspective corrigée déplaçable (APCD). Ces deux techniques se servent d'une sphère de 30 cm de diamètre comme périphérique d'entrée pour manipuler les objets virtuels. Elles permettent l'exploration des objets de manière très similaire à la réalité physique, contrairement aux environnements d'apprentissage plus traditionnels sur ordinateurs de bureau ou tablettes. Nous avons choisi le casque de RV comme l'un des systèmes HF les plus facilement exploitables actuellement en contexte écologique. Nous avons choisi l'APCD comme le système HF le plus performant à l'heure actuelle, notamment pour la manipulation 6D d'objets, comme nous l'avons montré dans le chapitre précédent.

Nous avons conçu un protocole d'apprentissage sur deux jours successifs, car de nombreuses études en sciences cognitives ont montré l'importance du sommeil sur la mémoire déclarative et procédurale, y compris pour des tâches visuospatiales [46, 113]. En parallèle, une étude de Cherdieu et al. suggère que l'avantage de l'implication du corps dans l'apprentissage des connaissances conceptuelles ne peut être mesuré que lorsque les participants sont testés à nouveau un deuxième jour [36].

Nous émettons l'hypothèse que le casque de RV et l'APCD devraient surpasser la tablette dans une tâche d'apprentissage d'objets 3D. Nous supposons en effet que le rendu plus réaliste et la manipulation plus intuitive permis par les techniques HF devraient faciliter la construction d'une représentation mentale de l'objet en réduisant la surcharge cognitive extrinsèque lors de l'apprentissage. Cette hypothèse a d'abord été exprimée, mais non testée, par Lee et al. [88]. Tester cette hypothèse est un objectif secondaire de notre étude.

La seconde contribution de cette thèse est d'apporter des démonstrations empiriques solides des bénéfices de techniques d'interaction HF dans une tâche d'apprentissage d'objets 3D. Dans ce chapitre, nous commençons par décrire l'étude utilisateur que nous avons réalisée. Nous présentons ensuite les résultats et les analyses de l'étude. Enfin, nous discutons de ce nouveau bénéfice mesuré des systèmes d'interaction HF, de la relation entre les capacités de RM et l'apprentissage d'objets 3D, et du protocole expérimental ambitieux que nous avons mis en place.

Sommaire

Introduction	103
4.1 Conception de l'étude utilisateur	105
4.1.1 Vue d'ensemble	105
4.1.2 Protocole expérimental	106
4.1.3 Conditions expérimentales	108
4.1.4 Installation technique	111
4.2 Résultats et Analyses	112
4.2.1 Méthodologie	112
4.2.2 Résultats de reconnaissance	113
4.2.3 Résultats des VMRT	115
4.2.4 Analyse de la quantité de rotation	116
4.2.5 Résultats subjectifs	117
4.3 Discussion	118
4.3.1 Supériorité des groupes HF	118
4.3.2 Apprentissage d'objets 3D et capacités de rotation mentale	119
4.3.3 Un protocole expérimental ambitieux	120
Conclusion	122

4.1 Conception de l'étude utilisateur

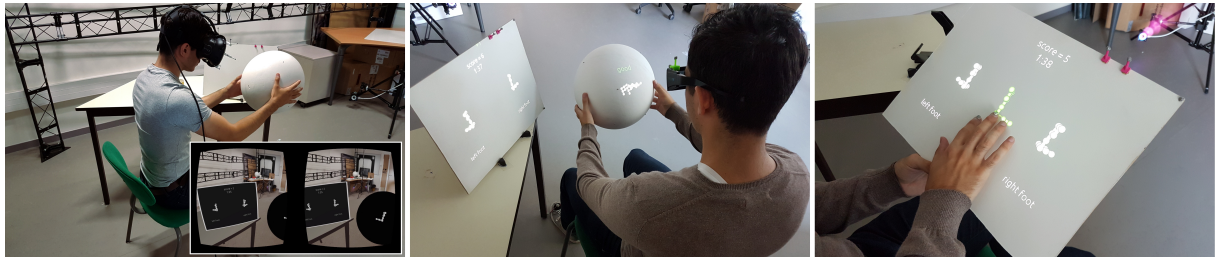


FIGURE 4.1: Les trois techniques d'interaction testées : la technique HF avec le casque de RV (à gauche), la technique HF avec l'APCD (au milieu), et la technique basse fidélité avec la tablette (à droite).

4.1.1 Vue d'ensemble

Nous avons conçu une étude utilisateur pour observer comment trois techniques d'interaction affectent l'apprentissage d'un objet 3D générique. Les techniques incluent deux interactions haute-fidélité, l'une utilisant un casque de RV, l'autre un APCD ; et une interaction basse fidélité utilisant une tablette tactile. Ces techniques sont illustrées dans la figure 4.1. Nous avons utilisé un seul objet afin de limiter la charge cognitive et la durée de l'expérience pour les participants. Nous appelons l'objet à apprendre le “**sphinx**” (c.f. Figure 4.2, à gauche).

Chaque participant est *entraîné* en utilisant l'une des trois techniques d'interaction. Nous avons choisi ce protocole inter-sujets pour éviter tout transfert de connaissances entre les sessions. Les participants se voient présenter à plusieurs reprises des objets vus depuis différents *points de vue initiaux*. Les objets sont soit le **sphinx**, soit un objet avec une forme légèrement différente. Les participants peuvent faire tourner ces objets librement avec la technique d'interaction qui leur est assignée afin de les observer depuis de nombreux points de vue. Ils doivent ensuite identifier si les objets sont le **sphinx** ou non.

Nous mesurons la reconnaissance du **sphinx** dans des sessions de *reconnaissance*. Les participants se voient présenter une série des mêmes objets que ceux utilisés dans l'*entraînement*, vus depuis divers points de vue *fixes* sur une tablette basse fidélité. Pour chaque objet présenté, les participants doivent décider si l'objet présenté est le **sphinx** ou non.

Enfin, un choix de conception majeur a été de répartir les sessions expérimentales des participants sur deux jours consécutifs afin de favoriser une stabilisation du processus d'apprentissage, et surtout l'intégration de l'expérience sensorimotrice vécue dans la représentation visuelle construite par le participant.

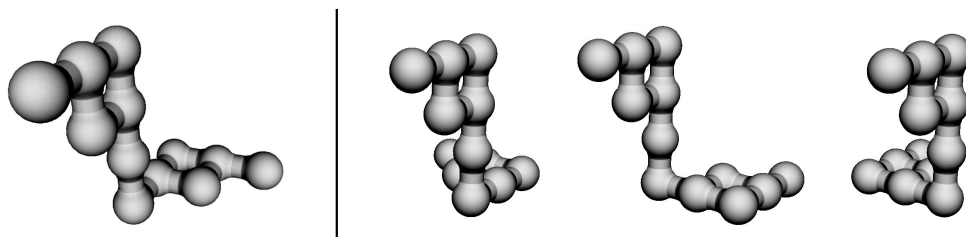


FIGURE 4.2: Le **sphinx** (à gauche) et les trois distracteurs (à droite).

Le sphinx et les distracteurs

De nombreux objets de la vie quotidienne peuvent être facilement reconnus depuis n'importe quel point de vue grâce à des indices visuels forts tels que des motifs de couleur, de taille ou de texture. Par conséquent, se souvenir de leur forme vue depuis n'importe quel point de vue n'est *pas* indispensable pour les reconnaître, et la demande en capacité de rotation mentale est probablement fortement réduite. Comme nous nous sommes concentrés sur l'influence de la technique d'interaction sur le processus d'apprentissage, nous avons utilisé un objet inconnu qui *n'a pas* de tels indices, comme un os vu sur des images radiographiques. Nous avons créé le **sphinx** comme un objet abstrait qui ne ressemble à aucun objet de la vie réelle, pour éviter tout biais potentiel induit par les différences d'expérience des participants. Nous nous sommes inspirés des objets utilisés dans le VMRT. Le **sphinx** est asymétrique de sorte qu'il a une apparence différente lorsqu'il est vu depuis différents points de vue. Nous avons créé trois objets qui ont été utilisés comme distracteurs à partir de symétries et de rotations partielles du **sphinx**. Les distracteurs sont illustrés à côté du **sphinx** sur la figure 4.2, à droite.

Le **sphinx** doit être distingué des distracteurs, soit après des rotations depuis un point de vue *initial* pendant les sessions d'entraînement, soit depuis un point de vue *fixe* pendant les sessions de reconnaissance. Pour les points de vue initiaux et fixes, nous avons utilisé 32 rotations à partir d'une *orientation de référence*, définies selon 8 axes de rotation : (1,0,0), (0,1,0), (0,0,1), (1,1,1) et leurs opposés, et selon 4 angles de rotation (40°, 80°, 120° et 160°).

4.1.2 Protocole expérimental

La figure 4.3 présente la chronologie complète de l'expérience. Chaque étape du protocole est détaillée ci-dessous :

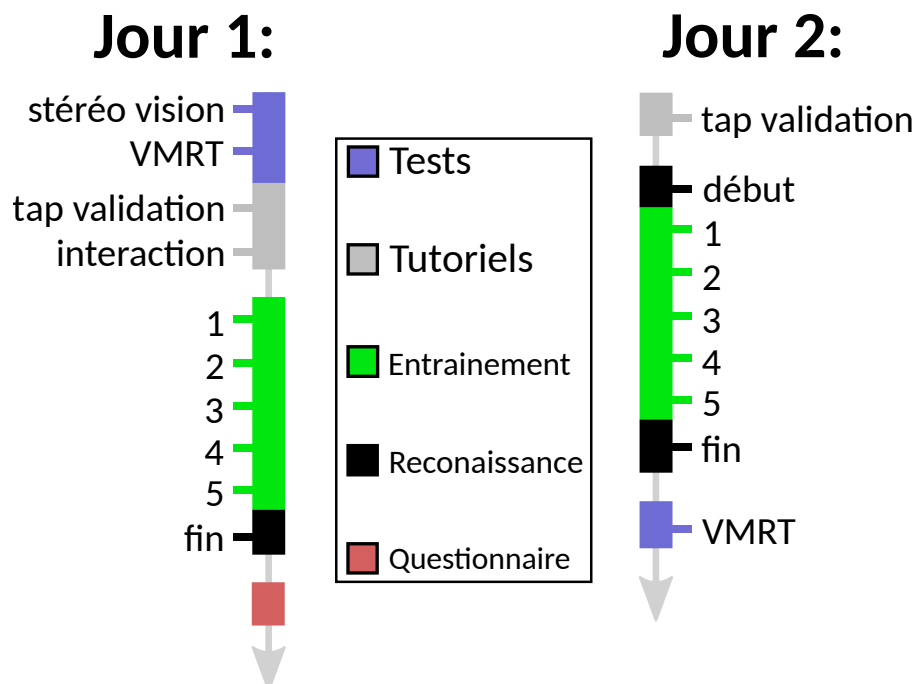


FIGURE 4.3: La chronologie de l'expérience par jour.

Pre-tests et post-tests

Comme deux des techniques d'interaction testées utilisent un rendu stéréoscopique, nous avons testé les participants pour nous assurer qu'ils avaient une bonne vision stéréoscopique en utilisant le test de vision stéréoscopique RANDDOT.

Avec un protocole inter-sujets, une différence de performance entre plusieurs groupes pourrait s'expliquer en partie par une différence de performance initiale de "reconnaissance d'objets 3D" des participants entre les groupes. Comme il n'y a pas de test standard pour cette capacité, nous avons choisi d'équilibrer les groupes en fonction de leurs capacités de RM en utilisant le VMRT [154]. Comme on l'a vu en section 2.2.2, des travaux antérieurs suggèrent que les capacités de RM devraient jouer un rôle essentiel à la fois dans l'apprentissage et dans la reconnaissance des objets 3D [64, 102]. De plus, en demandant aux participants d'effectuer un deuxième VMRT à la fin de l'expérience, nous pourrions évaluer comment les trois techniques d'interaction ont affecté leurs capacités de RM.

Tutoriels

L'étude demande aux participants de dire si un objet présenté est le **sphinx** ou non. Pour fournir cette réponse, nous avons demandé aux participants de taper une fois sur le sol avec le pied droit si l'objet présenté était le **sphinx**, et avec le pied gauche si ce n'était pas le cas. Nous avons choisi cette technique de validation afin d'éviter toute perturbation de la zone de travail qui impose l'utilisation constante des deux mains et une concentration visuelle permanente. Elle s'est également montrée plus rapide et plus robuste qu'une validation vocale dans un test pilote. Au début de chaque journée, les participants suivent un rapide tutoriel d'environ 1 minute pour s'habituer à cette technique de validation par "tap" avec les pieds (tap validation).

De plus, au début de la première journée, ils se familiarisent avec la technique d'interaction à laquelle ils sont affectés dans un autre tutoriel d'une minute environ, où ils pratiquent l'interaction avec des lettres 3D (des objets sans rapport avec le **sphinx**).

Entraînement pour apprendre le sphinx

Chaque jour, les participants exécutent cinq *sessions d'entraînement* de trois minutes. Dans chacune, ils effectuent autant d'*essais d'entraînement* que possible pendant les trois minutes. Un essai d'entraînement commence par l'affichage d'un objet *cible* dans l'une des trente-deux orientations initiales. L'objet cible est soit le **sphinx**, soit un distracteur. Les participants font pivoter librement l'objet cible jusqu'à ce qu'ils puissent répondre qu'il s'agit du **sphinx** ou non en utilisant la technique de validation par tap. Ils reçoivent immédiatement un retour visuel qui leur indique s'ils ont répondu correctement, et l'essai suivant commence. Et ainsi de suite jusqu'à la fin de la session.

Le nombre d'essais effectués dans une session d'entraînement dépend du rythme des participants : plus ils répondent rapidement, plus le nombre de réponses faites au cours de la session de trois minutes est élevé.

Lors de la première session d'entraînement, un seul distracteur est utilisé : les objets cibles sont toujours soit le **sphinx**, soit ce distracteur. La tablette affiche une aide dans laquelle le distracteur est représenté à gauche et le **sphinx** à droite, tous deux dans leur orientation de référence. Cette aide est illustrée sous diverses conditions sur la figure 4.4. Les participants **doivent** faire pivoter l'objet cible dans son orientation de référence avant de pouvoir répondre : le système interdit toute réponse tant que l'orientation de référence n'est pas atteinte. Cette première session d'entraînement a été conçue pour permettre aux participants de découvrir le **sphinx** et pour leur présenter une première stratégie d'identification simple : orienter l'objet cible pour obtenir un point de vue familier depuis

lequel le **sphinx** est facilement différenciable. On qualifie cette première session de session d’entraînement à manipulation dictée, par opposition aux sessions d’entraînement à manipulation libre (toutes les autres).

Au cours de la deuxième session d’entraînement, l’aide ne montre plus que le **sphinx** dans son orientation de référence. Les trois distracteurs sont utilisés et les participants peuvent répondre à tout moment. Cette session a été incluse parce que nous nous sommes rendu compte lors d’une étude pilote que cacher le **sphinx** tout en introduisant de nouveaux distracteurs était trop déroutant.

Au cours des trois dernières sessions d’entraînement, aucune aide n’est fournie : les participants doivent se fier uniquement à leur mémoire pour savoir si l’objet cible est le **sphinx**. Le nombre total de sessions d’entraînement a été limité à cinq afin de maintenir la durée de l’expérience en dessous de 45 minutes par jour.

Reconnaissance

La capacité des participants à reconnaître le **sphinx** a été mesurée lors de *sessions de reconnaissance* de trois minutes. Tous les participants, quel que soit leur groupe, ont effectué les sessions de reconnaissance avec la même modalité basse fidélité. Lors d’un essai de reconnaissance, un objet est affiché dans l’une des trente-deux orientations sur la *tablette*. Dès que les participants ont déterminé si l’objet est le **sphinx** ou non, ils répondent avec la technique de validation par tap et un nouvel essai de reconnaissance commence. Il n’y a aucun retour visuel sur l’exactitude des réponses fournies.

En plus des sessions de reconnaissance en fin de chaque journée, les participants ont effectué une session de reconnaissance au début du deuxième jour. Cette session nous a permis d’isoler l’impact d’une nuit de sommeil sur les connaissances acquises la veille.

Questionnaire

Nous avons recueilli les retours subjectifs des participants avec un bref questionnaire. Ils ont donné leur avis concernant cinq affirmations sur une échelle de 1 (“Je ne suis pas du tout d’accord”) à 5 (“Je suis tout à fait d’accord”). Les affirmations étaient : “L’expérience m’a fait mal aux yeux, m’a provoqué des maux de tête ou des nausées.” (fatigue mentale) ; “J’ai ressenti une fatigue musculaire.” (fatigue physique) ; “Cette expérience m’a plu.” (plaisir) ; “Je pouvais manipuler les objets instinctivement.” (facilité d’apprentissage) ; et “Je pouvais manipuler les objets avec précision.” (précision).

4.1.3 Conditions expérimentales

La condition RV

Les participants portaient un casque de RV opaque et étaient immergés dans une scène virtuelle reproduisant la salle expérimentale. La figure 4.4 montre un utilisateur portant le casque opaque (en haut à gauche), ainsi qu’une image de la scène virtuelle telle qu’elle est vue par l’un des yeux de l’utilisateur (en haut à droite). La table, la tablette et la sphère ont été traquées et reproduites à leur position réelle dans la scène virtuelle. De cette façon, les participants pouvaient saisir et orienter instinctivement la sphère virtuelle en manipulant la sphère physique. Pour améliorer le contraste, nous avons utilisé un fond noir pour la tablette et la sphère dans la scène virtuelle. Les objets cibles étaient affichés au centre de la sphère virtuelle, comme s’ils y étaient solidement attachés. La rotation de l’objet était donc isomorphe à la rotation de la sphère physique. Cette condition haute-fidélité offre un rendu stéréoscopique couplé à la position de l’utilisateur et un contrôle isomorphe des objets explorés.

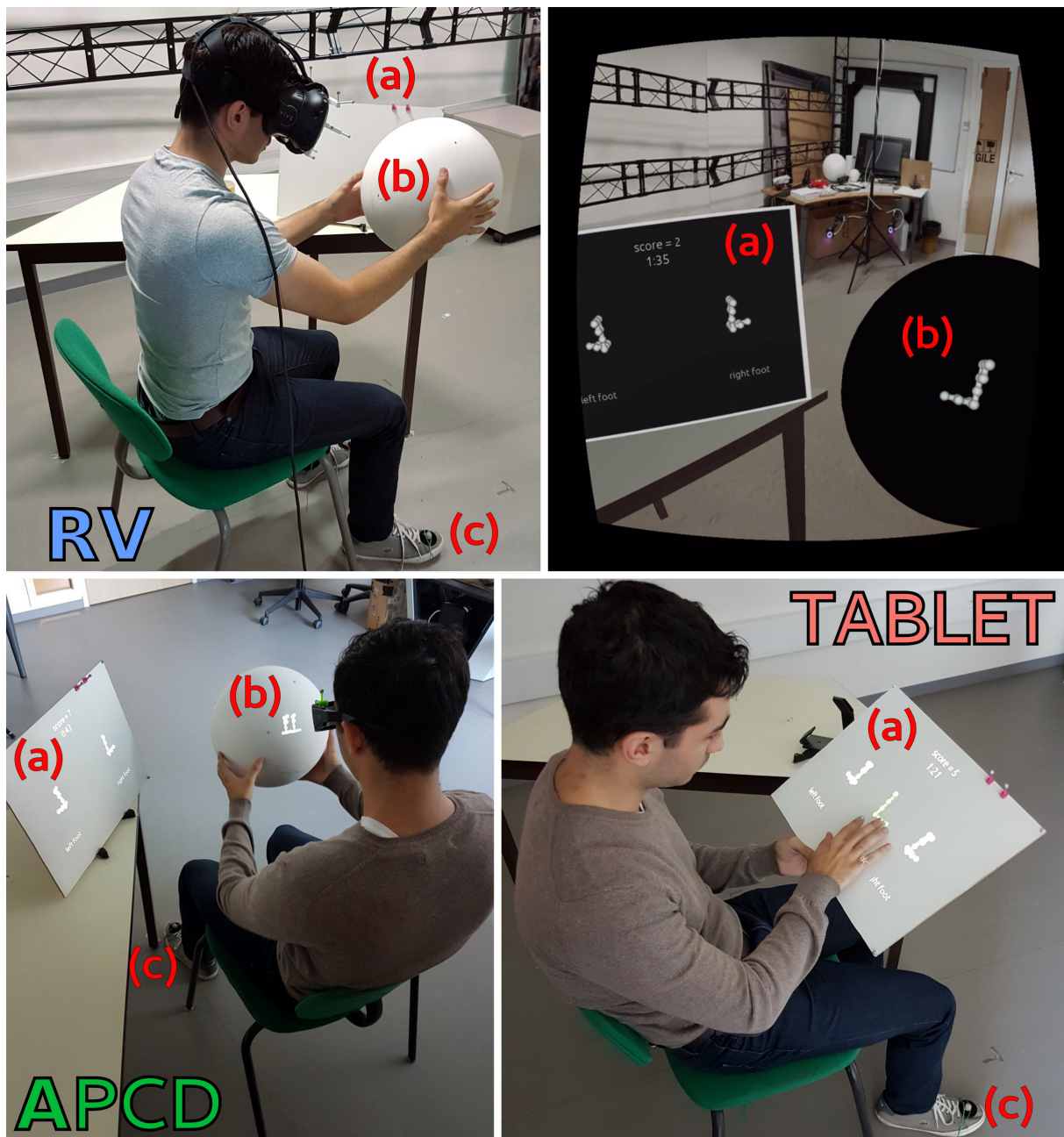


FIGURE 4.4: Configuration expérimentale montrant la tablette (A), le dispositif sphérique (B), et les marqueurs pour la validation par tap (C), pour tous les groupes, lors d'une session d'entraînement à manipulation dictée.

La condition APCD

Les participants tenaient la même sphère que dans la condition RV, mais la sphère était utilisée comme APCD : elle servait à la fois de périphérique de contrôle et d'affichage, comme cela est illustré sur la figure 4.4, en bas à gauche. Les participants portaient les lunettes à obturateur permettant l'affichage stéréoscopique actif à 60 Hz par œil. Comme pour la RV, il s'agit d'une condition haute-fidélité qui offre un rendu stéréoscopique couplé à la position de l'utilisateur et un contrôle isomorphe des objets explorés.

La condition TABLET

Avec cette condition, tous les retours graphiques ont été affichés sur la tablette. Lors d'une étude pilote, nous avons laissé les participants utiliser la tablette comme ils le

voulaient, mais une très grande majorité l’a saisie avec sa main non dominante pour mieux interagir avec sa main dominante, au lieu de la laisser posée sur la table. Ce n’est pas surprenant, compte tenu du modèle de la chaîne cinématique de Guiard, qui indique que la main non dominante positionne l’objet qui donne une référence spatiale pour la main dominante [62]. Lors de l’expérience, nous avons choisi d’imposer cette utilisation à deux mains dans un but d’uniformité.

L’objet contrôlable était affiché au milieu de l’écran comme illustré sur la figure 4.4, en bas à droite. Nous avons choisi d’émuler la tablette en utilisant notre système projecteur-caméra plutôt que de mettre en jeu une tablette tactile commerciale afin de maximiser les performances de contrôle d’objet de cette condition et de la mettre ainsi à niveau avec les deux autres. En effet, les tablettes commerciales ne permettent actuellement pas d’égaliser notre système en termes de latence ni de précision. Les différences de performance engendrées par ce choix sont détaillées dans le paragraphe “Installation technique”. Une interaction tactile avec un seul doigt contrôlait la rotation de l’objet autour des axes x et y de l’écran, tandis qu’une interaction avec deux doigts contrôlait la rotation dans le plan de la tablette. Pour être cohérent avec les tablettes que l’on trouve actuellement dans le commerce, le rendu n’était ni stéréoscopique, ni couplé à la position de l’utilisateur. Cette condition basse fidélité avec un rendu en simple perspective est représentative de systèmes largement répandus dans les établissements d’enseignement.

Informations secondaires

Dans toutes les conditions, la tablette a été utilisée pour afficher des informations secondaires pendant les essais (temps restant, indicateur de succès et aide lorsqu’elle est disponible), comme cela est illustré sur la figure 4.4. Les objets 3D de l’aide étaient affichés avec un rendu stéréoscopique couplé à la position de l’utilisateur pour les deux conditions haute-fidélité uniquement.



FIGURE 4.5: Une tâche de reconnaissance. Cette tâche est identique pour tous les groupes.

Les sessions de reconnaissance

Les objets étaient affichés au centre de la tablette sans rendu stéréoscopique ni couplage avec la position de l'utilisateur, comme cela est illustré sur la figure 4.5. Les objets n'étaient pas manipulables : les participants des groupes RV et APCD laissaient la sphère sur le côté et retiraient le casque ou les lunettes stéréo ; les participants du groupe TABLET remettaient la tablette sur la table. La seule information secondaire disponible était le temps restant avant la fin de la session, affiché au-dessus des objets à identifier.

4.1.4 Installation technique

Les différences de performances entre les systèmes pouvant impacter fortement les performances humaines mesurées, nous avons égalisé autant que possible les performances techniques des trois conditions. Toutes les actions des participants étaient suivies par le même système optique Optitrack utilisant dix caméras Prime 13 à 240 Hz. Le système optique offre des performances de suivi élevées en termes de précision et de stabilité. En particulier, son instabilité n'est pas perceptible à l'œil nu. Toutes les conditions ont été implémentées avec le même logiciel C++ fonctionnant sur un ordinateur Intel Xeon 3.7 GHz avec une carte graphique NVidia GeForce GTX 1080.

Les images étaient affichées par le même projecteur pour les conditions APCD et TABLET (un Barco F-50 de résolution 2560x1600 pixels @ 120 Hz). Les lumières étaient éteintes pour maximiser le contraste de la projection. La tablette a été simulée par une planche de bois de 60 cm × 40 cm peinte en blanc. Des marqueurs étaient attachés à la planche à des fins de suivi. Dans la condition RV, les images étaient affichées dans un casque HTC Vive (de résolution 1080x1200 pixels par œil @90 Hz avec un champ de vision de 110°) et le suivi était assuré par le système de suivi optique : nous n'avons pas utilisé le système de suivi moins précis du RV.

Nous avons utilisé le même système de suivi de mouvement dans toutes les conditions, et le même projecteur dans les conditions APCD et TABLET. Cela contribue à la validité interne de l'expérience, tout en évitant une baisse des performances prévisible qui serait induite par la latence du périphérique pour la condition TABLET [149]. En effet, notre tablette avait la même latence de 27 ms que l'APCD ; ce qui est notablement inférieur aux tablettes commerciales actuelles (environ 80 ms). La précision de capture du système, de l'ordre 0,022 mm, est clairement suffisante compte tenu que la précision des utilisateurs au toucher sur une surface tactile a été mesurée à 0.17 mm [11]. Nous avons estimé la densité de pixels à environ 90 dpi avec les conditions APCD et TABLET, ce qui est nettement inférieur à la résolution des tablettes commerciales actuelles. Nous avons considéré que cela ne poserait pas de problème dans le contexte de notre expérience où nous mettons en jeu uniquement des objets larges et sans détails fins.

Les taps avec les pieds ont été mesurés en fixant des marqueurs optiques sur chaque chaussure des participants, comme on en voit l'illustration sur la figure 4.4. L'interaction tactile dans la condition TABLET a été mise en œuvre en attachant des marqueurs sur les ongles de l'index et du majeur des participants.

Dans les conditions RV et APCD, nous avons utilisé la sphère en polystyrène de 30 cm de diamètre. Des marqueurs y étaient attachés à des fins de suivi.

4.2 Résultats et Analyses

4.2.1 Méthodologie

Trente personnes (dix-huit hommes et douze femmes) ont participé à l'étude, dont huit gauchers (âge moyen 27,1 [18, 40]). Après avoir passé le VMRT, ils ont été affectés à l'un des trois groupes de technique d'interaction. Au cours des deux jours, ils ont suivi un total de dix sessions d'entraînement de trois minutes chacune ($2 \text{ (jours)} * 5 \text{ (sessions)} * 3 \text{ (minutes)} = 30 \text{ minutes d'entraînement au total}$), et ont exécuté trois sessions de reconnaissance de trois minutes.

Notre objectif principal est d'étudier les résultats des sessions de reconnaissance, qui ont été conçues pour évaluer l'effet de la technique d'interaction sur l'apprentissage du *sphinx*. Nous avons également analysé la phase d'entraînement pour évaluer le processus d'apprentissage, son évolution au fil des sessions et selon les techniques.

Mesures

Les performances des participants à chaque session d'entraînement et de reconnaissance ont été caractérisées par différentes mesures qui ont permis d'analyser le compromis précision-vitesse à différents niveaux :

1. Le **temps de réponse**, calculé comme la durée entre l'apparition d'une cible et le moment où le participant fait un tap pour répondre, moyenné par session.
2. Le **taux de succès**, calculé comme le nombre de bonnes réponses divisé par le nombre total d'essais effectués au cours d'une session.
3. Le **score** calculé comme suit : un point est attribué pour chaque réponse correcte et un point est perdu pour chaque réponse fautive au cours de la session en cours. Ce score reflète le compromis précision-vitesse car les participants qui répondent lentement mais juste auront un score similaire aux participants qui répondent plus rapidement mais font quelques erreurs.
4. La **quantité de rotation** (uniquement pour les sessions d'entraînement), calculée comme la quantité totale de rotation appliquée à une cible durant un essai.

Enfin, nous avons analysé la différence de scores entre le pré-test et le post-test de VMRT pour étudier si l'apprentissage d'un objet 3D par la manipulation a un effet bénéfique sur les capacités de rotation mentale.

Facteurs et hypothèses

Les données ont été analysées en rapport avec notre hypothèse principale et en testant les effets des facteurs suivants sur les mesures :

- La **technique** d'interaction, avec 3 niveaux : TABLET, RV, et APCD (facteur inter-sujets) ;
- Le **numéro de session**, dont le nombre de niveaux dépend du type de session :
 - *Type R* : Les sessions de reconnaissance, avec trois niveaux de numéro de session : jour 1 fin, jour 2 début, et jour 2 fin (facteur intra-sujets) ;
 - *Type E-D* : Les sessions d'entraînement avec manipulation dictée, avec deux niveaux de numéro de session : jour 1 session #1 et jour 2 session #1 (facteur intra-sujets) ;
 - *Type E-L* : Les sessions d'entraînement avec manipulation libre, avec huit niveaux de numéro de session : jour 1 sessions #2 à #5 et jour 2 sessions #2 à #5 (facteur intra-sujets).

Les sessions de type R ont été impliquées dans l'analyse de reconnaissance du `sphinx`, les sessions de type E-L dans l'analyse de l'entraînement, et les sessions de types E-D et E-L dans l'analyse de la quantité de rotation.

Les sessions d'entraînement #2 sont légèrement différentes des autres sessions de type E-L, étant donné l'aide affichée. Pourtant, nous avons regroupé ces sessions parce que nous avons observé un comportement similaire de la part des participants (en particulier, aucun d'entre eux n'a jamais utilisé l'aide de la session #2 pendant le jour 2). Nous avons néanmoins vérifié que les analyses statistiques mettant en jeu les sessions de type E-L étaient similaires si les sessions d'entraînement #2 étaient exclues.

Aucune de nos hypothèses ne suggérerait a priori l'analyse des quantités de rotation. Néanmoins, les différences de cette mesure entre les sessions de type E-D et E-L ont apporté des indices essentiels sur le processus d'entraînement, et en particulier sur l'internalisation des stratégies. [37].

Nous nous attendions à ce que les performances de reconnaissance suivent l'ordre : $APCD > RV > TABLET$, que ce soit pour les scores, les temps de réponses ou les taux de réussite. Nous nous attendions également à ce que l'apprentissage de l'objet 3D ait un effet positif sur les scores de VMRT. Selon Cherdieu et al. [36], l'effet de la technique pourrait être plus clair au jour 2 qu'au jour 1, en particulier pour les scores de reconnaissance.

Statistiques

Nous avons effectué l'analyse des scores et des taux de réussite en utilisant des ANOVA mixtes et des t-tests par paire avec correction de Holm.

Pour l'analyse du temps de réponse, nous avons utilisé un modèle linéaire mixte en raison des grandes différences de variabilité entre les techniques d'interaction qui rendaient l'ANOVA inutilisable. Avec le modèle linéaire mixte (méthode `lme` du package R `nlme`), nous pouvions prendre en compte la disparité des variances. Le modèle prend la technique d'interaction et le numéro de session comme facteurs fixes et le participant comme facteur aléatoire. Nous avons calculé une analyse par paire à partir de ce modèle avec de multiples comparaisons de moyennes en utilisant les contrastes de Tukey (avec la méthode `glht` du package R `multcomp`, qui inclut une correction pour des comparaisons multiples).

Enfin, nous avons produit des scores de corrélation en utilisant des coefficients de corrélation de rang de Spearman pour analyser la corrélation entre le temps de réponse et la quantité de rotation effectuée pendant les tâches d'entraînement. Nous avons également utilisé des coefficients de corrélation de Spearman pour évaluer la corrélation entre les scores de pré-VMRT et les scores obtenus lors des sessions de reconnaissance, afin d'évaluer la pertinence du test de VMRT en tant que tâche d'équilibrage.

4.2.2 Résultats de reconnaissance

La figure 4.6 indique le taux de réussite des participants aux sessions de reconnaissance, ainsi que le temps moyen qu'ils ont passé pour répondre aux essais de reconnaissance. Les résultats sont regroupés en fonction de la technique d'interaction utilisée pendant l'entraînement.

Taux de succès

Nous avons observé un effet du numéro de session sur le taux de réussite ($F(2, 54) = 25,2, p = 1,8e-08$). Le taux de réussite s'est amélioré au cours de l'expérience. Cependant, il n'y a eu aucun effet du groupe ($F(2, 27) = 0,04, p = 0,96$), ni aucune interaction entre le groupe et le numéro de session ($F(4, 54) = 1,51, p = 0,21$). Par conséquent, tous les groupes ont amélioré leur taux de réussite au cours de l'expérience, à un rythme similaire

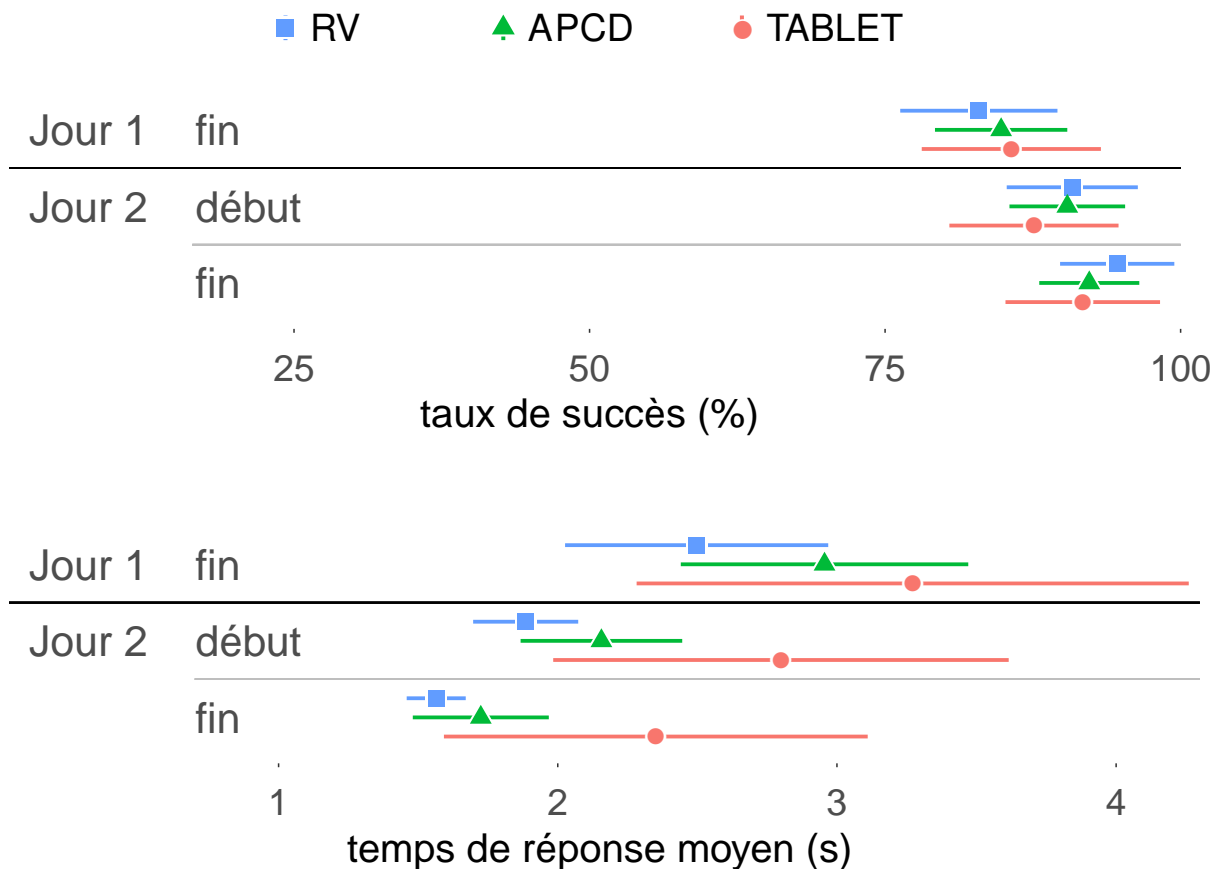


FIGURE 4.6: Taux de réussite (en haut) et temps de réponse moyen (en bas) au cours des trois sessions de reconnaissance, par groupe, avec les intervalles de confiance à 95%.

et atteignant des taux similaires. Le taux de réussite est supérieur à 80% après le premier jour et supérieur à 90% à la fin de l'expérience. Nous nous sommes donc concentrés sur l'analyse des performances de temps de réponse.

Temps de réponse

La figure 4.6 en bas, révèle des différences notables dans le temps de réponse entre les groupes, en particulier pendant le deuxième jour de l'expérience. Elle révèle également des améliorations notables du temps de réponse au fur et à mesure des sessions, dans tous les groupes. Les différences au cours du premier jour ne sont pas robustes. Nous avons exécuté un modèle mixte linéaire sur les données du jour 2. Il a confirmé l'effet global de la technique sur le temps de réponse ($p = 0,013$). L'interaction statistique : technique * numéro de session, a été supprimée du modèle car les données ne montrent pas d'effet ($p = 0,64$). Les tests par paires ont confirmé une différence robuste entre les groupes RV et TABLET ($p = 0,007$). La différence entre les groupes APCD et TABLET est quand à elle moins robuste ($p = 0,064$). En revanche, nous n'avons pas observé de différence robuste entre les groupes RV et APCD ($p = 0,23$). À la fin du deuxième jour, le temps de réponse moyen du groupe TABLET était respectivement 34% et 27% plus lent que pour les groupes RV et APCD.

Nous pouvons remarquer sur la figure 4.6 bas les intervalles de confiance très larges des temps de réponse moyens du groupe TABLET par rapport aux autres groupes. Nous présentons une explication possible dans la section Discussion.

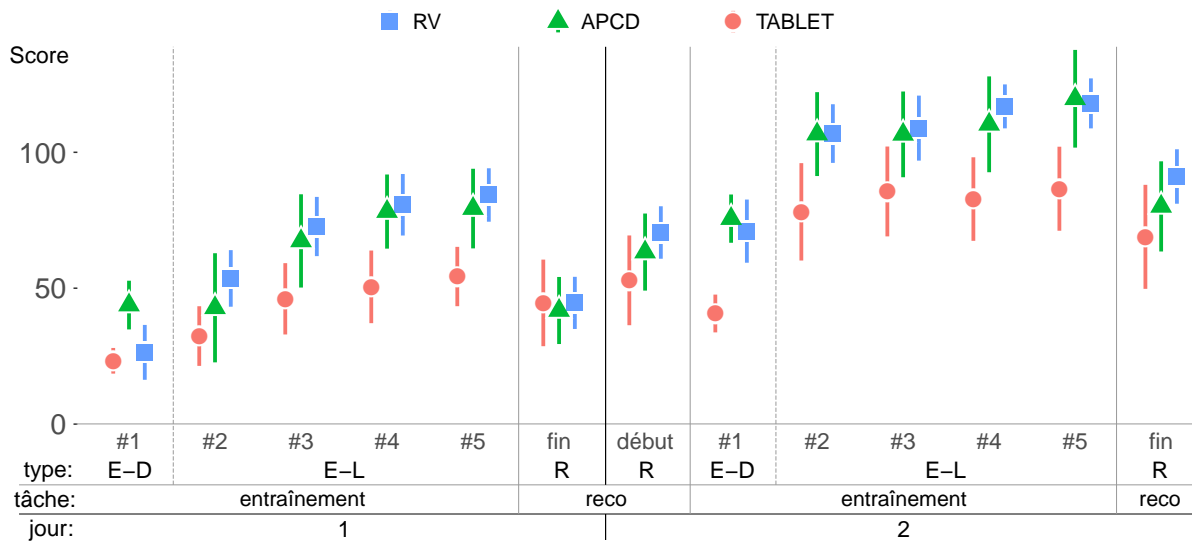


FIGURE 4.7: Score moyen des participants regroupés par technique d'interaction et session, avec les intervalles de confiance à 95%.

Corrélation entre entraînement et reconnaissance

La figure 4.7 montre les scores moyens des participants groupés par technique d'interaction, pour chaque session d'entraînement et de reconnaissance (reco). Elle donne un aperçu de la progression des participants au cours de l'expérience. Nous observons une amélioration globale pour chaque type de sessions. Les scores des sessions de reconnaissance (type R, sans manipulation) et des sessions d'entraînement avec manipulation dictée (type E-D) sont inférieurs à ceux des sessions avec manipulation libre (type E-L).

Comme prévu dans les expériences d'entraînement, les scores des participants ont augmenté au cours des huit sessions d'entraînement avec manipulation libre (effet global de la session : $F(7, 189) = 146,5, p < 2e-16$). Cette progression a suivi un schéma similaire pour toutes les techniques, sans interaction statistique technique * numéro de session claire : $F(14, 189) = 1,19, p = 0,283$.

Nous avons calculé la corrélation entre le score moyen au cours de l'entraînement et le score à la session de reconnaissance post-entraînement, pour chaque participant et chaque jour. Nous avons trouvé un coefficient de corrélation de Spearman de 0,78. Cela confirme un effet positif de l'entraînement pour apprendre le sphinx.

4.2.3 Résultats des VMRT

La figure 4.8 montre les scores moyens aux tests de VMRT pour chaque groupe. Comme le VRMT était notre test d'équilibrage des groupes, il n'est pas surprenant que

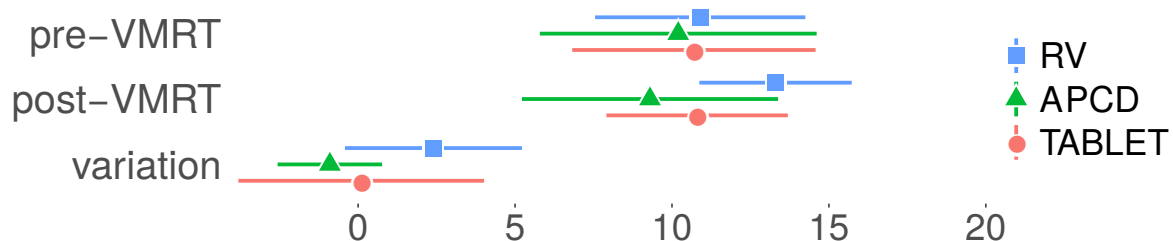


FIGURE 4.8: Scores moyens des pré-tests et post-tests de VMRT, et la différence entre les deux, par groupe, avec les intervalles de confiance à 95%.

	RV	APCD	TABLET	Tous
corrélacion score - VMRT	0.59	0.50	0.78	0.63
p-value	0.074	0.14	0.008	2e-04

TABLE 4.1: Coefficients de corrélation de Spearman (et p-valeur correspondantes) entre les résultats au pré-test de VMRT des participants et leur score de reconnaissance du jour 2, séparés par groupes ou tous regroupés.

les moyennes pré-test soient toutes très similaires. Cependant, nous n’avons trouvé aucun effet robuste du groupe sur la variation entre les pré-tests et les post-tests ($F(2, 27) = 1,267$, $p = 0,298$).

Le tableau 4.1 montre le coefficient de corrélation de Spearman entre les résultats au pré-test de VMRT des participants et leurs scores de reconnaissance moyens du deuxième jour, regroupés par technique d’interaction, et toutes techniques confondues. Le coefficient de corrélation général (0,63, $p = 2e-04$) indique une corrélation positive entre le test VMRT et les scores de reconnaissance. On observe également que la corrélation est plus forte pour le groupe TABLET que pour les deux autres groupes.

4.2.4 Analyse de la quantité de rotation

Le tableau 4.2 montre la corrélation entre le temps de réponse et la quantité de rotation effectuées pendant un essai, selon le type de session et le jour. Pendant les sessions d’entraînement avec manipulation dictée (type E-D : session #1 de chaque jour), le temps de réponse et la quantité de rotations sont fortement corrélés ($> 0,8$). La corrélation est plus faible pendant les sessions d’entraînement avec manipulation libre (type E-L) le premier jour, et elle s’abaisse encore au fur et à mesure de l’avancement de l’expérience, tombant à 0,56 le deuxième jour.

Ces résultats indiquent que plus les participants avancent dans l’expérience, moins ils ont besoin de faire pivoter l’objet cible avec la sphère pour parvenir à l’identifier. Vers la fin de l’expérience, le fait qu’il y ait pu avoir plus de temps d’investi dans la rotation mentale de l’objet que dans les manipulations physiques suggère une internalisation de la tâche [37]. Cette théorie indique que les gens effectuent spontanément des gestes lorsqu’ils résolvent des tâches de rotation mentale, que ces gestes facilitent les tâches, et qu’ils diminuent en quantité avec la pratique.

Bien sûr, la corrélation reste très élevée pendant la session #1 du jour 2, car la manipulation physique était imposée pour pouvoir répondre. Le taux de réussite lors des sessions de type E-D reste supérieur à 0,98 dans tous les groupes. Par conséquent, les faibles scores de ces sessions (visibles dans la figure 4.7) reflètent un temps de réponse plus lent que dans les sessions avec manipulation libre. Cela indique que les rotations physiques sont plus lentes à effectuer que les rotations mentales, quelle que soit la technique

	jour 1	jour 2
type E-D (manipulation dictée)	0.81	0.80
type E-L (manipulation libre)	0.71	0.56

TABLE 4.2: Coefficient de corrélation de Spearman entre le temps de réponse et la quantité de rotation effectuée pendant un essai, séparé par type de session et par jour. Toutes les p-valeurs sont $< 2e-16$.

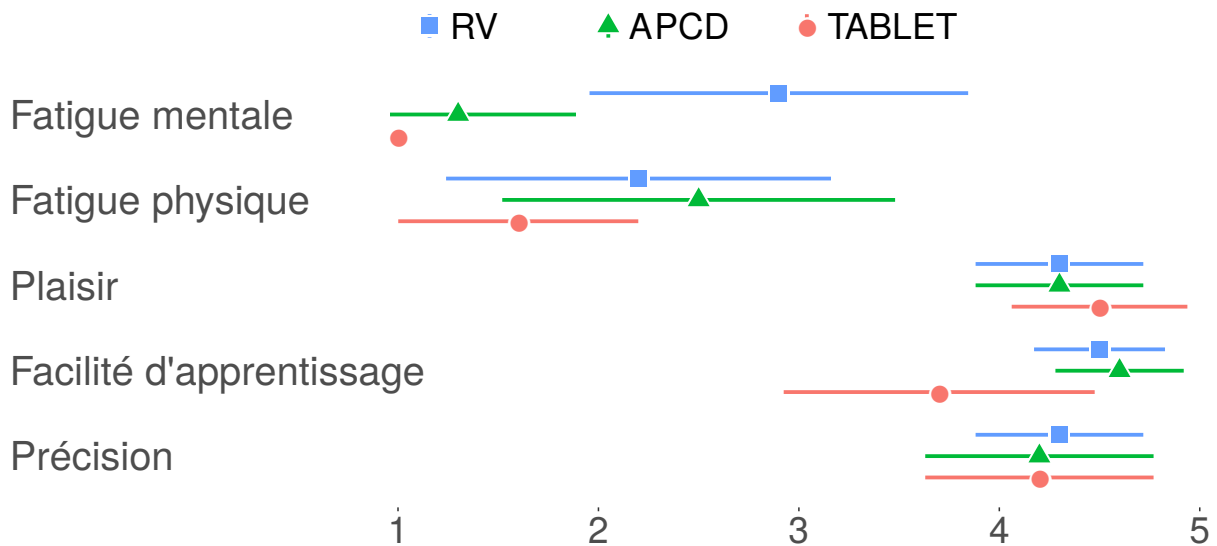


FIGURE 4.9: Notes subjectives moyennes des participants de chaque groupe, avec les intervalles de confiance à 95%.

d'interaction utilisée. De plus, le faible score du groupe TABLET dans la session #1 du jour 2 indique que la tablette est moins efficace pour contrôler les rotations des objets cibles que le casque de RV et l'APCD.

4.2.5 Résultats subjectifs

Les évaluations subjectives sont rapportées dans la figure 4.9.

Le casque de RV est la seule technique d'interaction signalée comme générant des maux de tête. Cela reflète une nouvelle fois cette limitation bien connue des casques opaques qui fait que certains utilisateurs se sentent nauséeux en les utilisant, en raison de la vision déformée par les lentilles et de vertiges de déplacement proches de ceux causés par le mal des transports. L'utilisation de la sphère semble être un peu plus fatigante que celle de la tablette (la note moyenne globale des groupes RV + APCD étant de $2,4 (\pm 1,57)$ contre $1,6 (\pm 0,97)$ pour le groupe TABLET). Les participants de tous les groupes ont trouvé l'expérience amusante, ce qui est encourageant pour une acceptation des systèmes par les étudiants dans des contextes écologiques d'apprentissage. Les retours suggèrent que les trois techniques d'interaction sont suffisamment précises pour exécuter les tâches, ce qui est cohérent avec les taux de réussite globalement élevés observés (c.f. Figure 4.6, en haut). Le casque de RV et l'APCD se sont révélés très instinctifs à utiliser, mais certains participants ont signalé des difficultés à faire pivoter les objets exactement comme ils le souhaitaient avec la tablette. Pour enquêter sur cette difficulté, nous avons analysé les traces enregistrées au cours des sessions d'entraînement avec manipulation dictée (première session d'entraînement de chaque jour). La quantité de mouvements non optimaux (c'est-à-dire les mouvements qui ne tournent pas l'objet cible vers son orientation de référence), était $2,16 \times$ plus élevée lors de l'utilisation de la tablette que lors de l'utilisation de la sphère.

4.3 Discussion

Notre étude utilisateur a été conçue pour évaluer l'effet des interactions 3D haute-fidélité sur l'apprentissage d'objets 3D et sur l'amélioration des capacités de RM. Dans cette section, HF fait référence aux deux groupes RV et APCD. Nos résultats suggèrent que l'apprentissage d'un objet 3D avec une interaction HF améliore la capacité des participants à identifier l'objet en le voyant depuis n'importe quel point de vue, par rapport à un apprentissage via une tablette tactile. Cet avantage n'a pu être observé que lors des tests effectués après une nuit de sommeil.

Les capacités de RM sont étroitement liées aux performances de reconnaissance dans le groupe TABLET contrairement aux groupes HF, ce qui suggère que l'utilisation de techniques d'interaction HF pourrait réduire les inégalités des élèves dans un contexte écologique d'apprentissage.

Nous allons maintenant discuter les résultats obtenus, nos observations en rapport aux capacités de RM, et le protocole mis en place.

4.3.1 Supériorité des groupes HF

Dans les sessions de reconnaissance du jour 2, les temps de réponse étaient plus rapides pour les deux groupes HF que pour le groupe TABLET. L'amélioration des performances mesurées est assez importante (plus de 27%).

Deux observations indiquent que ce résultat est robuste. Premièrement, le taux de réussite moyen du groupe TABLET était inférieur à celui des groupes RV et APCD lors de la dernière session de reconnaissance, c'est-à-dire lorsque l'amélioration du temps de réponse avec les techniques HF était la plus importante. Deuxièmement, la technique d'affichage utilisée pour mesurer les performances de reconnaissance était beaucoup plus proche de l'expérience d'entraînement du groupe TABLET que de celle des groupes HF. En particulier, les participants des groupes HF ont appris le *sphinx* à partir d'un rendu stéréoscopique couplé à la position de l'utilisateur, mais ont été invités à le reconnaître à partir de rendus en perspective simple tels que ceux utilisés dans l'entraînement du groupe TABLET. Les performances des groupes HF sont supérieures à celles du groupe TABLET *malgré* ce biais expérimental favorisant la condition TABLET.

En outre, cela indique que quelle que soit la représentation mentale du *sphinx* que les participants HF ont pu construire, celle-ci était robuste à un changement d'apparence. Ce signe encourageant indique que les interactions HF valent la peine d'être testées en situation écologique, par exemple pour former les étudiants en médecine à reconnaître des organes en utilisant des rendus virtuels haute-fidélité.

RV et APCD : deux alternatives aux performances comparables

Malgré les performances supérieures de l'APCD sur le casque de RV pour contrôler la pose des objets virtuels, reportées au chapitre précédent, les performances de l'APCD ne sont pas supérieures aux performances du casque de RV dans la tâche d'apprentissage d'objets 3D que nous avons évaluée dans cette expérience. Il semble que les informations contextuelles inhérentes à l'APCD ainsi que sa vitesse et sa précision dans les rotations d'objets n'étaient *pas* utiles dans cette étude où l'objet n'avait aucun rapport avec son environnement. Cependant, les participants du groupe RV peuvent avoir bénéficié d'un meilleur contraste autour de l'objet cible contrôlable, en raison de la nécessité d'afficher la sphère avec un fond noir dans cette condition.

Il convient de mentionner que les coûts de déploiement actuels des deux techniques HF que nous avons testées sont très inégaux. Une application de RV haute-fidélité est réalisable aujourd'hui avec un budget modeste et de nombreux logiciels facilement valorisables

tels que Unity 3D. La création d'un niveau similaire de fidélité avec un APCD a nécessité un budget et un effort de développement logiciel beaucoup plus importants. Cependant, l'état du marché des technologies de RV et de RA haute-fidélité évoluera certainement dans les prochaines années, par exemple en réduisant les céphalées et l'isolement provoqués par les casques opaques de RV, ou en réduisant le coût des APCD en utilisant des écrans OLED et un suivi inertiel. Dans tous les cas, notre étude s'est concentrée sur l'utilisabilité fondamentale des différentes approches que nous avons testées. Cela indique qu'il est intéressant de poursuivre les deux approches RV et RA dans un contexte d'apprentissage d'objets 3D, car elles apportent un avantage réel par rapport aux techniques d'interaction de basse fidélité.

4.3.2 Apprentissage d'objets 3D et capacités de rotation mentale

Apprendre avec de faibles capacités de RM grâce aux systèmes HF

Les scores de reconnaissance des participants du groupe TABLET sont fortement corrélés à leurs capacités de RM telles que mesurées par le pré-test de VMRT, résultant en une grande variabilité des scores de reconnaissance pour ce groupe.

Cependant, les coefficients de corrélation des deux autres groupes suggèrent que l'efficacité de l'entraînement HF dépend moins des capacités de RM des utilisateurs. Cela pourrait expliquer les différences de variabilité observées sur la figure 4.6 bas. Cela peut également être révélateur de la mise en place de différentes stratégies mentales dans la reconnaissance des objets 3D selon le dispositif utilisé pendant l'apprentissage, en accord avec les théories de la cognition incarnée ou ancrée [38, 8]. En particulier, nous avons émis l'hypothèse que la reproduction haute-fidélité d'un objet physique offerte par les techniques HF aide à créer une représentation mentale de l'objet, même pour les personnes ayant de faibles capacités de RM, grâce à une expérience sensorimotrice plus *naturelle*. L'expérience naturelle offerte par les systèmes HF réduit les demandes de RM et donc une partie de la charge cognitive extrinsèque nécessaire pour explorer et comprendre des objets 3D. Par conséquent, l'utilisation de systèmes d'interaction haute-fidélité pourrait faciliter l'apprentissage d'objets pour les personnes ayant de faibles capacités de RM, et contribuer ainsi à la réduction d'une inégalité dans les cours scientifiques. Ceci sera à démontrer dans des études ultérieures par exemple dans un cadre pédagogique concret tel que l'apprentissage de l'anatomie.

Le VMRT comme test d'équilibrage d'une tâche d'apprentissage d'objet 3D

Dans le cadre du protocole inter-sujets que nous avons mis en place, nous avons dû nous assurer que les groupes étaient équilibrés en ce qui concerne leur capacité à reconnaître les objets 3D. Nous ne connaissons aucun test qui prédise les performances sur une tâche cognitive aussi complexe, c'est pourquoi nous nous sommes rabattus sur l'utilisation du VMRT.

Les corrélations observées entre les scores de VMRT et les scores de reconnaissance lorsque toutes les conditions d'entraînement sont regroupées (cf. tableau 4.1, dernière colonne) indiquent que la rotation mentale a bien été impliquée dans la reconnaissance et que le VMRT était une mesure pertinente pour équilibrer des groupes vis-à-vis d'une tâche de reconnaissance d'objets 3D. D'autres capacités, telles que des capacités de perception dans l'espace, pourraient aussi être impliquées, en particulier avec les techniques HF, ce qui pourrait expliquer la moindre corrélation observée avec les groupes APCD et RV.

Améliorer les capacités de RM

Nous n'avons trouvé aucune preuve d'une amélioration des capacités de RM, telle que mesurée à l'aide du VMRT.

Nous avons choisi le VMRT pour mesurer la capacité de RM des participants car il est largement utilisé dans la littérature. Cependant, le test ne semble pas suffisamment sensible pour isoler un effet sur une période aussi courte. En effet, nous avons observé une grande variabilité de l'évolution des capacités de RM des participants, quel que soit leur groupe. Certains participants ont indiqué que le test était cognitivement épuisant, ce qui était encore plus problématique lors du post-test qui suivait une expérience déjà exigeante cognitivement. De plus, notre expérience a confirmé que la variabilité des capacités de RM chez les personnes est très importante : nous avons mesuré des scores de VMRT allant de 0 à 24 (la plage totale) lors du pré-test. Dans la littérature, les études utilisant le VMRT pour montrer un effet de différentes pratiques sur les capacités de RM utilisent de grandes cohortes (par exemple 200+ étudiants d'une classe), et/ou s'étalent sur une longue période d'entraînement (plusieurs semaines à un semestre) [110, 49, 72].

Nous pensons toujours, comme le suggère la littérature, que l'entraînement à la reconnaissance d'objets et le temps passé à travailler avec un environnement 3D virtuel pourraient améliorer les capacités de RM générales des utilisateurs [88, 39]. Mesurer un tel effet nécessiterait une expérience à long terme, avec de larges cohortes, qui ne permettrait pas d'atteindre le degré de contrôle des facteurs que nous avons mis en œuvre dans cette étude. Dans un futur travail, nous aimerions mesurer l'effet à long terme d'un apprentissage d'objets 3D qui mette en jeu un APCD, par exemple dans le cadre d'un vrai cours scientifique.

4.3.3 Un protocole expérimental ambitieux

Nous avons mis en place un protocole complexe afin de mesurer plusieurs effets à court et moyen termes, tout en garantissant un haut niveau de contrôle entre les conditions. Nous rapportons ici nos réflexions sur la plus-value d'expériences réparties sur plusieurs jours, et sur les difficultés de fournir un résultat généralisable avec un tel protocole.

Une expérience répartie sur plusieurs jours

Les études sur l'apprentissage dans le domaine éducatif s'étalent généralement sur plusieurs jours, par exemple pour étudier le rôle de l'implication de la gestuelle dans différentes tâches de mémoire [32]. Dans une étude de Cook et al., l'amélioration de l'apprentissage n'a pu être observée de manière robuste que plusieurs semaines après la session d'apprentissage [40]. Les expériences contrôlées sur plusieurs jours sont moins courantes dans le domaine de l'IHM. L'accent est souvent mis sur le contrôle moteur et les processus de perception de bas niveau, qui, en théorie, ne sont pas liées aux capacités cognitives de haut niveau. Pourtant, même les tâches de contrôle moteur peuvent induire des processus d'adaptation cognitive de haut niveau, comme le montrent Cattani et al. dans une expérience à long terme de suivi de cible [31].

Dans notre expérience, les résultats de reconnaissance des trois groupes étaient indifférenciables à la fin de la première journée. Si notre expérience n'avait duré qu'un jour, nous n'aurions pas été capables d'isoler un quelconque effet de la technique d'interaction sur l'apprentissage. La progression des performances de reconnaissance entre la fin du jour 1 et le début du jour 2 est frappante : même si les participants n'ont rien fait en rapport avec l'expérience entre les deux, leurs performances ont augmenté sensiblement dans toutes les conditions. Il semblerait que cet effet puisse être attribué à la consolidation

de la mémoire pendant le sommeil, un phénomène bien étudié en sciences cognitives. Cependant, l'amélioration du score de reconnaissance après une nuit pourrait également être liée à un effet test-retest. Un groupe témoin passant les test et re-test le même jour serait nécessaire pour conclure. Quoi qu'il en soit, cette observation incite à produire davantage d'études d'IHM qui s'étalent sur plus d'une journée, surtout lorsque la mobilisation de processus cognitifs complexes est attendue.

Généralisation des résultats en contexte écologique

Le **sphinx** a été conçu d'après les objets utilisés dans le VMRT. Ce test s'est avéré être en corrélation significative avec les capacités visuospatiales générales d'étudiants de STIM et en particulier avec leur taux de succès en examens d'anatomie [64]. Pour la tâche expérimentale, nous nous sommes inspirés d'une tâche fréquente pour les étudiants en médecine et en STIM : identifier des organes ou des pièces mécaniques à partir de dessins au crayon (c'est-à-dire des objets qui n'ont ni couleur ni texture) et d'un point de vue statique. Cependant, nous n'avons testé qu'un seul objet non écologique, en raison des contraintes de notre protocole expérimental et de la demande importante en temps et en efforts cognitifs pour les participants. Des études supplémentaires sont donc nécessaires pour tester la généralité des résultats de cette expérience.

Conclusion

Dans ce chapitre, nous avons présenté la première étude qui a permis de formellement mesurer les bénéfices de deux systèmes d'interaction HF dans une tâche d'apprentissage d'objets 3D. Dans une expérience utilisateur contrôlée répartie sur deux jours, nous avons comparé trois systèmes : un APCD, un casque de RV utilisant une large sphère tangible comme périphérique de contrôle, et une tablette tactile. L'étude d'un tel processus cognitif de haut niveau a nécessité la conception et la mise en œuvre d'un protocole expérimental conséquent. Elle a révélé une amélioration importante (27%) du temps de reconnaissance après un entraînement mettant en jeu des techniques d'interaction haute-fidélité (fournissant un rendu stéréoscopique couplé à la position de l'utilisateur et un contrôle isomorphe), par rapport à une interaction basse fidélité (avec un rendu en perspective simple et un contrôle tactile). Ce travail contribue à une meilleure compréhension des systèmes d'interaction HF en fournissant une démonstration empirique de leur bénéfice dans une tâche mettant en jeu un processus cognitif complexe. En outre, cette étude devrait créer une forte incitation à poursuivre l'utilisation des approches de RV et de RA dans des études d'apprentissage appliqué dans les domaines du médical et des STIM.

Troisième partie

Comment

Chapitre 5

Mesurer l'impact des paramètres techniques d'un système HF sur la présence

Introduction

Le sentiment de présence (défini section 1.1.4) est ressenti lorsqu'on est exposé à des réalités immersives à l'aide d'appareils haute-fidélité (HF). Il est recherché dans de nombreux domaines d'application comme les téléconférences, les thérapies par l'exposition, les simulateurs d'entraînement et les jeux vidéo, car il permet une interaction intuitive, et car il favorise l'implication émotionnelle. En effet, interagir avec un objet très présent est naturel pour l'utilisateur, ce qui réduit le temps d'apprentissage d'une technique d'interaction et améliore son acceptabilité. Lorsqu'on utilise des simulations pour entraîner des professionnels à gagner des compétences (par exemple pour l'entraînement de pilotes avec des simulateurs de vol), il est crucial que la reproduction virtuelle de la réalité physique soit aussi fidèle que possible, et par conséquent très présente. Cela permet de maximiser le transfert de l'apprentissage du virtuel vers le physique. Dans le domaine des jeux en réalité virtuelle (RV), les utilisateurs désirent être immergés dans le monde du jeu, ce qui nécessite une forte présence des entités virtuelles.

Pour afficher une scène avec un fort niveau de présence, il faut être capable de fournir des moyens d'interaction naturelles couplées à des retours sensoriels crédibles. À ce jour, les appareils HF capables de tels moyens sont en train d'émerger mais ils sont chers et se heurtent toujours à des obstacles techniques. Améliorer la présence d'une entité peut être réalisé en agissant à plusieurs niveaux d'abstraction, depuis les performances d'entrée et de sortie des appareils interactifs jusqu'au scénario d'utilisation. Un rendu pixelisé, par exemple, devrait produire moins de présence qu'un rendu haute résolution qui sature l'acuité visuelle humaine. Améliorer les performances techniques des systèmes HF est coûteux, et il est vraisemblable que certains paramètres soient plus critiques que d'autres à optimiser pour favoriser la présence ressentie. La littérature manque de travaux expérimentaux sur le sujet.

La présence étant un sentiment subjectif, elle n'est pas triviale à mesurer. Pour ce faire, nous avons vu en section 2.3 que de nombreuses études s'appuient sur des questionnaires subjectifs. Les rares travaux empiriques qui étudient l'impact des performances techniques des systèmes d'interaction sur la présence sont difficiles à comparer étant donné le manque de mesures objectives.

Dans ce chapitre, nous proposons une nouvelle méthode pour mesurer l'impact de divers paramètres sur la présence perçue d'une scène. Notre méthode repose sur une

répétition statistique de comparaisons entre des paires de conditions plutôt que sur un questionnaire direct sur le sentiment de présence. Nous avons conçu cette méthode dans le but d’obtenir une mesure de présence plus objective que les questionnaires. Notre méthode pourrait améliorer la comparaison des études futures et aider à comprendre les facteurs ayant un impact sur la présence de scènes virtuelles.

La troisième contribution de cette thèse consiste à avoir mis en œuvre cette méthode dans une étude utilisateur qui interroge l’impact de quatre paramètres majeurs influençant la perception visuelle d’une scène : la Résolution en pixels de l’écran, la Latence de bout en bout, la Fréquence de rafraîchissement de l’affichage et le Tremblement induit par le système de capture de mouvement (RLFT). Nous nous sommes concentrés sur les paramètres influençant la qualité du stimulus visuel perçu, en accord avec l’effet de capture visuelle mis en évidence dans notre état de l’art. Le système interactif testé est notre prototype d’APCD sphérique, qui est actuellement l’un des systèmes les plus performants au regard de ces paramètres techniques. Cette étude donne un premier aperçu de l’impact de ces quatre paramètres sur la présence.

Dans ce chapitre, nous commençons par expliquer la nouvelle méthode d’évaluation de la présence que nous proposons. Nous présentons ensuite l’étude utilisateur que nous avons mise en œuvre pour tester l’impact des paramètres RLFT sur la présence en utilisant la méthode d’évaluation introduite, ainsi que les résultats obtenus. Enfin, nous discutons ces résultats, notamment dans le but de fournir des lignes directrices de conception pour les constructeurs et les développeurs qui cherchent à optimiser le sentiment de présence ressenti par leurs utilisateurs.

Sommaire

Introduction	124
5.1 Une nouvelle méthode pour évaluer la présence	126
5.1.1 Une mesure relative plutôt qu’une échelle absolue	126
5.1.2 Présentation de la méthode	127
5.2 Conception de l’étude utilisateur	129
5.2.1 Choix des paramètres expérimentaux	129
5.2.2 Protocole expérimental	131
5.3 Résultats de l’étude	134
5.3.1 Méthodologie	134
5.3.2 Effet de l’entité	134
5.3.3 Résultats des comparaisons	135
5.3.4 Résultats subjectifs	136
5.4 Discussion	138
5.4.1 Analyse des résultats	138
5.4.2 Compromis de performance	140
5.4.3 Pertinence de la méthode d’évaluation de la présence présentée	140
Conclusion	142

5.1 Une nouvelle méthode pour évaluer la présence

5.1.1 Une mesure relative plutôt qu’une échelle absolue

Le plus souvent, la présence est évaluée à l’aide de questionnaires subjectifs. Les participants doivent noter la présence ressentie lors de différentes expériences sur une échelle absolue qu’ils ne comprennent pas forcément. Estimer la valeur absolue d’une expérience, ou une distance absolue entre deux expériences, sur une échelle inconnue est difficile. Cela entraîne une grande variabilité inter-participants qui ne permet pas de mettre en valeur les effets les plus discrets. De plus, chaque estimation n’est généralement effectuée qu’une seule fois par participant, ce qui empêche les calculs statistiques de base (comme la moyenne et la variabilité), qui améliorent la précision de la mesure et renseignent sur sa robustesse. À l’inverse, utiliser une technique qui permettrait de mesurer directement la présence ressentie par les participants, par exemple en se basant sur des signaux biologiques, semble illusoire, car la présence est un sentiment subjectif. Nous présentons ici une nouvelle approche fondée sur l’observation suivante : s’il est difficile de mesurer la présence sur une échelle absolue, il est beaucoup plus facile de classer deux expériences immédiates l’une par rapport à l’autre en fonction des sentiments de présence ressentis. Pour illustrer ce propos, faisons un parallèle avec la hauteur des notes de musique : reconnaître la hauteur d’une note unique nécessite de posséder l’oreille absolue, qui est un don rare. En revanche, la plupart des gens sont capables d’identifier la note la plus haute parmi deux notes entendues.

Notre approche a été inspirée par le protocole de la différence minimale perceptible (DMP, “just noticeable difference” en anglais) utilisé dans des recherches récentes en interaction Homme-machine pour mesurer la plus petite latence perceptible sur un système [107]. Le protocole DMP consiste à présenter des paires de stimuli et à demander aux participants d’en sélectionner un parmi les deux, par exemple en choisissant le meilleur ou le pire. Pour chaque bonne réponse, la différence entre les stimuli est légèrement réduite. En revanche, la différence est considérablement agrandie en cas de mauvaise réponse. Cela se produit lorsque les deux stimuli sont trop proches et que le participant ne peut plus percevoir la différence. Les changements dans les stimuli et l’ordre de présentation sont cachés aux participants afin que leurs réponses ne soient pas influencées par l’évolution des paramètres. Ce protocole converge de manière robuste vers la plus petite différence perceptible par le participant. Avec le protocole expérimental de Ng et al., l’un des stimuli de chaque paire est toujours la latence la plus faible atteignable avec leur système (1 ms). En supposant que 1 ms est inférieur au seuil absolu de latence perceptible par tous les participants, leur approche change l’objectif du protocole DMP de la mesure d’une plus petite *différence* perceptible entre deux latences à la mesure d’un *seuil absolu* de perception. Le protocole DMP peut aussi être considéré comme un moyen de mesurer objectivement une question subjective, qui est dans leur cas : “quelle est la plus petite latence du système que vous pouvez percevoir?”.

Le protocole DMP tel que mis en œuvre par Ng et al. est utilisé pour trouver avec précision un seuil pour un seul paramètre. Il n’est donc pas directement applicable pour mesurer l’effet relatif de divers paramètres hétérogènes sur le sentiment de présence. Un autre obstacle important est qu’un système qui crée la présence parfaite n’existe pas. Le monde physique ne peut pas être utilisé car ses paramètres tels que la latence ou la résolution ne peuvent pas être réduits sans que les participants ne s’en rendent compte. Néanmoins, nous gardons le principe d’une répétition de nombreuses comparaisons entre paires de stimuli et celui de l’ignorance par les participants du fait que tel ou tel paramètre soit modifié.

5.1.2 Présentation de la méthode

Le principe de notre méthode est donc d'exposer les participants à deux variantes d'une expérience virtuelle, que nous appelons *représentations*. Nous leur demandons laquelle des deux représentations génère le plus fort sentiment de présence. Pour faciliter cette *comparaison*, les participants peuvent basculer à volonté entre les deux représentations avant de donner leur réponse. À tout moment, lors d'une comparaison, les participants ont donc accès à deux commandes : l'une pour changer la représentation courante, l'autre pour identifier la représentation courante comme celle générant le meilleur sentiment de présence, comme schématisé dans la figure 5.1. Une seule comparaison ne donne qu'une réponse binaire. Cependant, nous faisons en sorte que les comparaisons soient rapides : de l'ordre de dix secondes. De cette façon, nous pouvons proposer de nombreuses comparaisons pour lesquelles nous contrôlons le paramètre modifié dans chaque représentation, et le *niveau* de modification appliqué ; ce qui permet une mesure plus fine des effets relatifs obtenus pour chaque paire de paramètres. Les différences entre les représentations ne sont pas expliquées afin que les participants ne puissent se fier qu'à leur sentiment pour faire leur choix. De plus, une même comparaison sera proposée plusieurs fois lors de la même session expérimentale. Cela améliore la précision de la mesure en permettant le calcul d'une moyenne, et fournit une estimation de la robustesse de la mesure en exploitant la variabilité des réponses fournies par un participant à chaque occurrence d'une même comparaison.

Chaque représentation correspond à l'une des *conditions expérimentales*. Une représentation est obtenue en optimisant chaque paramètre étudié sauf celui qui est altéré négativement jusqu'à un certain niveau. Deux conditions expérimentales diffèrent donc, soit par le paramètre qui est altéré dans chaque condition (comme par exemple la latence ou la résolution), soit par le niveau de l'altération pour un même paramètre. L'approche permet de tester des paramètres très différents, tels que les possibilités d'interaction ou l'utilisation d'un rendu stéréoscopique, tant que le participant peut basculer rapidement

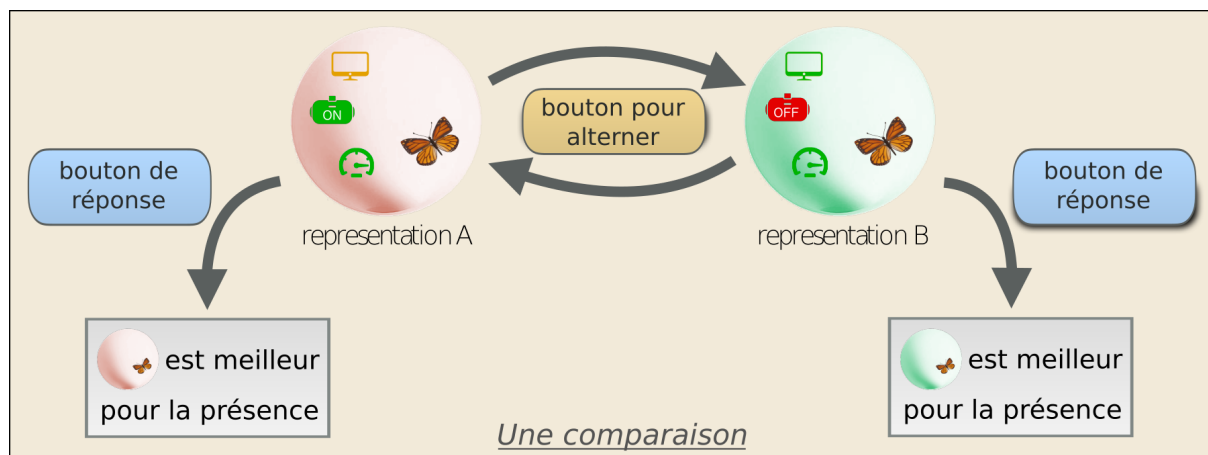


FIGURE 5.1: Représentation schématique d'une comparaison. Un bouton permet à l'utilisateur de basculer à tout moment entre les deux représentations. Un autre bouton permet d'identifier la représentation courante comme la plus présente. Pour simplifier, on considère dans cet exemple trois paramètres : la résolution, la stéréoscopie et la fréquence de rafraîchissement (icônes d'écran, de casque de RV et de compteur de vitesse respectivement). Dans la représentation A, la résolution est altérée au niveau orange. Dans la représentation B, la stéréoscopie est altérée au niveau rouge. La fréquence de rafraîchissement n'est pas altérée dans ces deux représentations. Le niveau vert correspond au niveau optimal de chaque paramètre. La scène présentée (un papillon) ne doit pas changer entre les deux représentations d'une comparaison.

et de manière répétée entre les représentations. Dans un souci de généralisation, cette méthode est répétée avec différentes scènes virtuelles.

À l’instar des méthodes psychophysiques, notre méthode s’appuie sur le sentiment de présence des participants pour évaluer l’influence de plusieurs paramètres. Mais contrairement aux méthodes présentées par Bergstrom et al. [12], les participants ne sont pas familiarisés avec les paramètres et ils ne peuvent pas les manipuler librement. En cachant les informations et le contrôle sur les paramètres, nous nous assurons que les participants ne sont pas influencés par leurs propres attentes concernant les paramètres.

5.2 Conception de l'étude utilisateur

Nous avons mis en place une étude utilisateur mettant en jeu la méthode d'évaluation de la présence présentée ci-dessus, afin de mesurer l'impact de différents paramètres sur le sentiment de présence. Dans cette section, nous détaillons nos choix de paramètres et de niveaux d'altération étudiés, puis nous présentons comment nous avons mis en pratique la méthode d'évaluation dans un protocole expérimental.

5.2.1 Choix des paramètres expérimentaux

Choix des paramètres techniques

Nous avons choisi d'utiliser notre prototype d'APCD pour cette étude. En effet, les valeurs optimales de notre prototype dépassent les casques de RV commerciaux en termes de tremblement du système de capture de mouvement et de résolution d'affichage et les égalisent en termes de latence. Afin d'isoler l'effet de chaque paramètre technique, nous avons implémenté une solution qui permet d'altérer indépendamment la valeur de chaque paramètre étudié par rapport aux valeurs optimales de référence de notre prototype.

Selon Cummings et al. [43], les paramètres techniques les plus étudiés et ayant le plus d'impact sur la présence sont la résolution d'affichage, la fréquence d'affichage, le tremblement du dispositif de capture (qui est responsable de l'imprécision du dispositif), le champ de vision et la stéréoscopie. Nous avons modifié cet ensemble de paramètres avec les justifications suivantes. Les entités virtuelles créées se comportent comme si elles étaient physiquement piégées à l'intérieur de l'APCD. Par conséquent, le champ de vision de l'APCD contient forcément entièrement l'espace disponible pour l'entité virtuelle, quelle que soit la taille du dispositif. Nous avons donc exclu le champ de vision de cette étude. Nous avons constaté dans une étude pilote que l'absence de stéréoscopie lors de l'utilisation d'un APCD brisait le sentiment de présence. Nous avons donc considéré que la stéréoscopie est une condition nécessaire pour permettre un bon niveau de présence, et nous l'avons retiré de l'étude. L'effet de la latence du système sur la présence n'a pas été étudié en profondeur, mais plusieurs études ont montré son fort impact négatif sur la qualité de l'interaction [30, 107, 112]. Nous soupçonnons un impact négatif similaire sur la présence et avons donc choisi de l'inclure dans l'étude. Nous avons donc testé l'effet de quatre paramètres sur la présence : la Résolution d'affichage, la Latence, la Fréquence de rafraîchissement et le Tremblement du système de capture de mouvement (RLFT).

Choix des niveaux d'altération

Pour tous les paramètres, nous avons testé au moins un niveau d'altération en plus du niveau optimal. Pour chaque paramètre, le niveau *optimal* était notre meilleur effort d'implémentation, c'est-à-dire le meilleur niveau que nous pouvions atteindre avec notre système expérimental. Pour toutes les représentations, l'entité était affichée dans la sphère de polystyrène de 30 cm de diamètre utilisée comme APCD sans fil. La capture de mouvement a été mise en œuvre avec le système de suivi optique Optitrack (8 caméras Prime 13 à 240 Hz). Nous avons enregistré les variations de position retournées par le système lors de la capture d'un marqueur statique pour évaluer le tremblement de capture et observé un bruit blanc gaussien de 0.022 mm d'erreur standard. Les images ont été créées par un projecteur Barco F-50 (2560x1600 pixels @ 120 Hz). La stéréoscopie active était mise en œuvre à l'aide des lunettes alternantes. Nous avons évalué la latence totale du système en utilisant l'approche de Cattan et al. [30]. Les niveaux optimaux suivants ont été atteints pour les quatre paramètres : résolution à 90 dpi dans la zone de mise au point (la résolution diminue lorsqu'on s'approche des bords de la sphère), latence à 27 ms, fréquence de

rafraîchissement à 60 Hz par œil et tremblement du système de capture à 0.022 mm.

Pour déterminer les niveaux d'altération pertinents selon l'état de l'art, nous avons diagnostiqué les performances RLFT des dispositifs de RV et de réalité augmentée (RA) présents sur le marché. Nous avons évalué deux casques de RV (HTC Vive et Vive pro), deux casques de RA (Metavision Meta 2 et Microsoft HoloLens) et une tablette de RA (Tango).

Les niveaux d'altération ont été choisis comme suit :

- **Résolution.** La majorité des tablettes compatibles avec la réalité augmentée, ainsi que le casque HoloLens, ont une résolution de l'ordre de 100 dpi ; ce qui est supérieur aux 90 dpi de notre implémentation APCD. La plupart des casques actuels ont une résolution beaucoup plus faible, de l'ordre de 50 dpi, bien que des casques de RV haute résolution deviennent disponibles. Nous avons choisi un premier niveau d'altération à 50 dpi. Nous avons également testé un deuxième niveau d'altération, reconnaissant qu'il est courant d'abaisser la résolution de rendu de manière logicielle pour conserver une fréquence d'affichage acceptable avec des scènes virtuelles de haute complexité. Nous avons choisi 35 dpi ($\sqrt{2} * 50dpi = 35dpi$) car cela correspond à la réduction de résolution nécessaire pour réduire de moitié le nombre de pixels à partir du premier niveau d'altération.
- **Latence.** La latence totale des systèmes utilisés pour créer des entités virtuelles n'est généralement pas publiée. Parfois, une prédiction logicielle est utilisée pour compenser une partie de la latence. Ce genre de prédiction a l'inconvénient de démultiplier les tremblements de la capture de position. Nous avons choisi d'éviter les prédictions dans cette étude pour nous concentrer sur l'impact fondamental de la latence et du tremblement. Néanmoins, la prédiction utilisée automatiquement dans l'HoloLens nous a empêchés d'estimer sa latence de base. Nous avons mesuré une latence de 28 ms pour le vive pro (presque identique à notre valeur optimale), et d'environ 60 ms pour le Meta 2. Ng et al. rapportent que la latence des appareils tactiles commerciaux est comprise entre 50 ms et 200 ms [107]. Nous avons choisi d'utiliser un seul niveau d'altération pour la latence à 60 ms.
- **Fréquence d'affichage.** Les taux de rafraîchissement sont faciles à évaluer, car ils sont gérés par le pilote de la carte graphique, qui les signale au système. Nous avons observé une fréquence de rafraîchissement native (c'est-à-dire optimale) de 60 Hz pour le HoloLens, 72 Hz pour le Meta 2, 90 Hz pour les Vive et Vive pro. La fréquence de rafraîchissement de la plupart des tablettes est de 60 Hz. Le matériel actuel offre donc un affichage stéréoscopique à au moins 60 Hz, mais le rendu de scènes complexes entraîne souvent une baisse de la fréquence de rafraîchissement atteignable. Nous avons choisi un niveau d'altération unique à 30 Hz pour illustrer les limites de rendu. Ce niveau de fréquence de rafraîchissement représente également ce que de la stéréo multiplexée dans le temps fournirait sur les tablettes à 60 Hz.
- **Tremblement du système de capture.** Nous avons mesuré une erreur standard de tremblement de 0.13 mm, 0.20 mm et 1.7 mm pour le Vive pro, le HoloLens et le Meta 2 respectivement. Nous avions d'abord l'intention de les utiliser comme 3 niveaux d'altération en plus du niveau optimal de notre système (0.022 mm). Cependant, nous avons déterminé grâce à des études pilotes que le tremblement de 0.13 mm était imperceptible et que celui de 1.7 mm rompait complètement le

sentiment de présence dans le cas de l'APCD. Nous avons choisi de ne pas inclure le niveau à 0.13 mm, de diviser par deux le niveau à 1.7 mm et d'inclure une valeur entre les deux niveaux d'altération restants. Nous avons donc testé 0.2 mm, 0.5 mm et 0.8 mm.

5.2.2 Protocole expérimental

Entités et tache

Dans un souci de généralisation, nous avons choisi deux entités de complexité différentes dans leur comportement et leurs possibilités d'interaction. La première entité était un fémur humain. Les participants pouvaient explorer l'os en tournant et en déplaçant la sphère, et l'os restait statique dans le système de coordonnées de la sphère, comme s'il était physiquement attaché à l'intérieur de celle-ci. Le second était un papillon qui agissait comme s'il était encagé dans la sphère : il pouvait marcher à la surface de la sphère ou voler à l'intérieur. Quand le papillon marchait, les participants pouvaient le forcer à prendre son envol s'ils secouaient la sphère. L'animation était aléatoire pour paraître plus réaliste. Les deux entités sont illustrées dans la figure 5.2. La tâche que nous avons demandé aux participants d'accomplir était d'explorer librement les entités présentées en changeant leur point de vue et en manipulant la sphère (y compris en la secouant).

Implémentation de la méthode d'évaluation

Nous avons implémenté la méthode d'évaluation de la présence présentée dans la section précédente. Chaque représentation était associée à un seul niveau d'altération d'un seul paramètre RLFT. En d'autres termes, chaque représentation était notre meilleur effort d'implémentation, à l'exception d'un paramètre qui était modifié. Le tableau 5.1 résume toutes les représentations utilisées dans l'expérience. Chaque représentation est nommée d'après son paramètre et son niveau altéré. Il convient de noter que le meilleur effort d'implémentation n'a pas été utilisé en tant que représentation.

L'ensemble de comparaisons présenté au participant était constitué de toutes les paires de représentations possibles, à l'exception des paires de représentations d'un même paramètre à différents niveaux. Une même comparaison était présentée 8 fois pour chaque

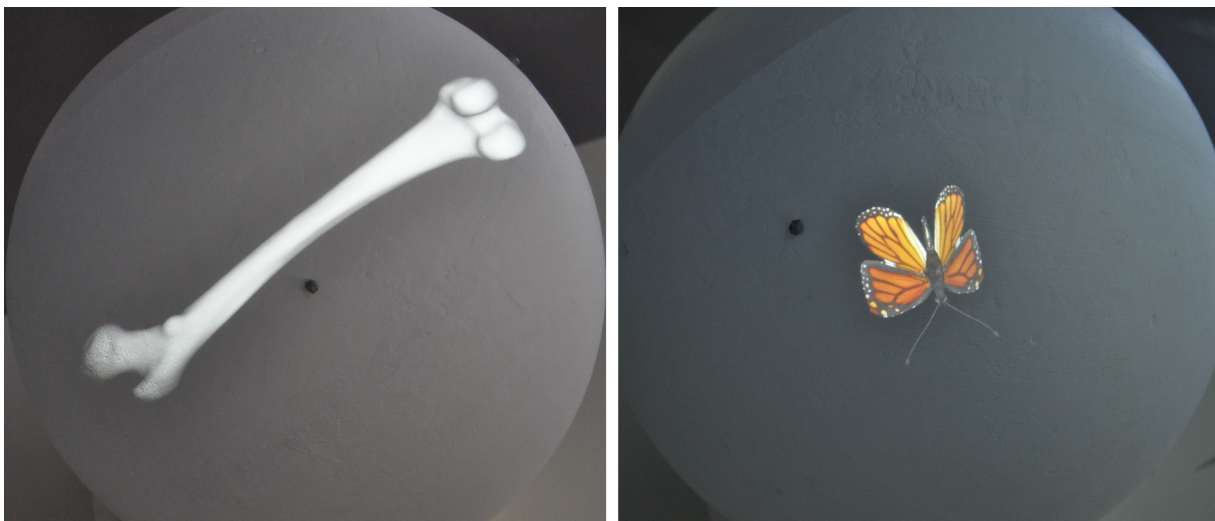


FIGURE 5.2: Les deux entités utilisées. À gauche : l'os statique. À droite : le papillon animé quand il vole à l'intérieur de la sphère. Pour mieux comprendre l'illustration, la stéréo a été désactivée sur cette image.

Dénomination	résolution (dpi)	latence (ms)	rafraîchissement (Hz/œil)	tremblement (mm)
Resolution 50	50	27	60	0.02
Resolution 35	35	27	60	0.02
Latence 60	90	60	60	0.02
Rafraichissement 30	90	27	30	0.02
Tremblement 0.2	90	27	60	0.2
Tremblement 0.5	90	27	60	0.5
Tremblement 0.8	90	27	60	0.8

TABLE 5.1: La valeur de chaque paramètre RLFT pour chaque représentation comparée. Les valeurs en bleu sont les valeurs optimales du paramètre donné.

participant (4 fois par entité), aboutissant à $17 \times 8 = 136$ comparaisons par participant. La séance expérimentale étant assez longue (de l'ordre de 90 minutes), nous l'avons fractionnée sur deux jours pour limiter les effets de la fatigue. Les comparaisons ont été présentées dans un ordre aléatoire et séparées en 8 blocs (4 par jour). L'entité affichée était la même au sein d'un bloc et était commutée entre chaque bloc. Les participants pouvaient prendre une courte pause entre chaque bloc.

En début de session expérimentale, nous expliquions la définition de la présence aux participants. Nous avons utilisé la définition vulgarisée suivante : "On dit d'une représentation qu'elle est plus présente qu'une autre si elle est plus fidèle à ce que vous pourriez observer dans le monde physique. Ainsi, vous ne pourriez pas différencier une scène virtuelle parfaitement présente d'une scène non virtuelle." À la fin de la première journée d'expérience, les participants remplissaient un court questionnaire avec trois questions conçues pour mettre en évidence les différences subjectives entre les deux entités présentées. Pour les deux entités, ils devaient rendre compte : de la variabilité des niveaux de présence vécus (sur une ligne droite sans échelle), de la facilité avec laquelle ils avaient pu faire leur choix, et du plaisir qu'ils avaient eu à explorer la scène (sur une échelle de 1 à 5 pour les deux dernières questions).

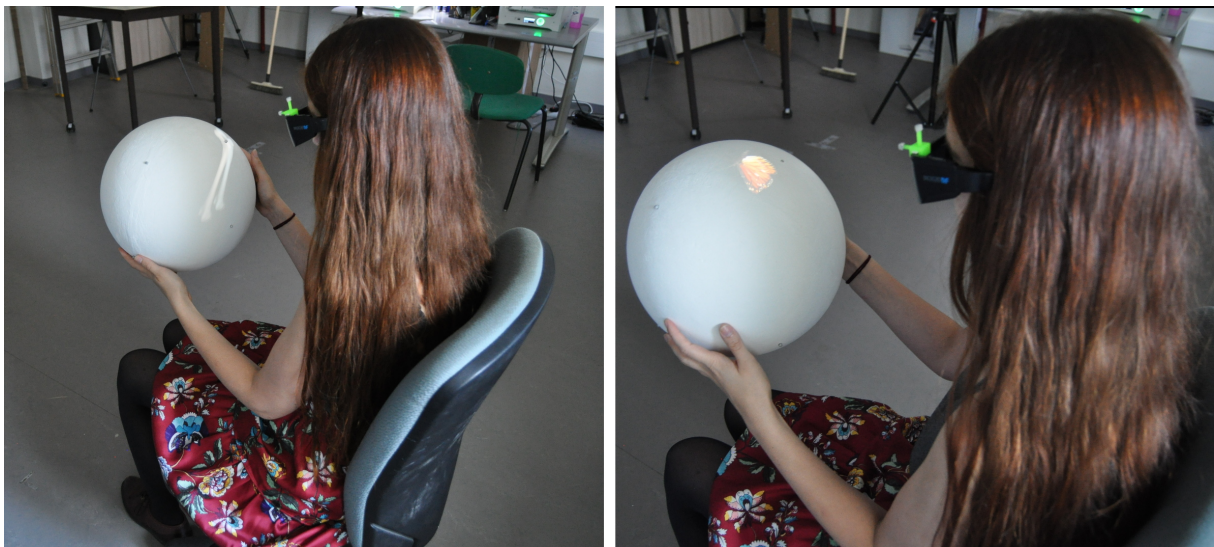


FIGURE 5.3: Une participante explorant une représentation de l'entité fémur (à gauche) et de l'entité papillon (à droite). La lumière du jour était couverte pendant l'expérience pour maximiser le contraste du projecteur.



FIGURE 5.4: Des marqueurs sont attachés aux chaussures de la participante pour détecter les “taps”.

Participants

Nous avons accueilli 16 participants : 8 femmes, 28.2 ans d’âge moyen (de 23 à 37 ans). Tous avaient une vision stéréoscopique correcte (testée avec le test de vision stéréoscopique optique RANDDOT), et aucun n’avait peur des papillons. La figure 5.3 montre une participante explorant deux représentations : une pour chaque entité. Tout en explorant, elle peut taper avec son pied droit à tout moment pour basculer entre les deux *représentations*. Elle pourra identifier celle qui génère le plus de présence en tapant avec son pied gauche alors qu’elle est active. Un tap est illustré dans la figure 5.4.

5.3 Résultats de l'étude

5.3.1 Méthodologie

Nous avons enregistré la préférence des participants en termes de présence pour chaque comparaison de deux représentations. Pour calculer la moyenne et les intervalles de confiance, nous avons associé des valeurs numériques aux réponses des participants. Étant donné une comparaison entre A et B, nous associons une valeur de 1 chaque fois que A a été préféré et une valeur de -1 chaque fois que B a été préféré.

Dans la figure 5.5, la figure 5.6 et la figure 5.7, chaque cellule représente une comparaison A vs B où A et B correspondent aux titres de la ligne et de la colonne de la cellule respectivement. Dans un souci de lisibilité, nous n'avons pas présenté d'informations redondantes : les représentations correspondant à une altération de résolution sont présentées uniquement sous forme de lignes et les représentations correspondant à une altération de tremblement uniquement sous forme de colonnes.

Nous avons calculé la probabilité d'explication des résultats par des erreurs d'échantillonnage et de mesure aléatoires (p-valeur de l'hypothèse nulle) avec des régressions logistiques pour chaque comparaison. Nous avons utilisé les méthodes R `glmer`, `anova` et `glht` des packages `lme4` et `multcomp`.

5.3.2 Effet de l'entité

Nous avons effectué une analyse de la déviance, avec correction en une seule étape, pour tester l'effet de l'entité sur les préférences des participants. Les résultats sont présentés dans la figure 5.5. Ils indiquent que le sentiment de présence des participants diffère en fonction de l'entité surtout lorsque la résolution est impliquée, mais également pour la comparaison fréquence 30 Hz versus tremblement 0.8 mm. Inversement, les résultats semblent être similaires pour les deux entités lorsque la résolution n'est pas impliquée.

Corrélation entre l'importance de l'effet et la durée de la tâche

Pour les deux entités, nous avons calculé le facteur de corrélation de Spearman entre l'importance de l'effet pour chaque comparaison et le temps moyen passé par les participants pour répondre à la comparaison. Nous obtenons un facteur de corrélation de -0,78

		latence	fréquence	tremblement		
		60 ms	30 Hz	0,2 mm	0,5 mm	0,8 mm
résolution	50 dpi	p= 6e-05	p= 0.005	p= 8e-05	p= 0.002	p= 1.000
	35 dpi	p= 8e-10	p= 9e-08	p= 1e-05	p= 0.003	p= 0.023
latence	60 ms		p= 0.470	p= 0.087	p= 0.111	p= 0.998
fréquence	30 Hz			p= 0.528	p= 0.160	p= 8e-04

FIGURE 5.5: Effet de l'entité : probabilité de l'hypothèse nulle de différence de préférence moyenne en fonction de l'entité, pour chaque comparaison. Plus une cellule est claire, plus un effet statistique de l'entité est probable pour la comparaison correspondante.

pour l'os et de -0,12 pour le papillon. Nous supposons qu'un effet important correspondrait à une comparaison facile, c'est-à-dire une comparaison où la meilleure représentation en terme de présence est évidente et donc nécessite un faible temps de réponse pour être identifiée. C'est ce que nous avons observé avec l'os, pour lequel le temps passé dans chaque comparaison est fortement lié à la difficulté de discernement des effets des altérations, mais nous n'avons pas observé cette corrélation avec le papillon.

5.3.3 Résultats des comparaisons

Pour chaque participant, nous avons calculé la réponse moyenne sur les quatre répétitions avec la même entité pour chaque comparaison. Nous avons ensuite calculé la moyenne et l'écart type entre les participants pour chaque comparaison, pour chaque entité. Ces résultats sont illustrés sur la figure 5.6. Nous avons calculé la p-valeur pour chaque comparaison en testant l'hypothèse linéaire des régressions logistiques. Les p-valeurs sont représentées dans la figure 5.7.

Une altération souvent préférée indique que le paramètre n'a pas un fort impact négatif sur la présence. Cependant, cela ne signifie pas que ce paramètre a un impact positif sur la présence car la présence résulte de la combinaison de tous les paramètres. Inversement, une altération rarement préférée indique que le paramètre a un fort impact négatif sur la présence. Nous nous concentrons donc sur les altérations les moins préférées.

Tremblement

Nous avons observé que le tremblement de 0.8 mm n'a jamais été préféré, ce qui indique que le tremblement est un paramètre critique pour la présence et qu'un tremblement élevé a un fort impact négatif sur celle-ci. Cet effet est robuste (avec de faibles p-valeurs) et important (avec des préférences moyennes contre cette condition à 0,78 ou plus), sauf en comparaison avec la fréquence de rafraîchissement de 30 Hz avec l'os.

Même avec un niveau d'altération de 0.5 mm, le tremblement a été jugé plus préjudiciable que la résolution et la latence pour les deux entités. Les effets observés sont moins importants qu'avec le niveau de tremblement le plus élevé mais ils restent robustes. Ce

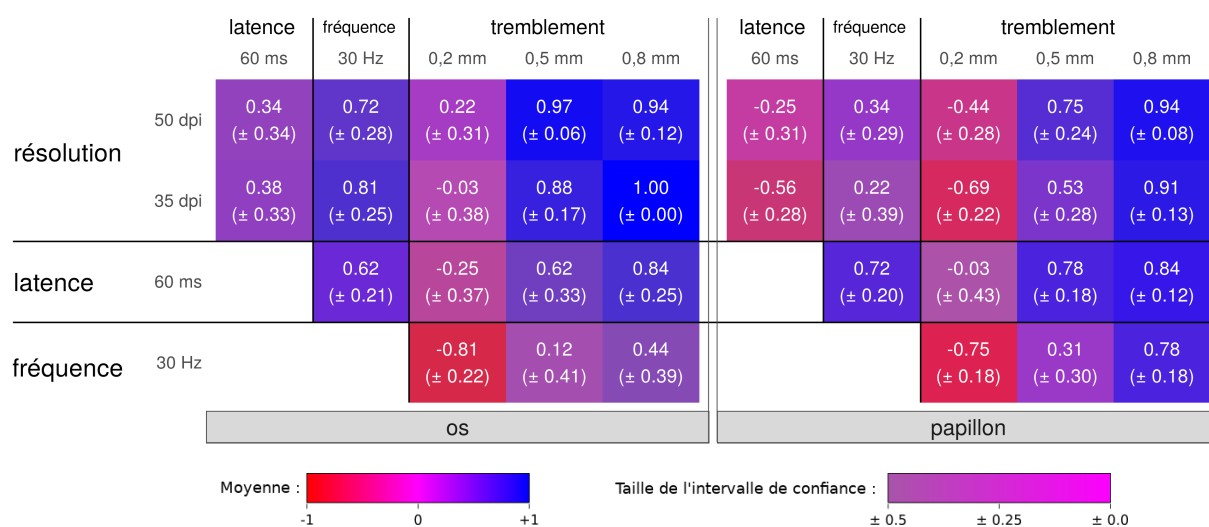


FIGURE 5.6: Préférences moyennes des utilisateurs par comparaison, avec les intervalles de confiance à 95%. Une valeur de 1 signifie que la condition définie par la ligne de la cellule a toujours été préférée à la condition définie par la colonne de la cellule. -1 signifie le contraire. À Gauche : résultats pour l'os. À droite : résultats pour le papillon. La teinte de la cellule code la moyenne et la saturation code la taille de l'intervalle de confiance.

		latence	fréquence	tremblement			latence	fréquence	tremblement			
		60 ms	30 Hz	0,2 mm	0,5 mm	0,8 mm	60 ms	30 Hz	0,2 mm	0,5 mm	0,8 mm	
résolution	50 dpi	p= 0.043	p= 6e-05	p= 0.117	p= 0.006	p= 0.006	p= 0.110	p= 0.038	p= 0.003	p= 0.031	p= 0.006	
	35 dpi	p= 0.030	p= 5e-04	p= 0.607	p= 1e-05	p< 1e-15	p= 0.001	p= 0.236	p= 4e-05	p= 0.003	p= 0.025	
latence	60 ms		p= 3e-05	p= 0.089	p= 0.005	p< 1e-15		p= 3e-06	p= 0.430	p= 6e-04	p< 1e-15	
fréquence	30 Hz			p= 1e-05	p= 0.264	p= 0.054			p= 1e-05	p= 0.047	p= 0.004	
						os	papillon					

FIGURE 5.7: Robustesse des comparaisons : probabilités de l’hypothèse nulle (p-valeurs) des préférences rapportées dans la figure 5.6. Plus la case est claire, plus le résultat est statistiquement robuste.

niveau de tremblement semble cependant avoir un effet proche de celui de la fréquence de rafraîchissement de 30 Hz : pour ces comparaisons, les préférences sont plus proches de 0, les réponses sont plus variables et la p-valeur indique un effet fragile ($p = 0,26$ pour l’os).

Le plus petit niveau de tremblement que nous avons testé (0.2mm) est beaucoup moins préjudiciable pour la présence. Il a été préféré à l’altération de la fréquence de rafraîchissement pour les deux entités et aux altérations de résolution pour le papillon. Les autres comparaisons sont moins claires, et indiquent un effet similaire à ceux des altérations de résolution pour l’os et à celui de l’altération de latence pour les deux entités.

Fréquence de rafraîchissement

La fréquence de rafraîchissement de 30 Hz est plus préjudiciable que la latence de 60 ms pour les deux entités, et plus préjudiciable que les deux altérations de résolution pour l’os. La fréquence de rafraîchissement de 30 Hz semble être plus néfaste que la résolution de 50 dpi avec l’entité papillon, même si ce dernier effet est plus petit et moins robuste (préférence moyenne de 0,34, $p = 0,038$). Enfin, la fréquence de rafraîchissement de 30 Hz semble équivalente à la résolution de 35 dpi avec l’entité papillon (préférence moyenne de 0,22, $p = 0,236$).

Latence

La latence est moins préjudiciable avec le papillon que les deux altérations de résolution, même si les résultats ne sont pas très clairs avec la résolution de 50 dpi (préférence moyenne de -0,25, $p = 0,11$). Avec l’os, la latence semble être légèrement plus préjudiciable que la résolution (valeurs de préférence moyennes de 0,34 et 0,38 pour les conditions de résolution de 50 dpi et 35 dpi, $p < 0,043$).

5.3.4 Résultats subjectifs

Nous avons demandé aux participants de représenter une plage de présence pour chaque entité, qui délimite toutes les représentations qu’ils ont explorées avec l’entité donnée, c’est-à-dire de la représentation qu’ils ont jugée la moins présente à celle qu’ils ont jugée la plus présente. La plage a été exprimée sur une échelle continue sans graduation allant de l’absence de tout sentiment de présence (min) à un sentiment de présence



FIGURE 5.8: Plage de présence de toutes les représentations explorées par entité, représentée par les utilisateurs. Les côtés gauche et droit des cases représentent respectivement la moyenne des limites minimale et maximale de la plage. Les intervalles de confiance à 95% sont représentés par les segments horizontaux colorés.

absolue (max). Les résultats sont reportés dans la figure 5.8.

La plage exprimée par participant est assez large : plus de 50% de l'échelle totale. Cela indique que la perte de présence dans les pires conditions a été clairement ressentie par les participants. Pour les deux entités, les résultats sont extrêmement similaires.

Pour les deux questions : “J’ai aimé explorer la scène” (plaisir d’exploration) et “En moyenne, je pouvais facilement identifier la représentation la plus présente dans chaque comparaison” (facilité d’identification), les participants ont répondu sur une échelle de 1 (pas du tout) à 5 (totale­ment). Les résultats sont illustrés sur la figure 5.9. Les participants ont légèrement préféré explorer la scène avec le papillon mais ils ont trouvé la tâche de comparaison plus facile avec l’os.

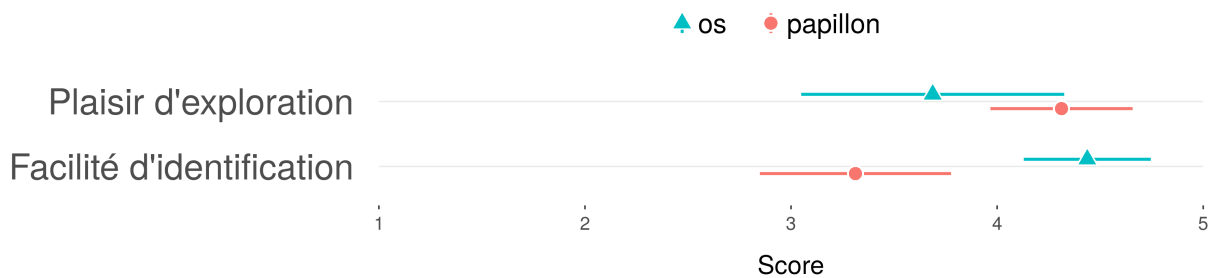


FIGURE 5.9: Plaisir subjectif moyen et difficulté de la tâche, par entité, avec les intervalles de confiance à 95%.

5.4 Discussion

5.4.1 Analyse des résultats

Effet de l'entité

Nous avons observé des résultats similaires avec les deux entités, sauf lorsque la résolution d'affichage est impliquée. Les deux entités étudiées sont assez différentes : l'os est un simple maillage non texturé tandis que le papillon est composé d'un maillage très détaillé avec des ailes texturées et très colorées. Cela pourrait expliquer pourquoi les altérations de résolution ont été plus préjudiciables à la présence dans le cas du papillon. Ce résultat souligne en quoi la nature de l'entité explorée peut imposer des exigences techniques spécifiques pour créer un fort sentiment de présence. De la même façon, les participants n'avaient pas à interagir de manière précise avec les entités présentées dans cette étude, ce qui peut expliquer pourquoi l'altération de la latence n'a pas eu un fort effet néfaste sur la présence. Un stylo virtuel, par exemple, pourrait avoir une exigence plus forte en termes de latence.

Analyse générale des comparaisons

En analysant les résultats des différentes comparaisons effectuées par les utilisateurs, on peut en déduire un ordre des différentes altérations testées en fonction de leur impact négatif sur le sentiment de présence perçu. Cet ordonnancement est présenté dans la figure 5.10. Les altérations y sont soit séparées par entité, soit regroupées lorsque les résultats ne dépendent pas de l'entité. Les rectangles regroupent les conditions qui partagent la même classe d'équivalence.

Le tremblement de 0.8 mm était la condition avec le plus d'impact négatif sur la présence ressentie, alors que le tremblement de 0.2 mm était l'une des conditions avec l'impact le plus faible, aux côtés de la latence de 60 ms. La fréquence de rafraîchissement de 30 Hz et le tremblement de 0.5 mm appartiennent à la même classe d'équivalence dans leur impact sur la présence avec les deux entités, et à la même classe que la condition de résolution à 35 dpi dans le cas du papillon.

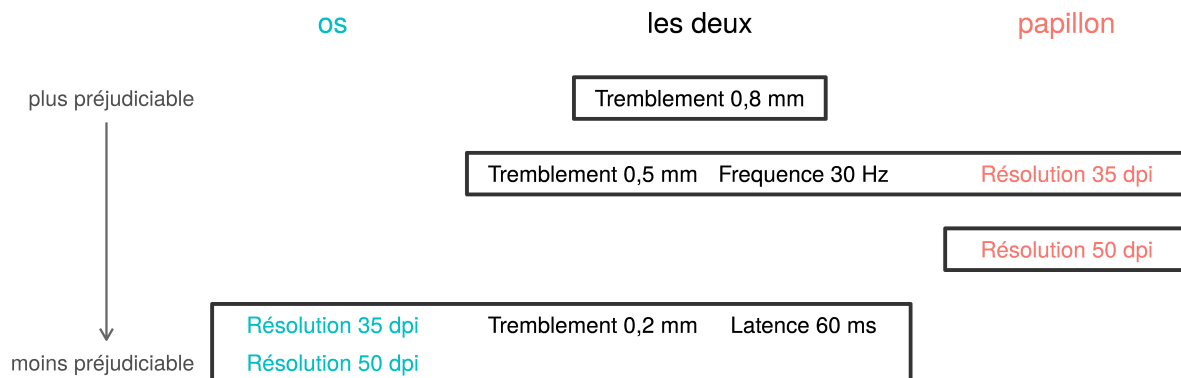


FIGURE 5.10: Altérations ordonnées par la force de leur impact négatif sur la présence, du plus préjudiciable (en haut) au moins préjudiciable (en bas). Les rectangles regroupent les altérations qui appartiennent à la même classe d'équivalence. L'effet des altérations de résolution dépend de l'entité.

Retours informels des participants

Certains participants ont rapporté que le papillon semblait plus présent que l'os car il se déplaçait de lui-même. Ils ont mis du temps à jouer avec lui et à observer ses mouvements, comme ils l'auraient fait avec de vrais papillons. Trois participants lui ont également donné un nom. Ces observations expliquent en partie la corrélation plus faible entre l'importance des effets mesurés et le temps de réponse des utilisateurs avec le papillon : même lorsqu'une comparaison était facile à résoudre, certains participants ne répondaient pas immédiatement car ils avaient envie de jouer avec le papillon. Cependant, certains participants ont été dérangés par l'animation des ailes : ils ont trouvé que l'animation n'était pas très crédible, réduisant la présence ressentie. Un participant a trouvé l'os extrêmement convaincant en tant qu'os en plastique imprimé en 3D, d'où le résultat obtenu d'une présence totale pour les meilleures représentations. Deux participants ont rapporté que la latence, en particulier avec l'os, produisait un effet de flottement, comme si l'os était immergé dans un récipient rempli de formol. Ceci est une autre illustration de la façon dont le cerveau humain parvient à produire une explication rationnelle à un stimulus étrange, comme avec l'effet ventriloque. Un autre participant a signalé que les "vibrations" de l'os lorsqu'il était censé être statique (c'est-à-dire avec un tremblement élevé) brisaient fortement la présence, car un objet statique qui se trouve en équilibre stable ne bouge jamais dans la "vraie vie". Dans ce cas, l'absence d'explication physique crédible semble avoir rompu la présence.

Plusieurs participants ont signalé que les mouvements du papillon rendaient difficile la comparaison des représentations, alors qu'ils pouvaient facilement reproduire un point de vue similaire pour les deux représentations à comparer avec l'os statique. Cela explique probablement pourquoi la tâche a été jugée plus difficile avec le papillon qu'avec l'os. Enfin, le fait que les participants aient également aimé explorer l'os statique sans s'ennuyer (score de plaisir supérieur à 3,5 sur 5) témoigne d'une forte implication dans la tâche lorsque l'utilisateur ressent un fort sentiment de présence, comme le suggèrent des travaux précédents [35].

Limites de l'étude

Comme dans toute étude utilisateur, le temps et l'endurance des participants sont limités. Nous avons donc dû limiter le nombre de conditions testées. Les résultats doivent être interprétés en tenant compte de ces limites. En particulier, nous n'avons testé les paramètres qu'un par un. Certains paramètres pourraient produire des effets qui dépendent du niveau d'autres paramètres. De plus, pour l'instant, nous n'avons obtenu des résultats que pour le cas d'une tâche particulière, avec deux entités spécifiques et un système d'interaction particulier. Des études supplémentaires sont nécessaires pour tester la généralité de ces résultats. D'autres systèmes HF plus répandus tels que des casques de RV et de RA pourraient être utilisés, bien que les limitations techniques actuelles de ces dispositifs puissent empêcher une comparaison directe des résultats observés avec les résultats obtenus au moyen de notre APCD sphérique. De plus, il serait particulièrement intéressant de tester d'autres tâches impliquant une interaction plus exigeante entre les participants et l'entité virtuelle, afin d'analyser plus largement le compromis latence-tremblement. Enfin, des entités plus complexes, telles que des avatars humains, pourraient être testées, bien que les principales limitations en termes de présence puissent provenir de la qualité du modèle humain plutôt que des paramètres techniques du stimulus visuel.

5.4.2 Compromis de performance

Résolution versus fréquence de rafraîchissement

L’affichage de scènes complexes peut réduire la fréquence de rafraîchissement affichable lorsque le temps de calcul de chaque image est supérieur à la période de rafraîchissement. Notre étude indique que la réduction de la résolution d’affichage pour conserver la fréquence d’image optimale est une bonne stratégie pour optimiser la présence ressentie, tant que l’utilisateur n’a pas un besoin d’interaction particulier avec des entités très détaillées. Malgré tout, même pour le papillon avec une texture détaillée, passer de 90 dpi à 50 dpi était moins préjudiciable pour la présence que de passer de 60 Hz à 30 Hz. Cela pourrait s’expliquer par le degré de plausibilité des deux phénomènes vis-à-vis des lois physiques. En effet, il existe de nombreuses explications physiques qui puissent justifier de ne pas voir tous les détails d’une entité observée : du brouillard, des lunettes ou une fenêtre sales, de l’humidité oculaire, etc. Ce n’est pas le cas pour les mouvements saccadés.

Latence versus tremblement de capture

L’étude a révélé que le tremblement du système de capture de mouvement pourrait avoir un effet néfaste important sur la présence. Une façon de réduire le tremblement est de lisser les données de suivi sur plusieurs fenêtres temporelles de capture, ce qui a pour contrepartie d’ajouter un délai supplémentaire à la latence totale du système. Notre étude indique que cela pourrait être une bonne stratégie en termes de présence, car l’altération de latence a eu un impact plus faible sur la présence que les deux plus grandes altérations de tremblement, et un impact similaire à la plus petite altération de tremblement. Dans notre expérience, les participants ont observé des entités virtuelles résidant à l’intérieur d’une sphère. La vibration des entités due au tremblement de capture ne semblait pas avoir d’interprétation physique plausible, tandis que le décalage du mouvement des entités créait l’illusion que les entités étaient immergées dans un fluide visqueux. Cependant, ce compromis peut dépendre de la tâche. Pour une tâche qui demanderait une haute précision d’interaction avec une entité, le compromis opposé peut être plus approprié, comme il a été montré en termes de performance de pointage [112]. De plus, nous n’avons testé qu’un faible écart de latence entre 27 ms et 60 ms. Il serait intéressant d’étudier dans quelle mesure le compromis tient lorsque l’on dépasse les 60 ms de latence.

5.4.3 Pertinence de la méthode d’évaluation de la présence pré-sentée

Nous avons observé que la méthode de mesure de la présence, basée sur des comparaisons relatives entre différentes conditions que nous avons introduites, avait un niveau élevé d’acceptation : tous les participants ont déclaré qu’ils comprenaient clairement la tâche. Dans le questionnaire subjectif, la question de la facilité à répondre à la tâche avait un score supérieur à 3,2 sur 5 pour les deux entités. La méthode a révélé plusieurs effets importants (par exemple, un score de 0,81 pour un tremblement de 0.2 mm par opposition à une fréquence de rafraîchissement de 30 Hz) qui semblent robustes ($p < 1e-05$ pour les deux entités).

Bien que nous ayons utilisé la méthode dans une étude centrée sur un petit ensemble de paramètres techniques, nous pensons qu’elle peut être utilisée pour tester de nombreux autres paramètres. Par exemple, elle pourrait être utilisée pour tester l’effet sur la présence de paramètres non techniques tels que la complexité de la tâche ou la plausibilité de l’entité. En outre, la méthode pourrait être utilisée pour tester en profondeur un seul

paramètre avec de nombreux niveaux d'altération, ce qui la rapprocherait du protocole DMP.

Cette première étude basée sur la méthode d'évaluation proposée montre la viabilité de l'approche. Cependant, des études complémentaires sur l'impact de ces paramètres sur la présence seront nécessaires pour s'assurer de la cohérence des résultats et ainsi tester la validité externe de la méthode.

Conclusion

Les améliorations récentes des appareils de RA et de RV haute-fidélité ont abouti à des expériences digitales qui produisent un sentiment de présence qui se rapproche du sentiment de présence absolu. Cependant, il reste à fournir une analyse méthodologique des facteurs entrant en jeu pour créer ce sentiment. Une façon de favoriser cette analyse méthodologique consiste à unifier le concept de présence dans les différents domaines de recherche concernés, afin que les résultats des études puissent être plus facilement comparés. La méthode d'évaluation de la présence que nous avons introduite n'est liée à aucun système, configuration interactive ou domaine de recherche particulier. L'étude utilisateur mettant en œuvre cette méthode a apporté deux résultats concrets : elle fournit une preuve de la viabilité et de l'utilité de la méthode, et une première mesure de l'effet relatif sur la présence de quatre paramètres importants du stimulus visuel dans le contexte d'un APCD sphérique. Ces mesures procurent des directives de conception pertinentes : éviter un tremblement de mouvement perceptible et une perte de fréquence d'affichage étaient les exigences les plus importantes avec notre configuration. Réduire la résolution ne réduit pas le sentiment de présence pour les entités qui n'ont pas de détails fins. Et contrairement à nos attentes, une latence perceptible du système n'a pas eu un fort effet négatif sur la présence. Cette contribution apporte des premiers éléments de réponse pour savoir comment concevoir des systèmes d'interaction HF qui favorisent un fort sentiment de présence.

Chapitre 6

Comment implémenter une application entièrement interactive avec un APCD

Introduction

Les appareils à perspective corrigée déplaçables (APCD) sont des écrans 3D manipulables, dont l’affichage est couplé à la fois à la position de l’utilisateur et à la position de l’appareil lui-même ; ce qui permet de créer l’illusion forte de tenir la scène virtuelle, ou de tenir un portail à travers lequel la scène virtuelle est visible, comme illustré dans la figure 6.1. Bien que leurs champs de vision soient restreints par rapport à ceux des casques de réalité augmentée (RA) et de réalité virtuelle (RV), ces appareils offrent un haut niveau de présence des scènes explorées. Aussi, nous avons discuté de leurs avantages par rapport à d’autres systèmes haute-fidélité (HF) dans le chapitre 1, et de leur supériorité dans une tâche de placement dans l’espace dans le chapitre 3. Ces résultats laissent penser que les APCD sont des options viables pour mettre en œuvre des applications autonomes réalistes telles que de la modélisation 3D ou des simulations pour la formation.

Cependant, à l’heure actuelle, seuls les éléments les plus fondamentaux de l’interaction 3D tels que la sélection et la manipulation d’objets ont été étudiés dans le contexte des APCD [74]. Pour concevoir des applications autonomes, il est nécessaire qu’un ensemble complet d’interactions soit disponible en plus du contrôle du point de vue et du pointage spatial. Mine [103] a identifié cinq formes d’interactions fondamentales dans les environnements 3D virtuels : la *mouvement* de l’utilisateur, la *manipulation* des entités, la *sélection*, la *mise à l’échelle* et la possibilité d’interagir via des *menus/widgets*. Les inter-



FIGURE 6.1: Un utilisateur manipulant un APCD.

actions via les menus et les widgets n’ont jamais été explorées pour les APCD et sont donc la principale motivation de cette contribution. Nous présentons également des solutions permettant une sélection et une manipulation efficace d’objets, et la mise à l’échelle de la scène.

La contribution de ce chapitre est double : nous contribuons à la conception d’une part, et à l’évaluation qualitative d’autre part, d’une première panoplie de techniques d’interaction, permettant l’utilisation des APCD dans des applications 3D complètes. En particulier, nous introduisons une combinaison fluide de pointage direct et de lancer de rayon pour la sélection d’objets. Nous proposons également un ensemble de techniques d’interaction pour la sélection de commandes dans un menu hiérarchique et le contrôle de paramètres. Ces techniques tirent parti des capacités rotatives de l’appareil. Nous rapportons une évaluation utilisateur qualitative, où nous avons testé un sous-ensemble des techniques conçues dans une application autonome d’apprentissage de l’anatomie.

Dans la première section de ce chapitre, nous commençons par discuter les particularités de notre contexte d’étude qui vont guider nos choix de conception. En particulier, nous présentons la métaphore du *hublot* permettant l’exploration d’une large scène virtuelle. Nous justifions le choix de nous concentrer sur des appareils sphériques tenus à deux mains, qui offrent notamment trois contrôles de rotation continus et indépendants (tangage/roulis/lacet). Nous présentons les techniques d’interaction issues de la littérature sur la HF que nous avons dû adapter pour répondre aux spécificités des APCD.

Nous décrivons ensuite les différentes techniques que nous avons conçues et mises en œuvre dans une panoplie interactive complète. Enfin, nous présentons l’application d’apprentissage d’anatomie que nous avons implémentée pour mettre en jeu la panoplie que nous avons conçue, ainsi que l’étude utilisateur qualitative que nous avons mise en place pour tester l’utilisabilité de cette application.

Sommaire

Introduction	143
6.1 Analyse du contexte d’utilisation ciblé	145
6.1.1 Un hublot volumique, tangible et portable	145
6.1.2 Interagir avec des environnements virtuels 3D	146
6.1.3 S’adapter aux contraintes physiques de l’approche	147
6.1.4 Compromis présence-efficacité	149
6.2 Conception d’une panoplie interactive pour APCD	150
6.2.1 Sélection d’objets	150
6.2.2 Manipulation d’objets	154
6.2.3 Zoom	155
6.2.4 Conteneurs de Widgets	157
6.2.5 Widgets	159
6.3 Démonstration d’usage de la panoplie interactive	164
6.3.1 Une application pour apprendre l’anatomie	164
6.3.2 Évaluation qualitative	166
6.3.3 Perspectives	168
Conclusion	170

6.1 Analyse du contexte d'utilisation ciblé

6.1.1 Un hublot volumique, tangible et portable

Notre travail se concentre sur les techniques d'interaction pour un affichage volumique tangible portable qui constitue à la fois un hublot et une poignée vers le monde virtuel.

Implémentation

Nous avons vu dans la section 1.2, que des travaux récents ont utilisé des volumes en plastique transparent pour créer un effet de hublot tangible dans différents contextes de RV. Cependant, l'utilisation d'objets trop lourds (plus d'1 kilo) rendait les interactions fatigantes. A notre connaissance, aucune étude de RA n'as mis en jeu cet effet. Des efforts de recherche ont expérimenté diverses surfaces tangibles pour l'exploration 3D en utilisant soit un casque de RA [92] soit de la projection externe [141]; mais les deux études se sont limitées à l'exploration spatiale et à la sélection d'objets.

Comme nous l'avons vu dans la section 1.3, l'utilisabilité des APCD pour réaliser des hublots volumiques tangibles a été montré théoriquement par Issartel et al. [74]; mais l'utilisation d'appareil cubiques est source d'occlusions et de discontinuité dans l'affichage. Des écrans sphériques ont été développés pour éliminer ces problèmes [9, 19, 52, 170]. Les écrans sphériques sont généralement construits à partir d'un ensemble de projecteurs assemblés à la base d'une sphère tronquée; ce qui ne confère pas (ou très peu) de mobilité à l'écran. La stéréoscopie étant difficile à mettre en œuvre avec des écrans cubiques ou sphériques, certains systèmes ne s'appuient que sur le couplage de l'affichage à la position de l'utilisateur pour créer une illusion de profondeur [170, 52, 148, 142]. Cependant, une étude récente a observé que la stéréoscopie avait un fort effet positif sur les performances des utilisateurs dans des tâches de sélection [51]. Implémenter un volume tangible portable qui combine les qualités d'un affichage sphérique et stéréoscopique reste un défi à l'heure actuelle, mais nous avons démontré que c'était possible à l'aide de projections externes : notre prototype d'APCD, qui réunit toutes ces conditions, est décrit dans la section 1.3.2.

Boule à neige versus hublot

Nous avons identifié deux modes principaux pour la relation spatiale entre l'appareil et la scène virtuelle. La scène virtuelle est soit *fixée à l'écran*, soit *fixée au monde*, selon le fait qu'elle soit respectivement fixée de manière rigide à l'écran ou au monde physique entourant l'utilisateur.

Lorsque la scène entière est suffisamment petite pour tenir à l'intérieur de l'appareil, le mode d'affichage fixé à l'écran crée la métaphore physique d'une *boule à neige* [25, 142]. Le changement de point de vue sur la scène se fait intuitivement en manipulant l'appareil. De nombreuses applications 3D, cependant, nécessitent de naviguer dans de grandes scènes virtuelles, c'est-à-dire plus grandes que l'appareil. Le mode d'affichage fixé au monde est plus approprié dans ce cas. Avec des scènes fixées au monde, l'appareil agit comme un *hublot* permettant de percevoir un monde virtuel autrement invisible, se superposant au monde physique de la même manière que le "boom chameleon" [150], mais avec une échelle 1 :1.

Ces deux modes peuvent être combinés pour pouvoir modifier la scène : les utilisateurs activent un déclencheur pour attacher un objet d'intérêt à l'appareil, tandis que tous les autres objets de la scène virtuelle restent fixés au monde [74]. Le déplacement de l'appareil déplace ainsi l'objet d'intérêt dans un mouvement 6D isomorphe par rapport à la scène jusqu'à ce que le déclencheur soit relâché.

Un hublot volumique tangible tenu à deux mains peut bénéficier à toute forme de systèmes d'interaction HF. À l'heure actuelle, l'approche la plus convaincante pour tester cette métaphore interactive est l'implémentation d'APCD avec projection externe. À l'avenir, une mise en œuvre plus performante pourrait provenir des améliorations des casques de RV et de RA, ou d'APCD sous la forme d'écrans OLED sphériques autonomes équipés de capteurs inertiels. Bien qu'elles soient principalement conçues pour les APCD, les interactions introduites dans ce chapitre devraient se généraliser à tout type d'implémentation de volumes tangibles tenus à deux mains.

6.1.2 Interagir avec des environnements virtuels 3D

Pointer un objet virtuel

Les systèmes d'interaction 3D HF utilisent généralement une technique de *pointage direct* : un curseur virtuel est mappé de manière isomorphe au mouvement de la main de l'utilisateur. Un objet est pointé lorsque le curseur entre en contact avec lui. Le pointage direct est souvent considéré comme le moyen le plus intuitif de pointer des objets pour sa similitude avec la façon dont nous atteignons les objets dans le monde physique [97]. Sur certains systèmes tels que la RA mobile, le pointage direct peut être un défi à mettre en œuvre, mais sa faisabilité a été récemment démontrée par Qian et al [120].

Le pointage direct ne permettant pas d'accéder à des objets hors de portée des bras, plusieurs techniques alternatives ont été introduites. Bowman et al. ont classé ces techniques en deux catégories principales : extension du bras et lancer de rayons [20]. Ils ont montré que le lancer de rayons était plus efficace qu'une technique d'extension du bras (go-go [116]) pour acquérir des cibles éloignées, mais qu'il était moins efficace pour la manipulation en raison de l'effet bras-levier. Grossman et al. ont également observé la supériorité d'une technique de lancer de rayon pour de la sélection en environnement dense dans leur aquarium de VR sphérique [61].

Interagir avec des widgets

Le pointage sur des menus et des boutons est très efficace dans les interfaces 2D d'ordinateur de bureau grâce à la précision et la stabilité de la souris. Les menus 2D peuvent être étendus à la 3D en utilisant une technique de pointage 3D [58], mais le pointage à main levée est moins précis que la souris et génère plus de fatigue [28]. Les applications utilisant les casques de RV grand public ont désormais tendance à exploiter des techniques de lancer de rayon pour pointer sur des widgets 2D à l'aide de contrôleurs 6D. Avec les casques de RA, Pourmemar et al. ont observé que la sélection d'éléments dans un menu radial était plus rapide et générerait moins de fatigue lorsque les gestes étaient effectués avec la tête plutôt que des mains [117]. Englmeier et al. ont testé la sélection tactile de commandes affichées à la surface d'une sphère en plastique tenue à deux mains [50]. L'affichage était réalisé avec un casque de RV. Ils signalent des plaintes concernant le pointage au doigt lorsque les deux mains doivent tenir la sphère. Pour éviter ce problème, nous utiliserons uniquement le toucher du doigt sur l'écran pour l'activation plutôt que pour le pointage.

La validation de la sélection d'un widget repose souvent sur l'actionnement d'un bouton physique, ce qui peut entraîner un décalage parasite du pointeur. Des alternatives ont été explorées pour éviter ce problème. Pourmemar et al. ont utilisé une validation temporisée avec un temps d'attente de 1.5 s (la validation se fait automatiquement si le pointeur reste dans la cible pendant tout le temps d'attente). Muller-Tomfelde a spécifiquement étudié le temps d'attente optimal pour valider des actions de pointage manuel et a recommandé une plage de 350 ms to 600 ms [106]. Il a également souligné l'importance d'un

retour visuel progressif. Une temporisation introduit un retard et nécessite que le pointeur reste stable pendant ce délai. L'interaction par franchissement évite ce délai en validant dès que le pointeur franchit les frontières de la cible [1]. Tu et al. ont récemment étudié l'interaction par franchissement avec un pointage par lancer de rayon dans un contexte de casque de RV [151]. L'interaction par franchissement s'est avérée plus efficace que la validation par appui sur un bouton du contrôleur.

6.1.3 S'adapter aux contraintes physiques de l'approche

Un portail vers un monde virtuel

Contrairement aux casques de RV qui offrent de grands champs de vision pour percevoir le monde virtuel, les APCD n'affichent le contenu qu'au niveau de leur surface, et donc avec un champ de vision restreint. Afin de naviguer dans la scène virtuelle, l'utilisateur doit constamment déplacer l'appareil, comme avec les systèmes de RA mobile. On décrit alors l'APCD comme un "portail" à travers le monde virtuel. Notez qu'il est également possible de tourner autour de l'écran pour déplacer le champ de vision, mais un tel déplacement nécessite un mouvement plus large et génère plus de fatigue physique.

Avec la métaphore du hublot, nous avons testé trois façons de découper la scène virtuelle. Avec un **clipping complet**, la scène n'est visible que là où elle intersecte le volume occupé par l'appareil, comme illustré dans la figure 6.2, à gauche. Les utilisateurs comprennent facilement l'illusion de conteneur qui en résulte, mais le volume visible est petit. Avec un **clipping avant**, la scène virtuelle est découpée uniquement entre l'appareil et l'utilisateur. L'écran montre ce qui est à l'intérieur et à l'arrière du volume de l'appareil avec une profondeur infinie, comme illustré dans la figure 6.2, au milieu. Ceci est perçu comme une fenêtre *fermée* vers le monde virtuel (c'est-à-dire que le monde virtuel ne passe pas par la fenêtre). **Sans clipping**, l'écran agit comme une fenêtre ouverte sur le monde virtuel : les objets *devant* l'appareil sont également visibles, comme illustré dans la figure 6.2, à droite.

Afin de fournir autant d'informations contextuelles que possible, nous avons implémenté le **sans clipping** dans l'application d'apprentissage de l'anatomie. Les appareils de RA mobiles sont généralement utilisés avec un **clipping avant**. Avec un rendu stéré-

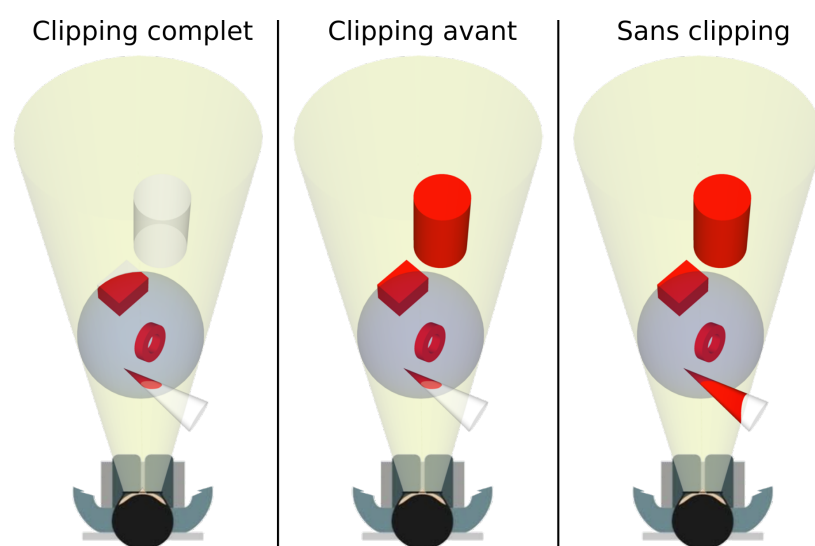


FIGURE 6.2: Couper la scène. L'APCD est représenté en gris clair, le cône de vue en jaune clair. Les parties visibles des objets de la scène sont rouges, les parties clippées sont blanches. À gauche : **clipping complet**. Au milieu : **clipping avant**. À droite : **sans clipping**.

réoscopique couplé à la position de l'utilisateur, ce dernier dispose d'indices de profondeur suffisants pour comprendre les positions relatives des objets virtuels et de l'appareil, de sorte que le clipping peut être évité. Dans une étude pilote, nous avons constaté que la métaphore de la fenêtre ouverte était assez intuitive pour ne pas imposer une perte gênante de présence par rapport aux autres façons de découper la scène virtuelle.

Une prise à deux mains pour un meilleur contrôle des rotations

Tenir une sphère à deux mains pour la manipuler présente plusieurs avantages par rapport à une prise à une main. Une prise à une main permet de manipuler confortablement une sphère de la taille d'une balle de Handball, tandis qu'une prise à deux mains permet de manipuler confortablement une sphère plus large : de la taille d'un ballon de basketball. Le fait d'avoir une sphère plus large agrandit le champ de vision de l'utilisateur. À une main, les rotations sont limitées aux contraintes articulaires du poignet et de l'avant-bras, et certains mouvements génèrent beaucoup d'occlusion. Deux mains peuvent coopérer pour effectuer des rotations amples en utilisant diverses formes d'embrayage : une main peut être sous l'appareil et le maintenir tandis que l'autre main se repositionne. Les doigts peuvent *se déplacer* sur la surface de l'appareil pour se repositionner sans lâcher l'écran. La fatigue musculaire causée par l'effet de bras de gorille [69] est moindre lorsque les deux bras partagent l'effort de maintien de l'appareil en l'air. Les rotations à deux mains ont l'avantage supplémentaire de faciliter les rotations *pures*, c'est-à-dire les rotations qui font pivoter l'appareil sans déplacer son centre. Les rotations pures sont utiles pour modifier un paramètre, par exemple un facteur de zoom, sans déplacer l'appareil, afin de ne changer ni le point de vue sur la scène, ni le champ de vision. En effet, le champ de vision dépend de la position de l'appareil par rapport à la scène virtuelle mais pas de son orientation.

En contrepartie, utiliser une sphère plus large et donc forcément plus lourde peut avoir un impact important sur la fatigue musculaire lors de la manipulation. Adopter une approche par projection permet de limiter cet impact car des volumes passifs légers peuvent être utilisés à la place d'appareils autonomes avec écrans actifs.

Il convient de noter que toutes les rotations n'offrent pas la même gamme d'orientations accessibles par des mouvements confortables. Nous avons constaté que la *rotation de tangage* (voir figure 6.3) est plus facile à effectuer à partir de la position de repos avec

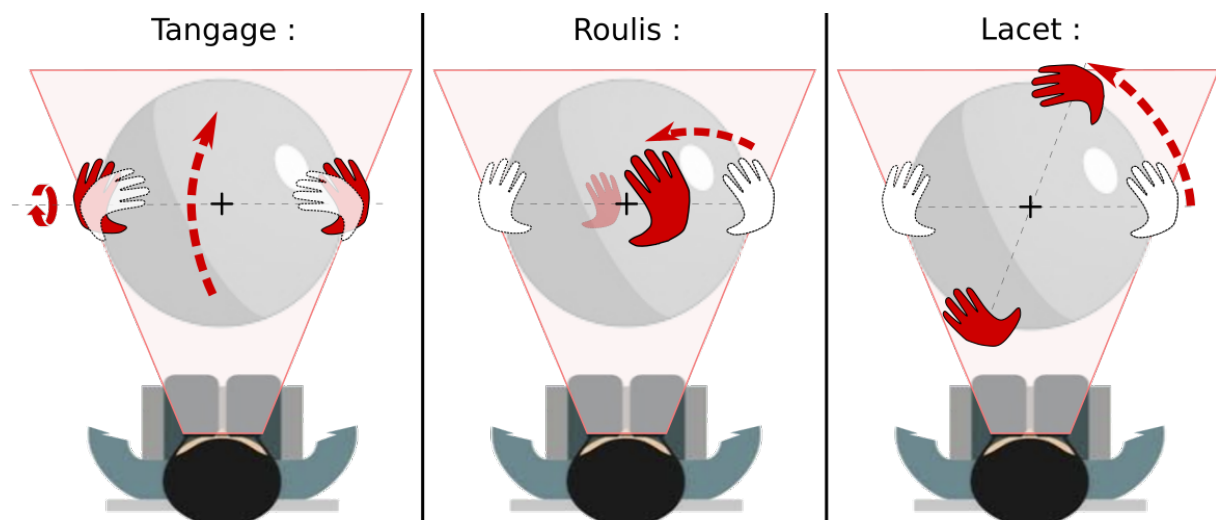


FIGURE 6.3: Faire tourner l'APCD, vu de *dessus*. Positions initiales des mains en blanc, positions finales en rouge.

les mains de chaque côté de la sphère. Depuis cette position, nous avons observé que trois doigts ou plus des deux mains peuvent coopérer pour maintenir la pression sur l'axe de tangage et faire tourner l'appareil constamment, même au-delà de 360°. Cela fait du tangage un candidat privilégié pour le contrôle de paramètres. Le roulis demande les mêmes mouvements que pour tourner un volant. Cette rotation est confortable mais ne permet qu'une gamme d'orientations confortables de 180°. Le lacet offre la même gamme d'orientations que le roulis mais il engendre plus de fatigue musculaire lors de mouvements larges et il provoque de l'occlusion lorsqu'une main est amenée devant la sphère, forçant un changement de prise en main. Les mouvements des mains nécessaires pour effectuer les rotations de tangage, de roulis et de lacet sont illustrés dans la figure 6.3.

Les variations d'orientation de lacet, de tangage et de roulis peuvent être mesurées indépendamment, permettant le contrôle intégral de trois paramètres indépendants. Nous avons testé le contrôle intégral de la largeur et du centre d'une plage de valeurs sur une échelle linéaire. L'interaction résultante s'est avérée difficile à contrôler finement. Jacob et al. suggèrent qu'un contrôle intégral ne peut être utilisé efficacement que lorsque les paramètres contrôlés sont perçus comme intégraux (comme par exemple les trois degrés de liberté d'un angle définis dans l'espace). [75]. Nous détaillons le contrôle intégral de l'orientation d'un rayon dans la section 6.2.5.

6.1.4 Compromis présence-efficacité

Comme nous l'avons vu dans le chapitre 5, les APCD peuvent générer un fort sentiment de présence des objets virtuels qui semblent être contenus dans l'appareil. Cela peut être exploité pour permettre aux utilisateurs d'anticiper le comportement d'une interface. Par exemple, lorsque les utilisateurs *ressentent* qu'un cylindre virtuel est fixé rigidement le long d'un rayon de la sphère physique qu'ils ont en main, il est intuitif pour eux d'orienter le cylindre en tournant l'appareil en conséquence. Cette approche est qualifiée par Jacob et al. d'interaction basée sur la réalité [76]. Ils notent que cette présence peut accélérer l'apprentissage du système pour les utilisateurs novices, et peut également améliorer les performances des experts dans des situations de surcharge d'informations. Cependant, les auteurs notent que "une grande partie de la puissance de l'utilisation des ordinateurs vient de [...] la capacité d'aller au-delà d'une imitation précise du monde réel". Par exemple, avoir un raccourci gestuel pour voir immédiatement le dos d'un objet virtuel est notablement plus efficace, bien que moins intuitif et limitant la présence, que de tourner l'affichage pour voir le dos de l'objet.

Au cours de notre conception, ce rapport présence/efficacité est rapidement apparu comme un compromis majeur. Dans ce qui suit, nous y ferons souvent référence et présenterons des conceptions alternatives qui tendent plus ou moins vers un côté ou l'autre de ce compromis. Dans l'application d'apprentissage de l'anatomie que nous avons créée pour l'étude qualitative, nous avons délibérément choisi de privilégier la présence à l'efficacité : les participants devaient être novices dans l'interaction avec un APCD et n'être amené à utiliser l'application qu'une seule fois, d'où l'accent mis sur l'intuitivité.

6.2 Conception d'une panoplie interactive pour APCD

Dans cette section, nous présentons une panoplie permettant de répondre à toutes les facettes de l'interaction avec des environnements virtuels. Nous traitons cinq aspects : la sélection, la manipulation d'objets, le zoom, les conteneurs offrant une interface pour widgets, et les widgets permettant différents contrôles de paramètres. Nous rendons également compte des observations recueillies en observant des participants utilisant ces interactions lors de premiers tests pilotes. Ces tests pilotes n'incluent pas l'étude qualitative, plus aboutie, qui sera présentée dans la section suivante.

6.2.1 Sélection d'objets

Effectuer une sélection dans un espace 3D nécessite deux éléments : 1) de pointer un objet avec un curseur, et 2) d'activer une commande. La position du curseur spécifie quel objet peut être sélectionné et l'activation spécifie quand la sélection a lieu. Nous avons testé les deux principales approches de pointage d'objets 3D : le pointage direct et une forme particulière de lancer de rayons, le picking.

Curseur 3D

Le curseur est ancré au centre de la sphère avec les deux approches. De cette façon, il est facile à déplacer vers n'importe quelle position accessible de l'espace en manipulant la sphère. Le centre de la sphère est au centre du focus visuel de l'utilisateur et son environnement proche est par conséquent toujours affiché. De plus, une rotation de la sphère n'induit pas de déplacement du curseur. Dans une sphère de 30 cm de diamètre, le curseur est représenté par une croix 3D de 2 cm d'envergure constituée de 3 cylindres de 1 mm de diamètre. L'épaisseur de la croix lui confère un volume qui contribue à sa présence.

Activation

L'APCD devant être tenu en permanence avec les deux mains, les utilisateurs ne peuvent par conséquent effectuer aucune action à main levée, ni utiliser aucun contrôleur manuel. Les APCD étant des écrans volumiques, ils peuvent être orientés dans tous les sens afin d'explorer un objet par exemple. Ceci est une différence majeure avec les appareils de RA mobile, qui ne fournissent qu'une seule surface plane d'affichage et ne peuvent donc pas être retournés. Par conséquent, des boutons physiques ne peuvent pas être utilisés pour activer des commandes, car l'utilisateur ne saurait plus où ils se trouvent au fur et à mesure des orientations de l'appareil. Nous avons étudié deux possibilités d'activation : en tapant avec un pied ou en touchant la surface de l'appareil avec un doigt dédié.

Nous avons testé l'activation au pied soit en appuyant sur des pédales physiques posées au sol, soit en traquant la position du pied avec un marqueur optique. L'activation au pied est largement utilisée dans divers domaines professionnels, et elle nous a été suggérée par deux pratiquants médicaux qui ont testé les premiers prototypes du système. Il s'agit par ailleurs de la technique que nous avons mis en place pour les études sur l'apprentissage d'objets 3D et sur la présence, que nous avons présentées dans les chapitres 4 et 5. Cette solution est viable dans un contexte d'utilisation où l'utilisateur est assis, ses pieds restant à peu près au même endroit durant la session d'utilisation. En revanche, elle semble invisageable pour une utilisation plus mobile qui pourrait être possible avec les futurs prototypes d'APCD. En position assise, taper avec les pieds n'engendre pas de

mouvements parasites dans le haut du corps et donc dans la position de l'appareil.

Tang et al. [148] ont développé un APCD cubique construit avec 6 écrans OLED tactiles, permettant d'interagir avec le contenu virtuel en touchant les surfaces du cube avec les doigts. Avec une sphère tactile, les événements de détection du toucher peuvent être déclenchés indépendamment de l'orientation de la sphère en touchant sa surface en n'importe quel point. Cependant, cela nécessite de consacrer au moins un doigt à l'interaction tactile quelle que soit la prise de l'utilisateur sur l'appareil. Nous avons observé que les pouces jouent un rôle essentiel dans la manipulation de l'appareil. Ainsi, aucune interaction tactile ne doit être associée aux mouvements des pouces. Les quatre doigts jouent un même rôle redondant : les utilisateurs ont tendance à tenir la sphère avec tous les doigts, mais nous avons observé qu'ils pouvaient facilement libérer un doigt (c'est-à-dire le garder en l'air, sans contact avec la surface de la sphère) sans altérer la manipulation. Nous avons observé que taper sur la sphère avec l'index ne générerait pas de mouvement parasite, grâce à la stabilité de la prise à deux mains.

Nous avons réussi à transformer notre prototype en une surface tactile en ajoutant un marqueur de capture de mouvement sur l'index de la main dominante de l'utilisateur, comme illustré dans la figure 6.1. De cette façon, aucun autre doigt ne peut déclencher un événement tactile involontaire lors de la manipulation de la sphère. En calculant la distance entre l'index et la surface de la sphère en temps réel, il est facile de simuler avec précision une surface sphérique tactile. Les futurs APCD pourraient détecter les événements tactiles sans technologie externe. En effet, Benko et al. ont déjà présenté un écran sphérique multi-touch entièrement fonctionnel [10]. En utilisant ce genre de technologie, il faut néanmoins être capable de différencier les événements tactiles déclenchés avec l'index pour activer des commandes, des événements déclenchés par les autres doigts lors de la manipulation de l'écran, et qui ne doivent pas être considérés comme des activations.

Nous avons implémenté l'activation au doigt dans l'application d'apprentissage de l'anatomie que nous avons utilisée pour l'étude qualitative, car nous avons observé qu'elle permet des fréquences d'activation plus élevées et parce qu'effectuer un tap avec le doigt nécessite moins d'énergie qu'avec le pied. Plusieurs modes d'activation, comparables aux boutons gauche et droit d'une souris d'ordinateur, peuvent être obtenus en utilisant à la fois le pied et le doigt comme mécanismes d'activation. Les futurs APCD pourront peut-être détecter le doigt qui effectue un tap et ainsi faire correspondre différents modes d'activation pour chaque doigt.

Pointage direct

Avec le pointage direct, les utilisateurs pointent un objet en déplaçant l'appareil jusqu'à ce que le curseur pénètre dans le volume de l'objet ciblé. Avec les casques de RV, le pointage direct est souvent considéré comme la technique de pointage la plus intuitive. Il permet également de bénéficier de la proprioception des utilisateurs : lorsqu'ils ont une bonne perception de la position des objets autour d'eux, leur proprioception leur permet d'améliorer considérablement leurs performances de pointage direct [104]. Cependant, l'acquisition de ces connaissances spatiales est entravée par un problème connu de sous-estimation de distances égocentriques en RV [97, 122]. Ce problème peut être atténué en utilisant un appareil stéréoscopique, et en affichant une réplique virtuelle de la salle physique dans laquelle l'utilisateur se trouve et un avatar de l'utilisateur [122]. Dans le cas des APCD, la stéréoscopie est accessible, et une perception idéale de la pièce et du corps de l'utilisateur est disponible naturellement. En effet, nous avons mentionné dans le chapitre 3 que la perception des informations contextuelles offertes par la réalité physique

peut avoir été un facteur important expliquant les performances élevées du APCD dans une tâche de placement. Tout cela indique que le pointage direct peut être particulièrement adapté pour interaction avec un APCD.

Lors de tests pilotes, nous avons observé que la plupart des utilisateurs parviennent à sélectionner des objets *efficacement* et *sans effort* grâce au pointage direct. Cependant, certains utilisateurs ont régulièrement besoin d'effectuer des gestes correctifs pour atteindre la profondeur de leur cible. Ces utilisateurs ont souvent une correction visuelle importante et/ou de faibles capacités de perception stéréoscopique (avec des scores inférieurs à 6/12 au test RANDOT), et expérimentent probablement le même problème de perception de profondeur dans la réalité physique. Une autre explication peut être que ces utilisateurs ont tendance à pointer à distance instinctivement au lieu d'effectuer un pointage direct. Cela pourrait venir d'un transfert de leurs attentes provenant d'une utilisation quotidienne d'ordinateur de bureau, où le pointeur reste toujours au niveau du plan de l'écran même lorsque l'utilisateur pointe des objets éloignés dans une scène 3D par exemple.

Quoi qu'il en soit, pour aider ces corrections, nous avons fourni des retours visuels clairs qui aident l'utilisateur à comprendre et à prédire ses possibilités interactives : les objets virtuels sont mis en évidence avec un contour jaune lorsqu'ils contiennent le curseur, comme illustré dans la figure 6.4, à droite. De cette façon, l'utilisateur est toujours au courant qu'un objet sera sélectionné s'il active la sélection. Pour indiquer qu'une activation a été détectée mais qu'aucun objet n'était pointé, le curseur clignote en rouge. Lorsqu'un objet est sélectionné, le contour prend une couleur verte ; de cette façon chaque action entraîne une rétroaction immédiate.

Cette technique est efficace et intuitive pour les objets à portée de main mais elle ne permet pas d'atteindre des objets éloignés, contrairement au picking.

Picking

Nous avons testé l'utilisation d'une technique de lancer de rayons pour sélectionner des objets qui sont hors de portée des bras de l'utilisateur. Une interaction de lancer de rayon classique met en jeu un périphérique d'entrée manipulable à une main qui définit l'origine et l'orientation 6D du rayon [61, 170]. Dans le contexte que nous étudions, l'utilisateur ne peut pas libérer ses mains pour pointer l'écran, nous avons donc implémenté un pointage par *picking*. Le picking est la forme dominante de pointage 2D sur les ordinateurs de bureau et les appareils mobiles. Avec cette technique, le rayon est défini par l'œil de l'utilisateur et le curseur, comme illustré dans la figure 6.4 à gauche. Comme pour toute forme de lancer de rayon, il pointe vers l'objet le plus proche qui intersecte le rayon depuis son origine. Cette technique permet de sélectionner des objets sans déplacer le curseur en profondeur et donc avec de moins grands mouvements que le pointage direct, résultant en des sélections plus performantes.

Dans les implémentations conventionnelles du picking, le curseur est toujours devant la scène, et pointe donc toujours vers des objets situés derrière lui. Dans le cas des APCD, nous proposons que le pointeur soit *à l'intérieur* de la scène virtuelle. Cela implique que le picking ait lieu *vers l'œil* lorsque l'objet le plus proche du curseur se trouve devant lui dans la scène. Bien que non conventionnel, nous avons observé que ce comportement était bien compris par les utilisateurs. Cela a un effet secondaire intéressant : lorsque le curseur se trouve à l'intérieur d'un objet, les techniques de picking et de pointage direct pointent toutes deux vers l'objet. En d'autres termes, le pointage direct n'est qu'un cas

particulier du picking où le pointeur entre en collision avec un objet. En conséquence, les deux techniques peuvent être proposées en parallèle sans avoir besoin d'une activation pour changer de mode. Pour aller plus loin dans la cohérence, nous affichons le même contour jaune que celui qui est utilisé dans le cas du pointage direct pour indiquer qu'un objet est sélectionnable par picking. Nous avons pu observer lors de tests pilotes que les participants effectuaient des transitions très fluides entre des sélections par picking ou par pointage direct.

Néanmoins, le picking présente certains inconvénients. En étendant la portée du curseur en dehors de sa limite visible, cette technique brise l'illusion de la présence physique du curseur. De plus, la manipulation d'objets qui ont été sélectionnés en dehors des limites de la sphère est difficile et nécessite la mise en œuvre d'une technique de manipulation avancée comme nous le détaillerons dans la section 6.2.2. Ainsi, le pointage direct semble plus pertinent à utiliser lorsque tous les objets sélectionnables restent dans une zone facilement accessible, ou si les performances de sélection sont moins importantes que le sentiment de présence du curseur.

Parallaxe de stéréoscopie

Le picking peut être ambigu lorsqu'il est utilisé avec un affichage stéréoscopique, car les deux yeux définissent chacun un rayon différent passant par le curseur et pouvant pointer vers des objets différents. Pour illustrer cette ambiguïté, imaginez que vous regardez un objet lointain et que vous essayez de pointer dans sa direction avec votre doigt. Vous verrez alors deux images séparées de votre doigt simultanément car vos yeux convergent loin de lui. Vous pouvez alors aligner l'objet lointain avec l'image de votre doigt vue par votre œil droit ou avec celle vue par votre œil gauche, ce qui entraîne une ambiguïté de sélection. Il en va de même lorsque vous essayez de sélectionner un objet éloigné avec le curseur 3D. La figure 6.4, à droite, montre cette parallaxe de stéréoscopie.

Valkov et al. [152] ont étudié ce problème dans un cas de rendu stéréoscopique sur une surface tactile. Ils ont constaté que les utilisateurs ont tendance à pointer entre les deux projections stéréoscopiques de l'objet mais plus près de la projection de l'œil dominant.

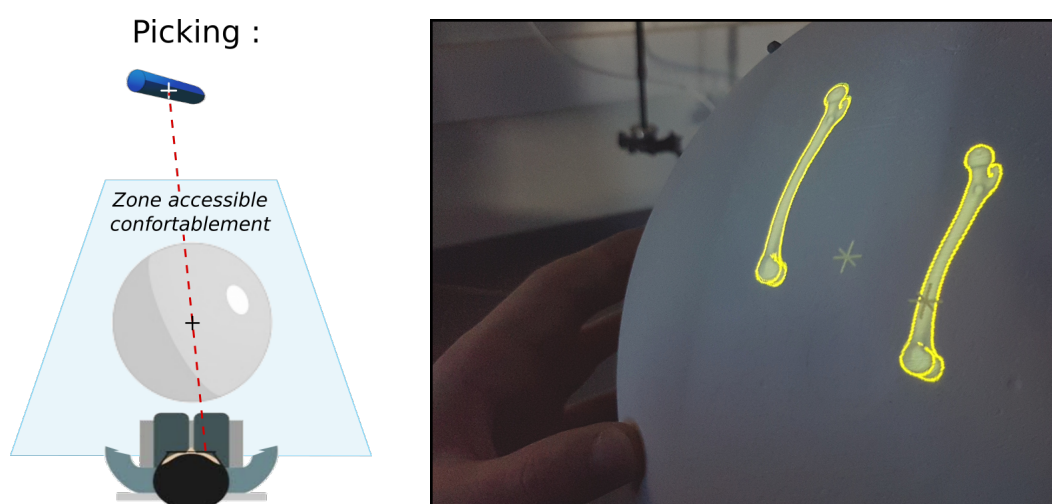


FIGURE 6.4: À Gauche : utilisation du picking pour sélectionner un objet en dehors de la zone accessible confortablement. À droite : l'utilisateur pointe vers une cible éloignée (l'os) avec le curseur. Les images du curseur et de l'os vues par son œil dominant sont alignées (à droite), mais pas les images vues par son œil non dominant (à gauche). L'objet sélectionnable est mis en évidence par un contour jaune.

Compte tenu de ce comportement, il est difficile de fournir un algorithme automatique de désambiguïsation. De plus, les utilisateurs peuvent volontairement lever l’ambiguïté de la sélection en fermant leur œil non dominant (une pratique courante dans le tir sportif). Nous avons implémenté ce comportement en calculant à tout moment les rayons de picking issus des deux yeux et en priorisant l’œil dominant en cas d’ambiguïté de sélection. Cette solution permet aux utilisateurs de sélectionner un objet éloigné quel que soit leur œil dominant, sauf dans le cas où le curseur est aligné avec un objet différent pour chaque œil. De cette façon, le système n’a pas besoin d’être configuré pour chaque utilisateur tant que la scène ne contient pas beaucoup d’objets qui se chevauchent. En effet, dans la majorité des cas de sélection, soit les deux rayons pointent vers le même objet, soit l’un des rayons pointe dans le vide. Une ambiguïté se produit uniquement si chaque rayon pointe vers un objet différent, dans le cas de scènes surchargées. Nous n’avons pas testé de telles scènes, et avons arbitrairement choisi de prioriser le rayon issu de l’œil droit pour désambiguïser la sélection si le cas se présentait avec notre application test. Des scènes denses avec de nombreuses cibles peuvent quant à elles générer de nombreuses ambiguïtés. Dans ce cas, il est approprié de configurer le système afin qu’il priorise l’œil dominant de l’utilisateur courant.

6.2.2 Manipulation d’objets

L’APCD fournit naturellement un contrôle 6D isomorphe pour des objets ancrés au centre de la sphère. Cette manipulation intuitive reproduit fidèlement le comportement qu’aurait un objet physique ancré à l’intérieur de la sphère.

Un glisser-déposer 6D

Dans les interfaces 2D conventionnelles, le glisser-déposer 2D est une interaction de manipulation courante. Dans l’espace 3D, il est possible de mettre en place une interaction de glisser-déposer 6D pour placer des objets : une fois qu’un objet a été sélectionné avec une interaction de sélection présentée ci-dessus, et alors que l’index de l’utilisateur reste en contact avec la surface de la sphère, l’objet est ancré dans le référentiel de la sphère. Ainsi, chaque mouvement de la sphère sera reproduit de manière isomorphe sur l’objet. Dès que l’utilisateur relève son index, l’objet est désélectionné et désolidarisé du référentiel de la sphère.

Manipuler des objets lointains

Lorsqu’un objet est sélectionné en dehors des limites de la sphère à l’aide de la sélection par picking, la rotation de la sphère provoque des déplacements importants qui peuvent amener l’objet manipulé à sortir de la zone visible, en raison d’un effet de bras de levier dû à l’éloignement du pivot de rotation et de l’objet. Cet effet de bras de levier illustré dans la figure 6.5 à gauche, rend l’orientation de l’objet sélectionné difficile et contre intuitive. Ce problème a déjà été identifié dans un contexte de RA mobile, comme illustré par Issartel et al. [74]. Nous avons testé deux techniques pour résoudre ce problème, soit en déportant le point de pivot de rotation (**curseur proxy**) soit en déplaçant l’objet lui-même (**curseur magnétique**).

Le **curseur proxy** est inspiré de la technique HOMER [20]. Après la sélection, l’objet est manipulé comme si un proxy du pointeur avait été créé au niveau du point de sélection : les rotations de l’appareil sont reproduites sur l’objet mais en utilisant le point de sélection comme pivot, comme l’illustre la figure 6.5, au milieu. Les translations de l’appareil sont également appliquées isomorphiquement à l’objet, mais avec une amplitude

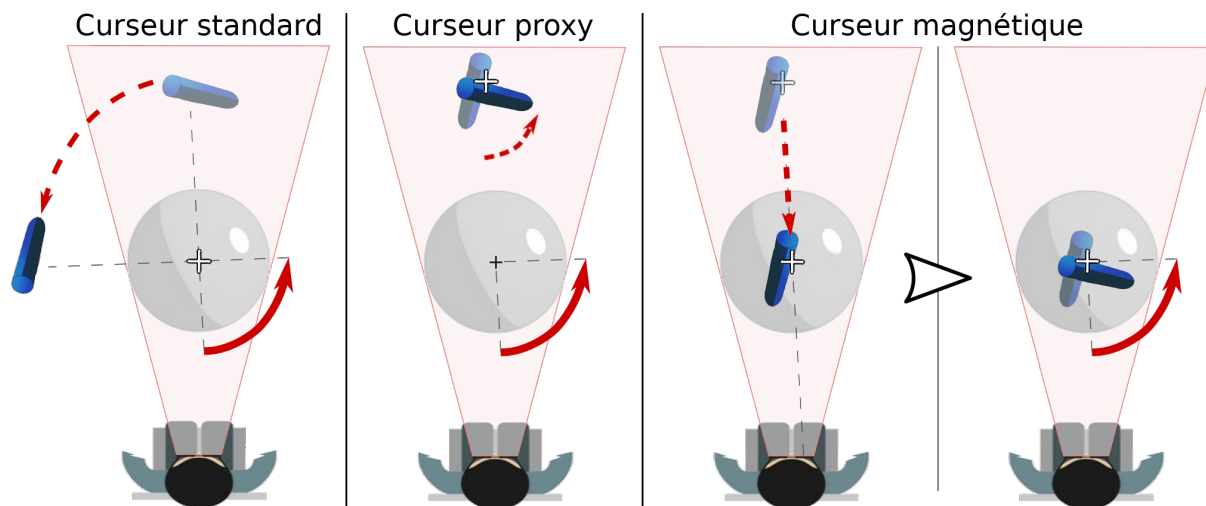


FIGURE 6.5: Manipulation d'un objet après une sélection par picking. À gauche : avec un curseur standard, l'objet tourne autour du centre de la sphère. Au milieu : avec le curseur proxy, l'objet tourne autour du point de sélection. À droite : avec le curseur magnétique, l'objet est attiré vers le curseur.

qui semble être réduite car l'objet est plus éloigné de l'utilisateur que l'APCD.

Pour des manipulations précises, il peut être plus approprié de rapprocher l'objet de l'utilisateur. Nous avons implémenté un **curseur magnétique** qui fonctionne comme suit : après sélection, nous animons rapidement (500 ms) une translation automatique de l'objet jusqu'à ce que le point de sélection coïncide avec le curseur, comme si l'objet était attiré par un aimant. La translation est continue pour éviter une discontinuité perturbante. Après ce déplacement, les utilisateurs manipulent l'objet de la même manière qu'après un pointage direct. Lorsque l'utilisateur termine la manipulation, l'objet reste en place (c'est-à-dire qu'il ne revient pas à sa position initiale). Le curseur magnétique est illustré sur la figure 6.5, à droite.

Ces deux techniques permettent de manipuler des objets lointains sans les perdre de vue, mais au prix d'une présence réduite. Le curseur proxy rompt la métaphore de manipulation naturelle du APCD en ajoutant une indirection dans la fonction de transfert du contrôle de la rotation. Le curseur magnétique, quant à lui, effectue automatiquement une translation qui ne correspond pas à une manipulation explicite de l'utilisateur. Dans les deux cas, le curseur physique agit sur un objet distant sans le toucher physiquement. La translation automatique effectuée par le curseur magnétique ne peut pas être inversée sans déplacer manuellement l'objet jusqu'à sa profondeur d'origine. Mais cet inconvénient pourrait ne pas être aussi gênant qu'il y paraît. En effet, une tâche nécessitant la manipulation d'un objet exige généralement une grande attention visuelle, et donc d'amener l'objet à manipuler au plus près de l'utilisateur. Pour examiner un objet éloigné, il serait nécessaire d'effectuer cette translation manuellement avec le curseur proxy, et cela pourrait nécessiter plusieurs débrayages. Nous avons utilisé le curseur magnétique dans notre application de test, car il permet d'amener rapidement des objets distants à l'intérieur de l'APCD.

6.2.3 Zoom

Une opération de changement d'échelle est définie par un centre 3D et un facteur de zoom 3D. Le centre est le seul point de la scène qui ne sera pas déplacé par l'opération de

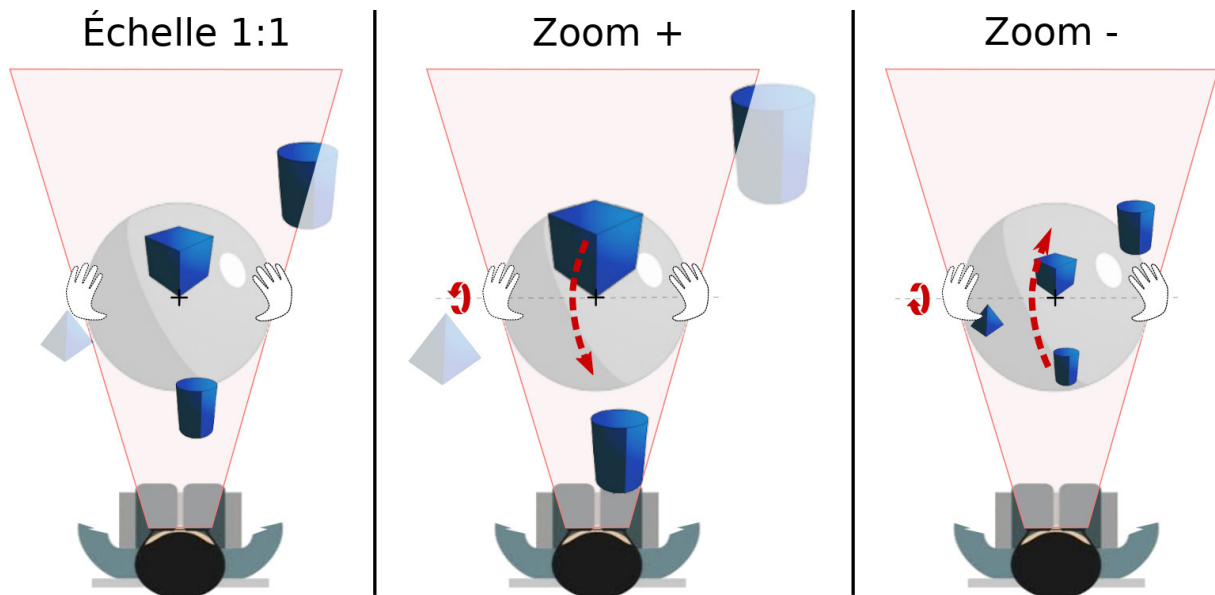


FIGURE 6.6: Rotations de la sphère effectuant des changement d'échelle de la scène centrées autour du centre de la sphère.

changement d'échelle. Le but principal d'un changement d'échelle dans les applications de RV est de zoomer sur la scène afin de se concentrer sur de petits détails ou de faire un zoom arrière pour avoir un aperçu de toute la scène. Habituellement, les utilisateurs ne cherchent pas à effectuer un changement d'échelle non uniforme qui déformerait les objets, mais préfèrent appliquer un zoom uniforme de la scène. Ainsi, le zoom peut être considéré comme une tâche à 1 (facteur) + 3 (position du centre) degrés de liberté.

Il semble naturel de manipuler le centre du zoom à l'aide d'un contrôle isomorphe en le maintenant au centre de la sphère. Ce choix de conception minimise également les déplacements des objets les plus proches du focus visuel de l'utilisateur lors d'un changement d'échelle de la scène. Le facteur de zoom peut être modifié simultanément s'il est lié à une rotation 1D de la sphère. Nous avons choisi de lier le facteur de zoom à la rotation de tangage, que nous avons identifiée comme la plus facile à effectuer. De plus, une graduation logarithmique devrait être utilisée au lieu d'une graduation linéaire pour homogénéiser les effets visuels induits par un zoom entre des petits et des grands facteurs. En effet, le même contrôle devrait être effectué pour passer d'un facteur de zoom de 0,1 à 0,2 comme pour passer d'un facteur de 10,0 à 20,0. Ainsi, une rotation continue de la sphère vers l'arrière réduira l'échelle de la scène de manière continue, tandis qu'une rotation vers l'avant augmentera l'échelle de la scène de manière continue, comme illustré dans la figure 6.6.

Une opération de zoom doit être initialisée par l'activation d'une commande, et doit également être finalisée par une activation (comme un tap avec l'index ou le pied). Pendant toute la durée de l'opération, le centre du zoom est continuellement mis à jour en fonction de la position du curseur, tandis que l'utilisateur ajuste le facteur de zoom avec une rotation de tangage. Dans des tests pilotes, nous avons observé que l'interaction était très intuitive et permettait aux participants d'affiner continuellement le centre de l'opération de zoom alors que la scène était agrandie et révélait plus de détails. Bien que la mise à l'échelle des objets n'ait pas été utilisée dans notre application de test, la même technique d'interaction pourrait être utilisée : l'activation du zoom permettrait de changer l'échelle de l'objet que l'on pointe avec le curseur, ou de changer l'échelle de la scène si le curseur pointe dans le vide.

6.2.4 Conteneurs de Widgets

Nous nous concentrons maintenant sur la conception de panneaux de contrôles, menus, boutons et sliders, et sur les conteneurs permettant l’affichage et les interactions avec ces widgets. Ce domaine reste largement inexploré dans le domaine des APCD. Comme pour la plupart des systèmes HF actuels, nous avons conçu des widgets 2D. Ces widgets sont présentés dans deux types de conteneurs planaires.

Des conteneurs planaires

Mine suggère que les contrôles qui ne nécessitent pas explicitement une manipulation 3D soient intégrés en tant que widgets 2D dans des panneaux de contrôle 2D virtuels, afin d’éviter d’ajouter une complexité inutile lors de l’interaction avec les widgets [103]. De nombreuses études ont montré les avantages des panneaux de contrôle 2D et des boîtes à outils dans des scénarios de RV immersifs, comme l’ont examiné Tu et al. [151]. Des panneaux de contrôle 2D ont également été mis en œuvre avec des casques de RA tels que le Microsoft HoloLens. Dans la plupart des cas, les utilisateurs interagissent avec les panneaux de contrôle 2D en pointant les widgets avec une technique de lancer de rayon, afin de minimiser les demandes de mouvement pour la sélection d’éléments.

Les travaux précédents sur les écrans sphériques utilisaient la surface de l’écran comme conteneur de widgets [10, 105]. Les widgets en surface peuvent ainsi être touchés directement, augmentant la cohérence de l’interaction tactile pour activer des commandes. Dans le cas des APCD *stéréoscopiques*, nous avons la possibilité d’afficher le contenu à l’intérieur de l’écran, et donc d’utiliser un conteneur plat qui soit toujours en face de l’utilisateur, à la place d’un conteneur hémisphérique où les widgets les plus éloignés se tournent le dos, nécessitent des mouvements de l’utilisateur ou de l’appareil pour être vus de face, et sont difficilement perceptibles en même temps. De plus, la position des mains de l’utilisateur étant contrainte par les besoins de manipulation de l’appareil, il n’est pas pertinent d’utiliser des widgets tactiles qui demanderaient un changement de prise en main de l’APCD pour être activés. Afficher les widgets à l’intérieur du dispositif fait comprendre instinctivement à l’utilisateur qu’il ne peut pas les toucher pour les activer. Nous avons donc conçu des conteneurs virtuels plats évoluant à l’intérieur de l’APCD. Nous avons testé l’utilisation de conteneurs placés à des emplacements fixes à côté de la zone de travail 3D, de la même manière que les boîtes à outils restent sur le côté des documents dans la majorité des interfaces d’ordinateurs de bureau. Cependant, bien que cette approche soit pertinente pour les casques de RV et de RA, nous l’avons trouvée inappropriée pour les APCD en raison du champ de vision notablement plus petit : l’utilisateur a régulièrement besoin de transférer son attention entre le panneau de contrôle et l’espace de travail, et ce transfert s’avère particulièrement coûteux lorsque l’un des deux espaces est hors de vue. Nous nous sommes donc concentrés sur des conteneurs contextuels qui apparaissent dynamiquement en face de l’utilisateur dans l’APCD.

Afficher les conteneurs contextuels

Pour afficher et masquer un conteneur contextuel, nous proposons d’utiliser un événement de double tap. Un double tap n’ajoute pas beaucoup de demandes motrices par rapport à un simple tap et évite des conflits avec l’interaction de sélection présentée ci-dessus. Il est également possible d’activer la commande d’appel du conteneur contextuel par un tap avec un autre doigt (ou l’autre pied), mais multiplier les parties du corps qui peuvent exécuter des actions engendre une surcharge cognitive ; et dans le cas des doigts, cela augmente le nombre de doigts indisponibles pour manipuler la sphère.

L'utilisation de transitions animées pour améliorer la compréhension des comportements d'une interface est devenue une pratique courante pour les concepteurs d'interface utilisateur. Dragicevic et al. ont démontré que les transitions Slow-in/Slow-out (SI/SO) aident les utilisateurs à anticiper le début et la fin d'une animation en allouant plus d'images aux stades précoce et tardif [48]. Nous avons implémenté des animations SI/SO rapide (0,3 seconde) pendant les transitions d'apparition et de disparition du conteneur contextuel. Lors de son apparition, le conteneur surgit du centre de la sphère (c'est-à-dire de la position du curseur de sélection lorsqu'il est affiché) avec une taille initialement nulle, et bouge jusqu'à atteindre sa profondeur et sa taille finales. L'animation inverse est utilisée lorsque le conteneur contextuel disparaît.

Interagir avec les conteneurs

Les APCD fournissent des contrôles précis des translations et des rotations 3D, conduisant à deux méthodes potentielles pour gérer la manipulation d'un pointeur dans un conteneur 2D.

L'approche la plus directe consiste à ancrer un pointeur au centre de la sphère afin qu'il puisse être manipulé intuitivement en déplaçant la sphère dans l'espace, c'est-à-dire un peu comme le curseur 3D de l'interaction de sélection. Si le conteneur reste fixe dans la scène 3D, la translation de la sphère déplacera le pointeur dans le référentiel du conteneur afin de sélectionner des commandes par picking. Avec cette approche, l'utilisateur peut interagir avec des widgets 2D et sélectionner des objets 3D avec la même interaction de sélection cohérente, améliorant l'intuitivité et la compréhension de l'interface. Cependant, cela pourrait conduire à des ambiguïtés dans le cas de widgets superposés à des objets sélectionnables. Avec des conteneurs contextuels, cette ambiguïté peut être résolue en désactivant la sélection d'objets pendant que le conteneur contextuel est affiché.

Néanmoins, cette approche a ses limites pour interagir avec un panneau de contrôle complexe. En effet, les écrans sphériques volumiques actuels ont une qualité d'affichage assez limitée : de 77 dpi pour Grossman et al. [61] à 90 dpi pour notre prototype. En plus de leur petit champ de vision (pas plus de 30 cm de diamètre autour du centre de focus visuel), fournir un panneau de contrôle lisible avec des cibles suffisamment grandes pour être facilement sélectionnées peut demander des panneaux de contrôle aussi larges que le champ de vision. Certaines parties du panneau de contrôle sortiraient alors de la zone affichée lorsque la sphère est translattée pour déplacer le pointeur vers la partie opposée du panneau, comme illustré dans la figure 6.7.

Une deuxième approche pour interagir avec un conteneur contextuel consiste à ancrer le conteneur lui-même au centre de la sphère et à utiliser les rotations de la sphère pour manipuler un pointeur dans le référentiel du conteneur. De cette façon, le conteneur 2D flottera dans la scène 3D lorsque la sphère sera déplacée et restera toujours dans la zone de focus visuel de l'utilisateur (c'est-à-dire à l'intérieur de la sphère), maximisant l'espace qui peut lui être alloué tout en restant toujours visible. L'orientation du conteneur, cependant, doit toujours rester stable dans le référentiel de l'utilisateur, de sorte que le conteneur reste vertical et lisible à tout moment. Avec un tel conteneur flottant, les translations de la sphère peuvent être utilisées pour naviguer dans la scène, pendant que ses rotations sont utilisées pour déplacer le pointeur à l'intérieur du conteneur, sans surcharger la modalité de translation. Les utilisateurs devraient pouvoir effectuer des rotations de manière intuitive comme l'ont observé Miyafuji et al. : en évaluant la manière dont les utilisateurs interagissent avec leur écran sphérique portable, ils ont noté que les utilisateurs effectuaient naturellement des rotations sans explication préalable [105].

Nous avons testé deux types de conteneurs contextuels : un marking menu et un panneau de contrôle, qui utilisent respectivement les translations et les rotations de l'APCD pour manipuler le pointeur 2D.

6.2.5 Widgets

Un marking menu qui utilise les translations de l'appareil

Nous avons créé un marking menu comme un moyen d'offrir une sélection de commandes efficace parmi un large ensemble de commandes. Suite aux recommandations des recherches récentes, nous affichons le menu immédiatement après son invocation [67]. Les conteneurs étant plans, il serait inefficace d'utiliser le pointeur 3D pour viser des cibles 2D. Nous avons réduit la difficulté de la tâche de pointage en définissant un pointeur 2D dans le menu, comme l'intersection du plan du menu avec l'axe œil-curseur (en utilisant l'œil droit). Le pointeur 2D est représenté par une petite sphère rouge sur le plan du menu et laisse une marque fine (une condition des marking menus). Ce pointeur 2D est contrôlé par les déplacements 2D de l'APCD dans un plan parallèle au menu, comme l'illustre la figure 6.7. Le menu disparaît dès que le pointeur 2D entre dans une feuille de la hiérarchie, ou s'il est annulé en tapant sur l'affichage avec l'index avant de sélectionner une commande.

Nous avons testé un menu à deux niveaux hiérarchiques. Nous avons observé qu'une sélection rapide des commandes était réalisable mais seulement après une formation importante pour la plupart des utilisateurs. Nous avons attribué la forte courbe d'apprentissage à la combinaison de deux facteurs : la faible expertise des participants dans l'utilisation des marking menus et la faible présence du pointeur 2D. Comme les participants se sont rapidement habitués à la présence du curseur 3D à partir de son volume et de son contrôle naturel, le passage au pointeur 2D moins intuitif s'est avéré difficile. De plus, l'utilisation des translations de l'affichage pour interagir avec des widgets présente un inconvénient important : elle modifie la position du curseur et le point de vue sur la scène. L'utilisation d'un petit rayon d'activation pour le menu permet de maintenir le déplacement acceptable pour de nombreuses tâches. Des tests pilotes nous ont orientés vers un rayon d'activation de 14 mm comme choix équilibré permettant une sélection efficace des commandes tout en limitant les fausses activations. Cependant, certaines commandes telles que le zoom peuvent se rapporter à un emplacement précis du pointeur lorsque la commande est dé-

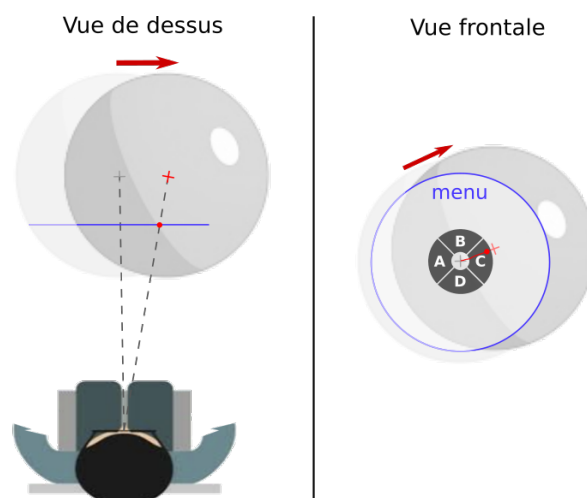


FIGURE 6.7: Pointage dans le marking menu : la translation de la sphère déplace le pointeur 2D dans le marking menu fixe (le point rouge représente le pointer 2D, la croix rouge représente le curseur 3D).

clenchée, et même de petites translations du curseur peuvent être préjudiciables. Cela peut être résolu en utilisant les rotations de l’affichage pour la sélection des éléments.

Un panneau de contrôle qui utilise les rotations de l’appareil

Comme alternative au marking menu, nous avons créé un panneau de contrôle avec l’aspect de présence comme objectif principal de conception. La sélection dans le panneau de contrôle se fait par franchissement des limites des widgets avec un rayon de sélection. Le rayon est matérialisé par une tige rouge de 2 mm de diamètre. La tige apparaît comme provenant du curseur 3D et étant fixée rigidement sur un point de la surface de l’appareil : le long d’un rayon de la sphère. L’orientation de la tige se fait donc intuitivement en tournant l’APCD. La tige est initialement orientée vers l’œil et coupe le panneau de contrôle en son centre. L’intersection est matérialisée par un disque de 10 mm de diamètre représentant le pointeur 2D dans le panneau. Pour optimiser l’espace dans le champ de vision limité de l’écran, le panneau a une forme de disque, les widgets ont des formes d’arc et ils sont répartis dans une disposition d’anneaux concentriques. Il y a une zone vide au centre du panneau pour empêcher l’activation involontaire du widget lorsque le panneau apparaît. Le panneau, les widgets, la tige et son contrôle sont illustrés dans la figure 6.8. La partie de la tige qui se trouve devant le panneau de contrôle pourrait cacher des informations visuelles du panneau derrière elle. Pour éviter ce comportement, nous ajustons la taille de la tige pour qu’elle ne dépasse jamais la profondeur du panneau de plus d’un centimètre, comme on peut le voir dans la figure 6.8. Nous avons gardé cette petite marge car elle permet de visualiser plus facilement l’angle entre la tige et le plan, ce qui est un indice visuel important. Cette implémentation permet une interaction semblable à un menu circulaire à 360°, particulièrement adaptée à un panneau de contrôle en forme de disque.

Le panneau est fixé rigidement en son centre sur l’axe œil-centre de la sphère. Par conséquent, il est toujours face à l’utilisateur et reste entièrement contenu dans la sphère

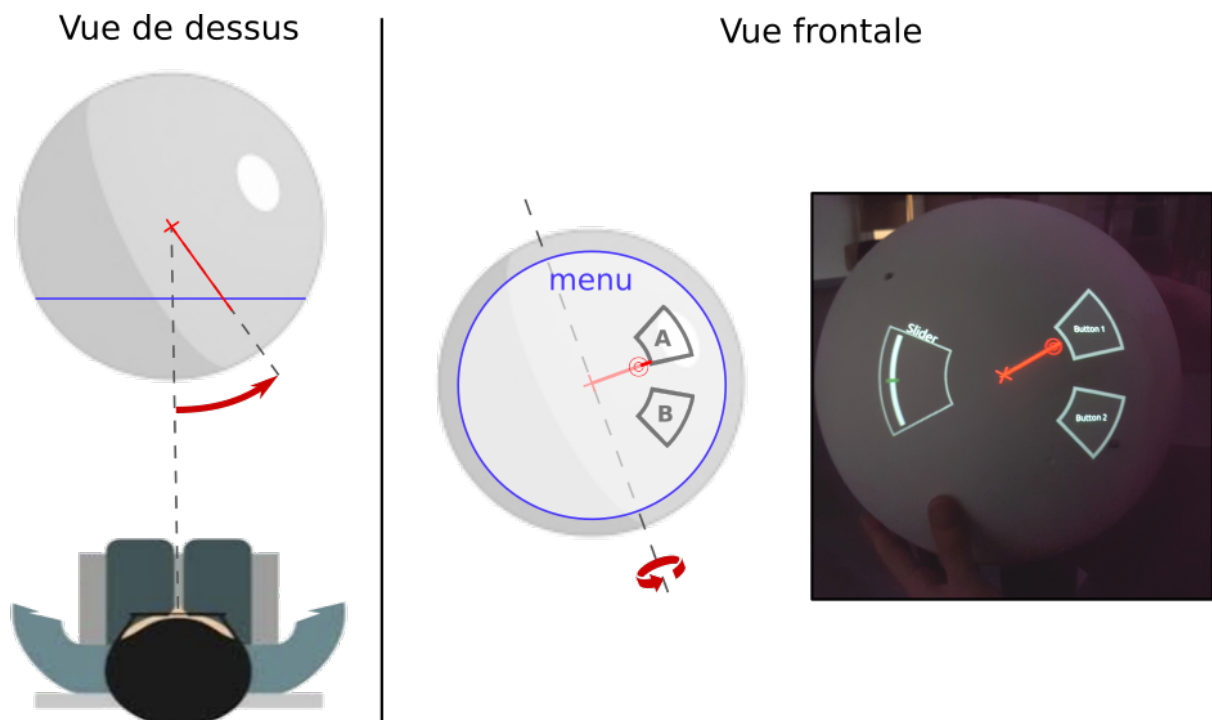


FIGURE 6.8: Pointage dans le panneau de contrôle : la rotation de la sphère déplace le pointeur 2D dans le panneau qui reste toujours centré dans la sphère. À des fins d’illustration, la stéréo est désactivée dans toutes les images du système réel, sauf dans la figure 6.1.

lorsqu'elle est déplacée pour naviguer dans la scène. La tige étant également fixée à l'appareil, la translation de ce dernier a un effet minimal sur l'intersection entre la tige et le panneau. Ces comportements, combinés avec l'arrière-plan transparent du panneau, permettent d'interagir avec le panneau de contrôle tout en naviguant dans la scène sans conflit d'interaction. De plus, des commandes d'activation peuvent être exécutées *sans* déplacer le curseur 3D. Par exemple, la commande de zoom peut être activée tout en maintenant le curseur au centre de zoom voulu. Nous avons testé trois types de widgets dans ce panneau de contrôle : les boutons, les sliders et les menus radiaux hiérarchiques.

Cette approche pourrait permettre un contrôle plus efficace que l'utilisation des translations (telles qu'elles sont utilisées dans le marking menu). En effet, de petites rotations induisent de grandes translations avec moins de demandes motrices lorsqu'elles sont projetées sur un plan perpendiculaire loin du centre de rotation avec l'effet de bras de levier. Cet effet géométrique est le principal avantage du pointage par lancer de rayon et peut être exploité même à petite distance du centre de rotation : une rotation de 10° induira une translation de 2.6 cm à une distance de 15 cm. Comparer les performances de ces deux approches dans le contexte des APCD serait intéressant, et nécessiterait une étude utilisateur contrôlée.

Boutons

Nous avons conçu des boutons en forme d'arc plutôt que rectangulaires. Ainsi, les frontières des widgets sont définies dans un système de coordonnées cylindriques, par des

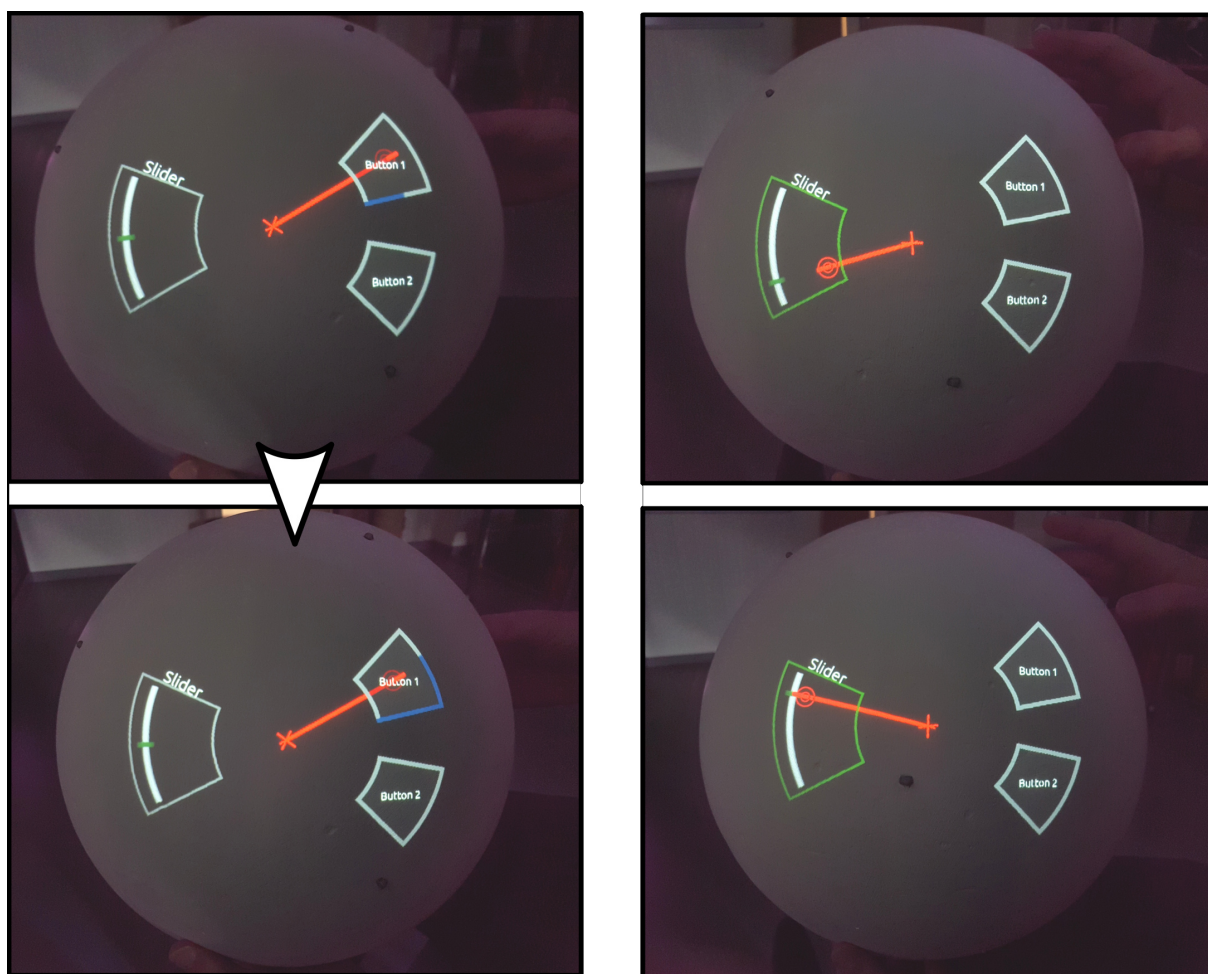


FIGURE 6.9: Interagir avec les widgets. À gauche : retour visuel progressif lors de l'activation d'un bouton. À droite : interaction par franchissement avec un slider.

angles minimum et maximum et par des rayons minimum et maximum. Nous avons testé deux modes d'activation pour les boutons. Avec une activation instantanée, un bouton est activé dès que la tige franchit sa frontière. Tout en étant très efficace, il s'est révélé perturbant pour les utilisateurs novices. La plupart des utilisateurs n'ont aucune expérience de l'interaction par franchissement. En transférant leur expérience à partir des interfaces d'ordinateur de bureau, beaucoup s'attendent à pouvoir survoler les boutons et que seul un toucher du doigt active le bouton. Cela peut entraîner des activations involontaires qui pourraient être stressantes. Nous avons donc introduit un mode d'activation par temporisation. Lorsque la tige entre dans un widget et tant qu'elle reste à l'intérieur, le contour du widget s'allume progressivement en bleu, comme illustré sur la figure 6.9, à gauche. Le bouton n'est activé que si la tige reste à l'intérieur jusqu'à ce que le contour soit entièrement allumé. Après des tests pilotes et en cohérence avec les conclusions de Muller [106], nous avons utilisé un temps d'attente de 0.5 s.

Sliders

La valeur d'un slider est constamment mise à jour dès que la tige entre dans le widget. Elle est définie par l'intersection de l'axe en forme d'arc du slider et d'une ligne entre le centre du panneau et le pointeur 2D, comme l'illustre la figure 6.9, à droite. Apitz et al. indiquent que l'interaction par franchissement permet de viser directement la valeur souhaitée sans avoir à acquérir d'abord le curseur d'un slider [3]. Le choix des frontières du widget offre l'avantage supplémentaire de permettre à l'utilisateur de contrôler le compromis vitesse-précision lors de l'interaction : plus le pointeur 2D est éloigné du centre du panneau, plus il est précis, mais au prix de mouvements plus importants. Lorsque les utilisateurs ont défini la valeur souhaitée, ils doivent déplacer le pointeur 2D hors des limites du widget. Ce mouvement de sortie peut créer un changement de valeur involontaire. Lorsqu'une valeur précise est nécessaire, les utilisateurs peuvent soit garder le pointeur près de la limite du curseur, soit faire un tap avec le doigt. Le tap ferme le panneau contextuel tout en gelant la valeur du curseur, mais demande à l'utilisateur un double tap supplémentaire pour rouvrir le panneau s'il souhaite interagir avec d'autres widgets.

Nous avons testé l'utilisation d'un slider pour contrôler le facteur de zoom de la scène comme alternative à la rotation directe de l'appareil dans un mode spécial présenté dans la section 6.2.3. Cela a l'avantage d'offrir une plage visuelle des valeurs de zoom possibles et de permettre un accès direct au zoom tout en restant dans le panneau de contrôle, mais au prix d'une certaine occlusion de la scène par les widgets du panneau. Là encore, la position du curseur 3D est utilisée pour définir en continu le centre du zoom car l'interaction avec le panneau n'implique pas de translations de l'appareil. Nous avons préféré mettre en jeu un slider de zoom plutôt que d'utiliser la rotation directe dans notre application de démonstration afin d'homogénéiser les différents contrôles de l'application au sein du panneau de contrôle unique.

Menu radial

Après avoir étudié le comportement des utilisateurs experts avec les marking menus, Kurtenbach et al. [85] ont suggéré de limiter les menus à huit éléments par niveau de hiérarchie. Au-dessus de cette limite, le taux d'erreur augmente et devient de moins en moins acceptable. Nous avons implémenté un menu radial [125] dans le panneau afin d'offrir un mécanisme de sélection de commandes plus intuitif que le marking menu présenté précédemment. Afficher le menu à l'intérieur du panneau de contrôle est également un moyen de regrouper tous les contrôles d'application au sein d'un seul panneau. Le menu affiche les niveaux hiérarchiques de commandes sur des arcs concentriques particulièrement adaptés au panneau en forme de disque, comme illustré dans la figure 6.10. Nous avons ajouté

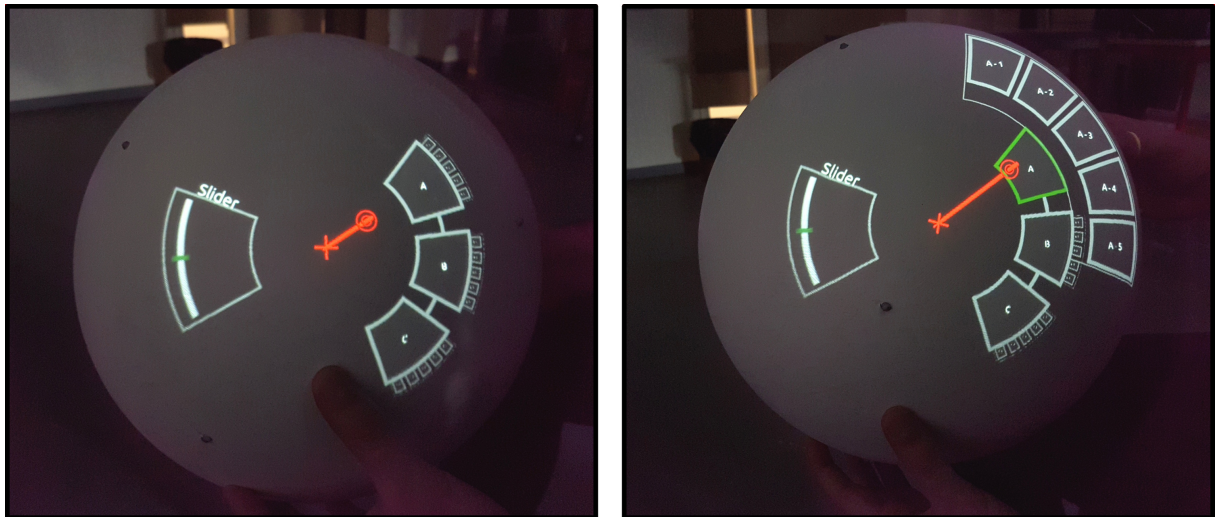


FIGURE 6.10: Un menu radial et un slider dans le panneau de contrôle. À gauche : tous les sous-menus affichent uniquement des aperçus réduits. À droite : le nœud “A” est sélectionnée et son sous-menu est développé (révélant tous ses nœuds fils), tandis que les sous-menus de ses nœuds frères (“B” et “C”) restent réduits.

un temps d’attente de 200 ms entre la sélection de deux éléments frères pour empêcher la fermeture d’un sous-menu pendant le survol d’un frère qui se produit quand on trace une diagonale pour aller sélectionner un élément fils. Une version miniature des sous-menus est affichée en bordure externe de leur nœud parent comme illustré dans la figure 6.10. Bien que ces miniatures ne permettent pas la lecture des éléments, elles servent à deux fins : elles informent que le parent a un sous-menu et elles informent sur le nombre d’options. Lors du franchissement de la frontière interne d’un nœud parent, le sous-menu est animé dans un mouvement SI/SO pour atteindre son échelle normale, soulignant la cohérence de l’interface. Il revient à sa taille miniature dans une animation inverse si un autre nœud est sélectionné. Les feuilles de la hiérarchie se comportent comme les boutons présentés précédemment, c’est-à-dire avec un éclairage progressif de leur contour. L’exploitation de l’interaction par franchissement, en particulier pour les nœuds de la hiérarchie, évite la nécessité de plusieurs activations pour sélectionner un élément.

6.3 Démonstration d’usage de la panoplie interactive

6.3.1 Une application pour apprendre l’anatomie

Nous avons implémenté une application de test pour montrer comment les différentes modalités d’interaction présentées dans ce chapitre peuvent être utilisées pour développer une application autonome, qui met en jeu la facilité de manipulation et l’intuitivité des APCD. Nous avons choisi de développer une application permettant l’apprentissage de l’anatomie. Ce domaine a été identifié par Chen et al. [33] comme l’un des domaines médicaux qui pourrait grandement bénéficier de la réalité mixte et augmentée et pour lequel ces systèmes restent aujourd’hui sous-exploités en raison de limitations techniques. De plus, Guillot et al. [63] suggèrent que l’apprentissage de l’anatomie à partir d’objets virtuels 3D réalistes pourrait réduire les exigences sur les capacités de rotation mentale, qui, variant considérablement entre les élèves, conduisent à des inégalités dans le processus d’apprentissage. L’apprentissage de l’anatomie est un cas appliqué de l’apprentissage d’objets 3D tel que nous l’avons défini et étudié dans cette thèse, et pour lequel nous avons montré les bénéfices des APCD dans le chapitre 4. Le rendu et la manipulation haute-fidélité fournis par les APCD semblent ainsi offrir une approche intéressante pour apprendre l’anatomie en manipulant et en assemblant des éléments anatomiques virtuels aussi naturellement que possible. Enfin, une telle application ne nécessite pas une large gamme de fonctionnalités pour présenter un intérêt pour les utilisateurs, ce qui la rend adaptée pour une étude qualitative avec des participants novices.

L’application demande aux étudiants de positionner correctement des structures anatomiques 3D dans un membre inférieur cible, en effectuant une série de tâches de placements à 6 degrés de liberté (6D). La cible est composée d’un maillage transparent de la peau et de la structure à placer. Selon la tâche, certains os sont également affichés pour fournir un contexte anatomique cohérent, comme illustré dans la figure 6.11 à gauche pour la tâche du péroné et à droite pour la tâche du fémur. Pour chaque tâche, l’étudiant doit d’abord reconnaître et sélectionner la structure correspondante puis la positionner correctement dans la cible.



FIGURE 6.11: L’application de démonstration. À gauche : identifier et sélectionner une structure anatomique à partir de sa forme. Au milieu : choisir une pièce à partir de son nom. À droite : la structure choisie dans le menu hiérarchique apparaît dans la scène. Le contour jaune indique la partie actuellement pointée.

Sélectionner une structure anatomique

L'application peut être configurée dans deux modes de sélection des structures anatomiques : le mode *menu* (basé sur la connaissance du nom de la structure) et le mode *picking* (basé sur la reconnaissance de sa forme).

Dans le mode menu, un menu hiérarchique contenant toutes les structures classées par catégorie (c.-à-d. os, muscle ou ligament) et triés par ordre alphabétique est utilisé. Il est contenu dans un panneau de contrôle contextuel. Le menu est illustré dans la figure 6.11, au milieu. Quand l'utilisateur sélectionne un élément du menu, la structure correspondant au nom de l'élément apparaît dans la scène, centrée sur l'emplacement actuel du curseur (c.-à-d. le centre de la sphère), comme illustré dans la figure 6.11, à droite. Ce mode a été conçu pour tester si les utilisateurs parviennent à interagir avec le panneau contextuel et le menu hiérarchique facilement et intuitivement.

Dans le mode picking, toutes les structures de la catégorie de la tâche en cours sont initialement affichés à côté de la cible, comme illustré dans la figure 6.11, à gauche. Toutes les structures anatomiques sont sélectionnables, et l'utilisateur doit reconnaître quelle structure correspond à la cible actuelle. Ce mode a été conçu pour tester la sélection et la manipulation d'objets 3D. En affichant plusieurs objets sélectionnables en même temps, dont certains plus éloignés que d'autres, nous pourrions observer si les utilisateurs ont des difficultés à sélectionner l'objet qu'ils souhaitent, y compris pour les plus éloignés. Nous avons mis en place le curseur magnétique, car il semble particulièrement bien adapté à cette tâche qui nécessite de sélectionner des objets à distance puis de les manipuler précisément dans une zone confortable où la cible a été positionnée.

Placer une structure anatomique

Le membre peut être déplacée librement pour faciliter la tâche de placement. Pour résoudre la tâche, les utilisateurs effectuent une série de glisser-déposer dans l'espace, en alternant au besoin entre des déplacements de structures anatomiques et du membre cible. Une fois la bonne structure placée au bon endroit dans le membre, celui-ci clignote en bleu pour indiquer la réussite de la tâche, et la tâche suivante commence. La tolérance d'acquisition appliquée est de 30° en rotation et 1 cm en position à l'échelle initiale, pour laquelle le membre inférieur mesure 40 cm de hauteur. La tolérance est assez permissive car l'objectif pédagogique est lié au placement relatif des structures et à l'orientation de leurs différentes parties plutôt qu'à leur placement précis.

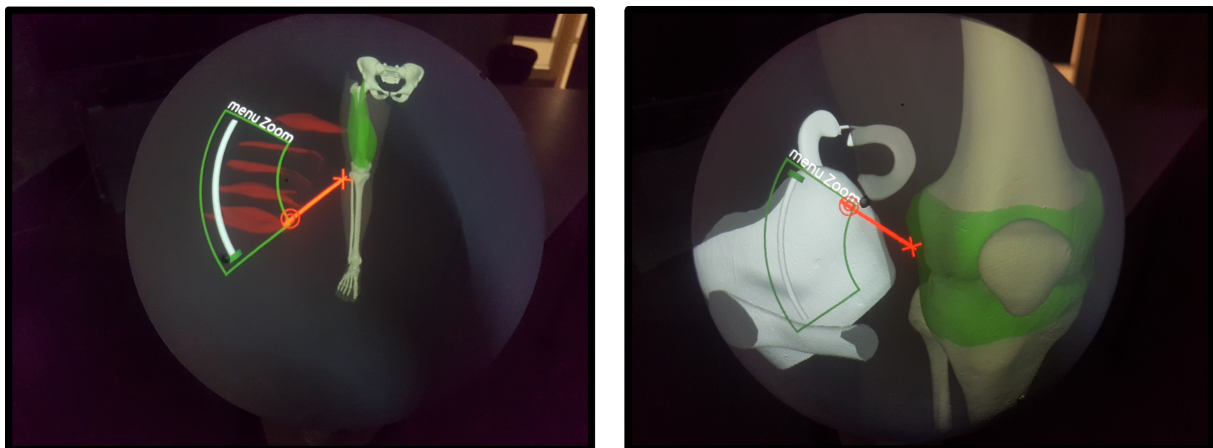


FIGURE 6.12: Adapter l'échelle de la scène à la tâche. À gauche : les muscles nécessitent un petit facteur de zoom. À droite : les ligaments nécessitent un facteur de zoom important.

Contrôler le niveau de zoom

Comme les muscles sont bien plus gros que les ligaments, les utilisateurs peuvent modifier l'échelle de la scène à tout moment pour l'adapter à leurs besoins. Nous avons ajouté un slider dans le panneau contextuel pour contrôler le niveau de zoom de la scène sur une plage logarithmique allant de 0,3 à 3,0. Nous avons choisi d'implémenter le zoom de cette façon afin de tester à la fois la mise à l'échelle et l'interaction avec les sliders. La figure 6.12 illustre la nécessité d'une adaptation du niveau de zoom en fonction de la tâche. Nous avons contrôlé que les utilisateurs utilisaient bien cette fonctionnalité.

6.3.2 Évaluation qualitative

Protocole expérimental

Nous avons accueilli huit étudiants (quatre femmes, âge moyen 23,5 [20, 26]) pour avoir un premier aperçu de l'intuitivité et de l'efficacité des techniques présentées dans ce chapitre. Nous nous sommes concentrés sur les étudiants car ils représentent le public cible d'une application d'apprentissage de l'anatomie. Trois participants n'avaient aucune expérience de l'interaction 3D, trois en avaient une expérience limitée et deux une expérience régulière. Cinq participants n'avaient jamais utilisé d'APCD auparavant, les trois autres avaient déjà essayé notre système mais seulement avec des démos simples. Aucun participant n'avait testé les techniques d'interaction conçues dans cette panoplie interactive au préalable.

Les participants étaient assis sur une chaise. Après une brève introduction sur le contexte de l'expérience, nous leur avons seulement dit qu'ils pourraient interagir avec du contenu virtuel 3D grâce à la sphère. Nous les avons informés qu'ils pourraient déplacer et tourner la sphère, et utiliser leur index pour effectuer des tap et des double tap en touchant la sphère comme s'il s'agissait d'une surface tactile. Sans plus d'explications, ni sur l'interactivité de l'application, ni sur la présence d'un panneau de contrôle, nous avons lancé l'application avec le modèle du membre cible et la possibilité de sélectionner des structures anatomiques dans le menu hiérarchique (mode menu). Nous leur avons laissé découvrir l'application par eux-mêmes pendant quelques minutes avant de révéler les possibilités interactives qui leur avaient échappé. Ensuite, nous leur avons demandé d'effectuer dix tâches de placement dans un premier mode de sélection, suivies des dix mêmes tâches de placement dans l'autre mode de sélection. Nous avons équilibré l'ordre de présentation des modes entre les participants.

Enfin, nous avons recueilli les commentaires subjectifs des participants dans un questionnaire. Ils ont évalué le plaisir d'utilisation de l'application, son utilité pour apprendre l'anatomie, l'inconfort induit par l'utilisation de l'index pour l'activation de commandes, la fatigue musculaire et la douleur mentale ressentie (maux de tête, mal aux yeux). Les notes étaient exprimées sur une échelle allant de 1 (pas du tout) à 5 (totalemment). Les séances ont duré entre trente minutes et une heure au total.

Observations et retours utilisateurs

Tous les participants ont découvert et compris toutes les techniques d'interaction en moins de trois minutes. Un seul participant a eu des problèmes d'interaction avec le panneau de commande, mais cela a été résolu dès que nous lui avons suggéré d'essayer de faire tourner la sphère. Trois participants ont compris et commencé les tâches d'ancrage avant qu'on le leur demande. Après avoir effectué quelques essais, tous les participants étaient confiants dans l'utilisation du système, y compris pour la gestion du niveau de zoom, qu'ils ont utilisé pour résoudre les plus petites tâches ligamentaires.

Contrairement au marking menu, nous avons observé que le menu radial était très intuitif à utiliser, car tous les participants le maîtrisaient facilement. Pourtant, certains participants ont relevé quelques difficultés dans son utilisation. Trois participants ont signalé que le délai empêchant les sélections involontaires en survolant des éléments du menu était trop court, provoquant des activations involontaires. Un participant a indiqué qu'il préférerait une validation explicite par tap avec l'index. Deux participants ont eu du mal à sélectionner le premier et le dernier élément des sous-menus, car ils nécessitaient une rotation importante de l'appareil. Malgré tout, ces problèmes sont restés isolés, avec seulement quatre occurrences dans une même session dans le pire des cas.

Une participante a mentionné qu'elle estimait que son efficacité dans la manipulation de la sphère était en quelque sorte entravée par le fait de devoir garder son index au-dessus de la surface de l'APCD. Une participante a indiqué des activations involontaires causées par l'oubli qu'il fallait garder son doigt au-dessus de la surface au repos. Il convient de noter que l'activation ne se produit que lorsque le doigt *atterrit* sur la surface, donc l'index peut rester en contact sans déclencher de nouvelles activations. Cependant, cela nécessite de garder l'index en contact permanent y compris lors de rotations de l'appareil pour éviter les fausses activations.

La moitié des participants ont capitalisé sur le comportement flottant du panneau de contrôle : ils interagissaient avec les widgets tout en déplaçant l'appareil pour naviguer dans la scène. Cela a été utilisé, par exemple, pour regarder les détails de la cible tout en considérant quelle structure sélectionner dans le menu.

À plusieurs reprises, nous avons observé que des participants utilisaient des mouvements physiques au lieu d'utiliser les outils numériques mis à leur disposition : un participant a déplacé sa tête autour de la sphère pour changer son point de vue sur un objet au lieu de tourner l'objet. Trois participants se sont rapprochés de la sphère pour regarder de plus près plutôt que de changer le niveau de zoom de la scène. Un participant a indiqué qu'il déplaçait inconsciemment la sphère vers des objets éloignés plutôt que d'utiliser le curseur magnétique, bien qu'il ait essayé d'éviter ce réflexe pendant l'expérience car il se rendait compte que c'était une perte de temps. Nous considérons ces comportements comme un signe d'un fort sentiment de présence des objets et d'un haut niveau de présence de l'interaction, favorisant la reproduction de comportements habituels effectués pour interagir avec des objets physiques. Un participant a également signalé que le curseur magnétique se comportait comme les rayons tracteurs illustrés dans de nombreuses œuvres de science-fiction, et qu'il était étrangement naturel et satisfaisant à utiliser. Cela pourrait être le résultat d'un effet de halo des interactions les plus naturelles sur les interactions les moins naturelles.

Résultats du questionnaire subjectif

La figure 6.13 montre les notes des participants concernant l'intuitivité et l'efficacité ressenties par rapport à chaque partie de l'interface. Tous les participants ont jugé toutes les parties de l'interface intuitives (note moyenne > 4), ce qui est cohérent avec nos observations. Toutes les parties de l'interface ont également été identifiées comme étant efficaces (note moyenne $> 3,6$). Le tap avec l'index a été identifié comme légèrement moins efficace que les autres interactions. Nous avons effectivement observé certaines activations involontaires du menu contextuel se produisant à la place d'une sélection d'objet lors d'un second tap effectué rapidement après une sélection manquée. Ce cas peut être considéré comme une erreur d'implémentation, et peut facilement être évité en corrigeant le code.

La figure 6.14 rapporte les notes des participants concernant leurs sensations après les sessions expérimentales. Ils ont tous indiqué qu'ils avaient aimé utiliser l'application et qu'ils avaient appris certaines choses sur l'anatomie du membre inférieur. En général,

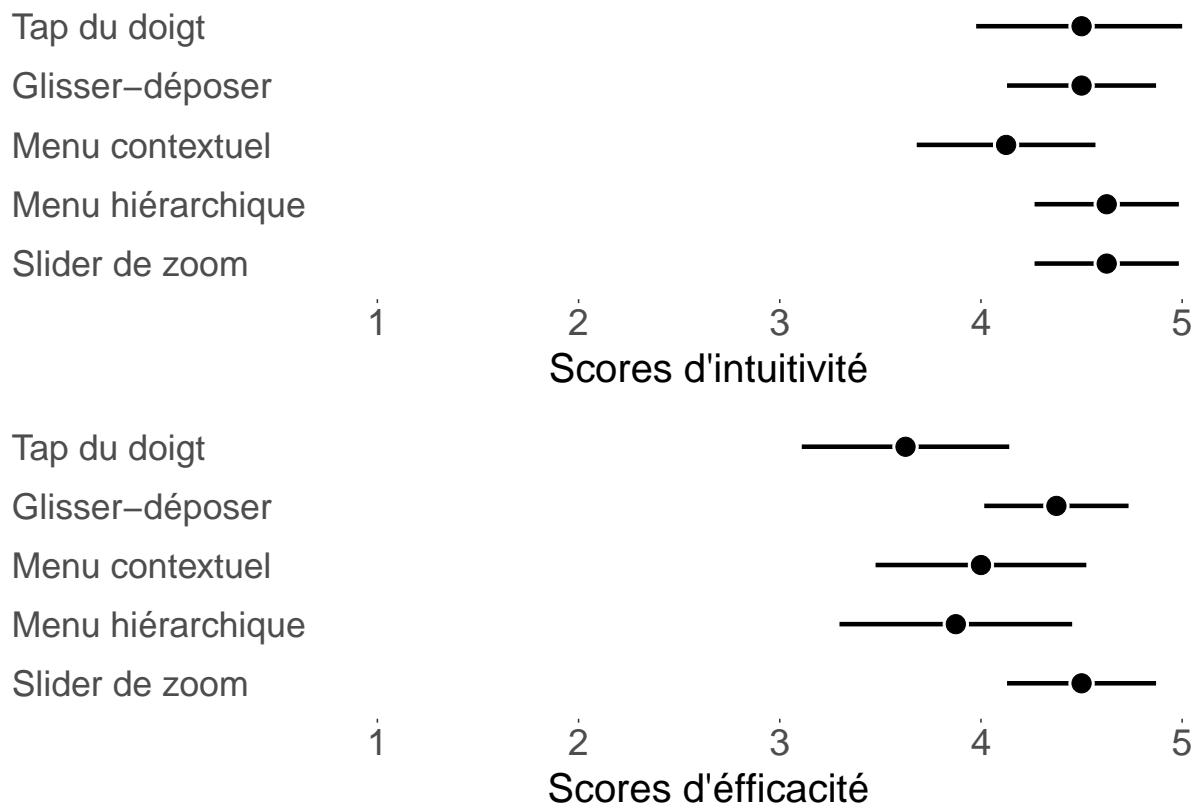


FIGURE 6.13: Notes subjectives des participants concernant l'intuitivité (en haut) et l'efficacité (en bas) de l'interface, avec les intervalles de confiance à 95%.

les participants n'ont pas expérimenté de fatigue musculaire ni de douleur mentale, et le maintien de l'index en survol au dessus de la surface de l'appareil n'était pas un problème. Il y avait cependant quelques exceptions. Deux participants qui tenaient la sphère plus haut que les autres (à hauteur de leurs yeux, plutôt que juste au dessus de leurs genoux) ont signalé un effet de bras de gorille. Un participant avec une myopie corrigée a signalé une légère douleur aux yeux après une session de cinquante minutes.

6.3.3 Perspectives

Notre étude s'est concentrée sur l'intuitivité dans un contexte d'utilisation novice limité dans le temps et avec un petit ensemble de fonctionnalités. Cet effort initial devrait être étendu à des applications plus complexes qui impliquent un contexte d'utilisation plus

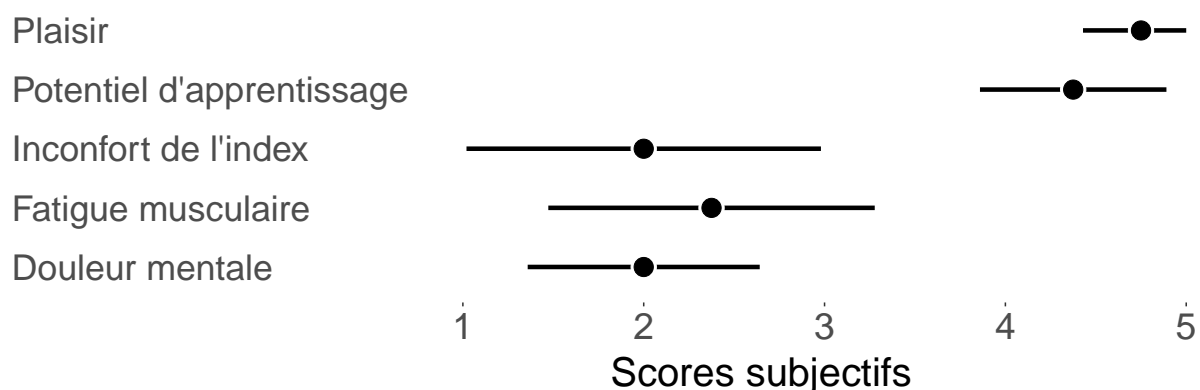


FIGURE 6.14: Ressentis des participants, avec les intervalles de confiance à 95%.

intensif. En particulier, nous avons l'intention d'étudier la saisie de texte et de valeurs numériques. Si l'on met de côté la détection vocale, un clavier virtuel semble être la solution la plus évidente s'il est associé à des techniques de pointage appropriées telles que les pointeurs 2D testés avec le marking menu et le panneau de contrôle. Des approches plus spécifiques peuvent être expérimentées. Par exemple, nous prévoyons de tirer parti du contrôle efficace des rotations de l'appareil pour la saisie numérique : le tangage et le roulis pourraient être combinés, le tangage offrant à la fois une plage étendue et une bonne précision pour définir une valeur, tandis que le roulis pourrait être utilisé pour définir rapidement un gain appliqué à la plage de valeurs. Des études comparatives sont également nécessaires pour fournir des informations plus approfondies sur l'efficacité relative des diverses techniques d'interaction dans le contexte des APCD. Par exemple, nous nous attendons à ce que l'amplitude moindre des gestes nécessaires pour utiliser le marking menu offre de meilleures performances que l'utilisation du menu radial, mais seulement après un entraînement important.

Compromis intuitivité-efficacité

L'utilisation de retours visuels et de transitions animées aide l'utilisateur à comprendre l'interaction. Avec de longues transitions, les utilisateurs novices ont le temps d'analyser et de comprendre le changement qui résulte d'une entrée donnée. Avec un long temps d'attente pour la validation, les utilisateurs ont toujours le temps de corriger un mouvement erroné avant qu'il n'entraîne une sélection involontaire. Avec des widgets larges et éloignés, les utilisateurs n'ont pas besoin d'être précis pour atteindre leurs objectifs. Tous ces choix de conception favorisent l'intuitivité de la technique d'interaction, au détriment de son efficacité. En effet, au fur et à mesure que les utilisateurs se sentent en confiance avec le système, la surcharge de retours visuels n'est plus nécessaire, et chaque temps passé à attendre une animation de transition ou un temps d'attente est une perte de temps. Au fur et à mesure que les utilisateurs deviennent plus précis, de plus petites distances entre les widgets réduiraient le temps passé à déplacer le pointeur.

Les concepteurs doivent garder à l'esprit ce compromis lors de la conception d'applications pour APCD. En vue de favoriser le caractère naturel de l'interaction, de réduire la charge cognitive et d'éviter une phase d'apprentissage fastidieuse, les concepteurs devraient se concentrer sur l'intuitivité de l'interface. Pour une application qui cible un usage professionnel fréquent ou qui nécessite continuellement un grand nombre d'activations, les concepteurs devraient se concentrer sur l'efficacité de l'interface.

Conclusion

Les APCD sont une approche prometteuse pour la visualisation et la manipulation de scènes 3D. Pour développer des applications complètes, déplacer naturellement l'écran ne suffit pas. Il faut également pouvoir sélectionner, manipuler, mettre à l'échelle et interagir avec des panneaux de contrôle et des widgets. Dans ce chapitre, nous avons discuté de la façon dont les APCD peuvent être rendus entièrement interactifs au sens de Mine [103]. Nous fournissons une première proposition pour une panoplie de techniques d'interaction sur lesquelles les concepteurs peuvent s'appuyer pour aller au-delà des démonstrateurs technologiques et vers des applications autonomes.

Cette panoplie est le résultat de nos expériences avec diverses formes de sélection d'objets, de mise à l'échelle de scènes, de paramétrage continu et de sélection de commandes. Bien que largement inspirées par la littérature, nos conceptions incluent des adaptations distinctes pour tirer parti des spécificités d'un écran tenu à deux mains et agissant également comme le principal périphérique d'entrée. En particulier, nous avons constaté que le contrôle efficace des rotations de l'appareil rend celles-ci appropriées à l'interaction avec des menus et des widgets.

Nous avons combiné certaines de ces conceptions dans une application d'apprentissage d'anatomie que nous avons mise entre les mains de huit participants novices. Les participants ont rapidement compris les interactions et ont pu les utiliser pour réaliser des tâches complexes. Les utilisateurs interagissaient intuitivement et avec succès avec toutes les fonctionnalités fournies. Nous avons observé certains comportements réflexes qui résultent de leur expérience de la réalité physique, comme se rapprocher d'un objet pour explorer ses détails ; qui témoignent d'un fort sentiment de présence de la scène et des widgets.

Ces observations suggèrent que les APCD peuvent être utilisés dans un contexte écologique avec des étudiants pour des enseignements liés à la 3D comme les cours d'anatomie ou de géométrie.

Conclusion

Dans cette thèse, nous avons apporté des mesures robustes de certains bénéfices des systèmes haute-fidélité (HF), et nous avons contribué à identifier les caractéristiques importantes dans la réalisation d'un type de système HF particulier : les appareils à perspective corrigée déplaçables (APCD). Nous espérons que ces travaux aideront la communauté à concevoir les techniques d'interaction "farfelues" de demain.

Résumé des contributions

Pourquoi utiliser des systèmes haute-fidélité ?

Nous avons présenté deux contributions qui ont permis de montrer les bénéfices des systèmes d'interaction haute-fidélité.

Notre première contribution, présentée dans le chapitre 3, consistait à quantifier les différences de performance de quatre techniques d'interaction dans une tâche fondamentale de l'interaction 3D : le placement d'objets dans l'espace. Dans une étude utilisateur contrôlée, nous avons montré la supériorité de deux techniques haute-fidélité : un APCD et un casque de réalité virtuelle (RV), par rapport à une technique basse fidélité mettant en jeu un bras articulé et un écran plat. L'APCD et le casque de RV se sont respectivement montrés 43% et 29% plus efficaces que l'utilisation du bras articulé. Nous avons également introduit une méthode pour évaluer l'indice de difficulté des tâches de placement 6D, permettant le calcul d'un indice de performance des techniques testées. Nous avons ainsi calculé des indices de performance de 4,4 et 5,4 bit/s pour le casque de RV et l'APCD, respectivement.

Notre seconde contribution, présentée dans le chapitre 4, consistait à mesurer formellement les bénéfices de deux systèmes d'interaction HF dans une tâche d'apprentissage d'objets 3D. Dans une étude utilisateur contrôlée répartie sur deux jours, nous avons comparé les deux mêmes systèmes HF utilisés pour notre première contribution et une tablette tactile. Cette étude a révélé une amélioration importante (27%) des performances de reconnaissance d'un objet après un entraînement mettant en jeu les techniques d'interaction haute-fidélité, par rapport à l'utilisation de la tablette.

Ces travaux contribuent à une meilleure compréhension des systèmes d'interaction HF en fournissant des preuves empiriques de leurs bénéfices dans deux tâches mettant en jeu des processus cognitifs très différents.

Comment implémenter des systèmes haute-fidélité ?

Nous avons présenté deux contributions apportant des solutions pour faciliter l'implémentation des systèmes d'interaction haute-fidélité.

Notre troisième contribution, présentée dans le chapitre 5, consistait en une étude utilisateur conçue pour mesurer l'effet de quatre paramètres techniques sur le sentiment de présence éprouvé lors d'une expérience interactive haute-fidélité. Cette étude met en jeu une méthode d'évaluation de la présence que nous avons introduite et qui est basée sur des comparaisons relatives de représentations. Les performances des paramètres évalués sont différentes pour chaque représentation. Nous avons testé l'effet de la résolution, de la fréquence d'affichage, de la latence et de la stabilité du système de suivi dans le contexte des APCD sphériques. Les paramètres qui se sont montrés les plus critiques étaient la stabilité du système et la fréquence d'affichage. En revanche, réduire la résolution ne diminuait pas le sentiment de présence pour les objets sans détail fin. Enfin, la latence ne s'est pas révélée critique avec les configurations testées. Ces mesures fournissent des guides pour la conception pertinents pour savoir comment concevoir des systèmes d'interaction HF qui favorisent la présence.

Notre quatrième contribution, présentée dans le chapitre 6, consistait à proposer des solutions pour pouvoir sélectionner, manipuler, mettre à l'échelle et interagir avec des panneaux de contrôle et des widgets, dans le contexte d'un APCD sphérique. Nous avons présenté une panoplie de techniques d'interaction sur lesquelles les concepteurs peuvent s'appuyer pour implémenter des applications autonomes adaptées à ce système. Nos conceptions tirent parti des spécificités d'un écran manipulé à deux mains, notamment du contrôle efficace des rotations de l'appareil pour interagir avec des menus et des widgets. Nous avons testé cette panoplie dans une application d'apprentissage d'anatomie. Une évaluation qualitative nous a permis d'observer que les utilisateurs interagissaient intuitivement et efficacement avec toutes les fonctionnalités fournies.

Ces travaux fournissent des lignes directrices matérielles et logicielles permettant de concevoir des expériences haute fidélité optimales.

Publications

Les contributions de cette thèse ont fait l'objet des publications suivantes :

- Louis, T., & Berard, F. (2017, Octobre). Superiority of a handheld perspective-coupled display in isomorphic docking performances. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (pp. 72-81).
- Louis, T., Troccaz, J., Rochet-Capellan, A., Hoyek, N., & Bérard, F. (2020, Septembre). When High Fidelity Matters : AR and VR Improve the Learning of a 3D Object. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*.
- Louis, T., Troccaz, J., Rochet-Capellan, A., & Bérard, F. (2019, Novembre). Is it Real? Measuring the Effect of Resolution, Latency, Frame rate and Jitter on the Presence of Virtual Entities. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (pp. 5-16).
- Louis, T., Troccaz, J., Rochet-Capellan, A., & Bérard, F. (2020, Octobre). Gyro-Suite : General-Purpose Interactions for Handheld Perspective Corrected Displays. In *Proceedings of the 2020 ACM Symposium on User Interface Software and Technology*.

Perspectives

Cette thèse devrait contribuer au développement de l'utilisation des systèmes d'interaction haute-fidélité.

Nous avons tout d'abord mis en place une méthode pour mesurer l'indice de performance des techniques d'interaction dans la réalisation de tâches de placement 6D [93]. L'utilisation de cette métrique permettrait de comparer de manière robuste les résultats de différentes études à venir, afin de converger vers les techniques les plus efficaces pour manipuler des objets dans l'espace.

Ensuite, l'apport des APCD pour l'apprentissage d'objets 3D [96] suggère que ce type de système HF pourrait contribuer à améliorer différents enseignements scientifiques. En particulier, ce résultat devrait inciter la communauté à démocratiser l'utilisation de ces systèmes dans différents enseignements scientifiques. En effet, nous avons observé que les APCD étaient rapidement pris en main et bien acceptés par les étudiants dans un contexte d'apprentissage d'anatomie. Nous avons fourni les moyens d'implémenter des applications adaptées à ces systèmes [95]. L'intégration de systèmes HF dans des salles de classe semble être un investissement accessible et pertinent pour les universités dès les années 2020. Cela pourrait faire gagner un temps significatif lors de la formation de médecins ou d'architectes, mais également permettre aux étudiants possédant de faibles capacités de rotation mentale de se former efficacement.

La méthode que nous avons introduite pour évaluer l'impact de différents paramètres sur la présence [94] n'est limitée à aucun système, configuration interactive ou domaine de recherche particulier. Cette méthode pourrait être utilisée pour tester rapidement de nombreux paramètres dans divers contextes d'utilisation, afin d'identifier les performances critiques à atteindre avec différents systèmes HF pour favoriser la présence. Cette connaissance est importante pour rendre rapidement accessibles au plus grand nombre les systèmes qui rendront possible les meilleures expériences interactives.

Comme nous l'avons montré dans notre état de l'art, le contexte d'application est déterminant lorsqu'on recherche à identifier les systèmes interactifs les plus adaptés. Face à la multitude de besoins, il est probable que différents systèmes se développeront considérablement durant les prochaines décennies, que ce soit pour augmenter notre environnement avec des informations pertinentes, ou pour produire des expériences fortement immersives.

Dans la vie de tous les jours, de nombreuses augmentations non intrusives peuvent ajouter un confort interactif important. Nous commençons déjà à le voir dans une moindre mesure avec les écrans publicitaires et les affichages qui remplacent les affiches papier dans les métros ou dans les fast-foods. Nous avons déjà l'habitude de pointer ces affiches pour en discuter ou pour passer une commande. Ce geste naturel pourrait être pris en compte pour générer des contrôles intuitifs et efficaces. L'augmentation personnalisée peut également offrir une plus-value importante. L'affichage de direction à suivre directement sur la route ou sur le trottoir permet un guidage GPS intuitif et non ambigu. La traduction automatique d'affiches et de panneaux en superposant le texte traduit à la place du texte original semble pertinente dans un contexte de mondialisation. Pour ces applications, l'exploitation de casques de RA pourrait se répandre, à condition que soit atteint un tel niveau de performance et de miniaturisation que leur port pendant toute une journée serait aussi confortable que celui de lunettes de vue. Décrire un tel niveau de performance peut sembler farfelu en 2020, mais qu'en sera-t-il d'ici dix ou vingt ans ?

Être capable de produire des expériences très immersives pourrait offrir des alternatives écologiquement moins coûteuses pour réaliser de nombreuses activités dans un contexte climatique incertain. De nombreux voyages en avion pourraient être évités grâce à la proposition d'un système de téléconférence où les utilisateurs se sentiraient haute-

ment présents et pourraient collaborer de manière aussi efficace que durant une réunion physique. Utiliser des simulateurs haute-fidélité pour pratiquer des activités à fort impact écologique telles que le ski alpin ou le saut en parachute semble envisageable à la condition que l'expérience soit aussi gratifiante que l'expérience physique simulée ; tant en termes de dépense sportive, que de panorama et de partage. Pour ces applications, il semble capital d'être capable de produire des expériences virtuelles réalistes que l'alternative physique ne surpasse pas, afin que les utilisateurs finaux s'y intéressent et s'y attardent.

Ma conviction est que l'être humain doit rester au centre des préoccupations de conception. C'est en adaptant les systèmes à l'utilisateur, et non pas l'inverse, que les gains finaux seront les plus perceptibles. Je pense que le meilleur moyen de s'adapter à l'humain est de s'inspirer de son environnement naturel. Au-delà de chercher à optimiser les performances techniques brutes, chercher à offrir toujours plus d'intuitivité et de confort aux utilisateurs est ce qui leur permet de vivre des expériences sereines et agréables.

Annexe A

Implémentation du prototype d'APCD de Bérard et Louis

Dans cette annexe, nous détaillons l'implémentation logicielle du prototype d'APCD de Bérard et Louis [25], que nous avons décrit dans la section 1.3.2, et que nous avons utilisé dans les différentes études de cette thèse. Nous présentons les différentes phases de calibration semi-automatique qu'ils ont développées puis l'algorithme utilisé pour calculer en temps réel les images que le projecteur doit afficher en fonction des poses de l'APCD et des lunettes à obturateurs.

A.1 Calibration

Calibration du projecteur

Afin de savoir quel pixel du projecteur doit être allumé pour éclairer un pixel précis de la surface de l'APCD, il faut calibrer les paramètres intrinsèques et extrinsèques du projecteur dans le référentiel du système de capture de mouvements. Étant donné que les paramètres intrinsèques du projecteur (distance focale et champ de vision) peuvent être modifiés et ne sont pas accessibles logiciellement, il ne suffit pas de connaître la pose du projecteur dans le référentiel du système de capture pour le calibrer. Cette calibration du projecteur devant être effectuée après chaque calibration du système de capture et devant



FIGURE A.1: À gauche : une grille de marqueurs est placée dans le champ de projection pour la calibration semi-automatique. À droite : la caméra fixée sous le projecteur voit les marqueurs qui reflètent la lumière envoyée par le projecteur.

être très précise, il est important de mettre en place une calibration semi-automatique.

Une grille de marqueurs rétro-réfléchifs ainsi qu'une caméra placée juste en-dessous du projecteur sont utilisées, de sorte que les marqueurs reflétant la lumière émise par le projecteur soient visibles par la caméra. La figure A.1 illustre cette configuration. En maximisant la luminosité du projecteur, les marqueurs illuminés par le projecteur seront visibles comme des taches blanches facilement identifiables sur les images caméras, tandis que les marqueurs non illuminés seront confondus avec le reste de l'image. Il est alors aisé de déterminer un seuil permettant d'identifier la position des marqueurs dans l'image. La calibration se fait alors en deux phases.

La première phase consiste à identifier la position des marqueurs dans le référentiel image de la caméra en allumant tous les pixels du projecteur afin de s'assurer que tous les marqueurs sont visibles. Cette phase permet d'isoler les zones de l'image caméra correspondant à chaque marqueur, afin d'effectuer la phase suivante pour tous les marqueurs simultanément.

La seconde phase consiste à allumer progressivement les pixels du projecteur étape par étape, et à consulter l'image caméra à chaque étape pour en déduire si un ou plusieurs marqueurs y sont illuminés. Cette seconde phase peut être effectuée naïvement en allumant chaque pixel du projecteur l'un après l'autre, mais cela demanderait des millions d'étapes. À la place, il est possible de coder chaque pixel de l'image projecteur à l'aide d'un code de Gray (produisant un code en 22 bits pour une résolution full HD, et en 24

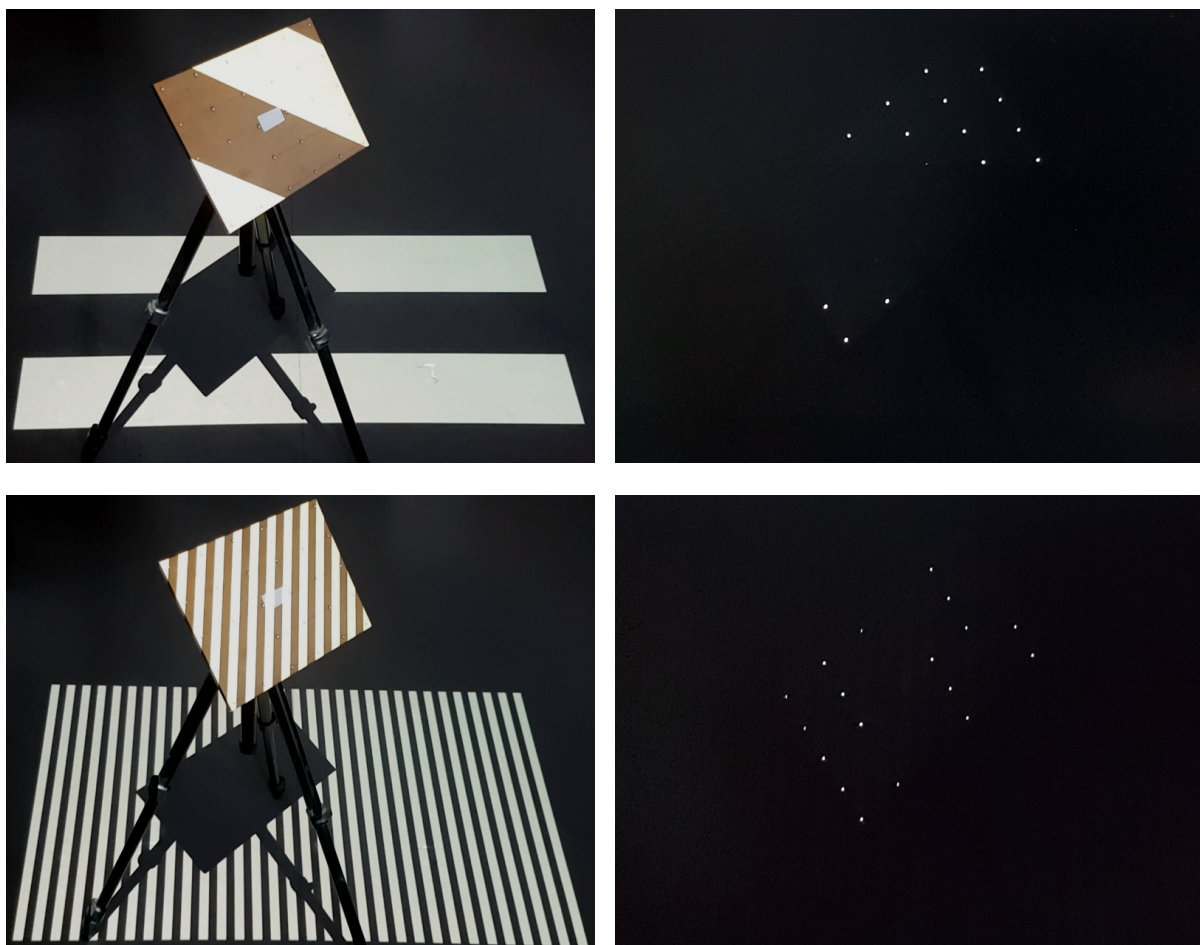


FIGURE A.2: Deux étapes de l'identification par code de Gray de la position des marqueurs dans le référentiel du projecteur. À gauche : les images projetées. À droite : la vue caméra correspondante montrant les marqueurs illuminés.

bits pour une résolution 4K). Une séquence de n images est alors affichée, dans lesquelles chaque pixel est allumé ou non en fonction du n^{eme} bit de son code de Gray. Cette méthode permet d'identifier tous les pixels en une vingtaine d'étapes. La figure A.2 illustre deux de ces étapes. L'utilisation d'un code de Gray permet que chaque erreur d'identification n'induit pas plus d'une erreur d'un pixel. Cette méthode a déjà été utilisée dans un contexte de calibration projecteur-caméra et détaillée par Yamazaki et al. [165]. Après avoir affiché les motifs de codes de Gray, cette phase de calibration est raffinée avec une méthode naïve à proximité des coordonnées trouvées. Pour gagner du temps, il est possible de raffiner en affichant les voisinages ligne par ligne, une dimension après l'autre, au lieu de le faire pixel par pixel, comme l'illustre la figure A.3. Au final, cette phase de calibration aura nécessité une soixantaine d'étapes, et aura duré moins de vingt secondes.

Ces deux phases aboutissent sur une grille de correspondances entre des points 2D dans le référentiel image projecteur et des points 3D dans le référentiel du système de capture optique (qui utilise justement les mêmes marqueurs rétro-réflexifs). Cette calibration doit être répétée en déplaçant et tournant la grille de marqueurs dans l'espace de projection afin d'obtenir quatre ou cinq grilles de correspondances dont les points 3D sont répartis dans tout le volume de projection. Finalement, la fonction OpenCV `cv::calibrateCamera` calcule les paramètres intrinsèques et extrinsèques du projecteur de sorte à optimiser les grilles de correspondances. Ce processus de calibration dure au total environ cinq minutes et permet d'atteindre un niveau de précision d'affichage en-dessous du millimètre dans tout le volume de projection. Nous avons mesuré cette précision informellement, en projetant un point au niveau d'un marqueur qui a été situé à divers emplacements de l'espace de travail.

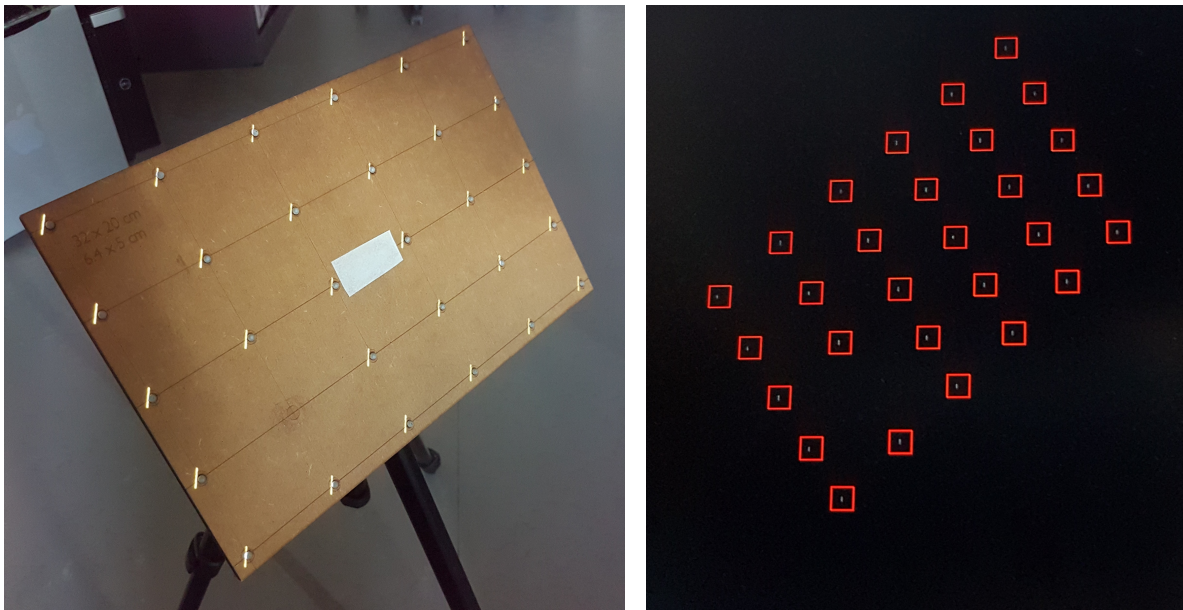


FIGURE A.3: Une étape de raffinement de l'identification de la position des marqueurs en projetant des lignes localement. À gauche : une image projetée. À droite : la vue caméra correspondante. Les carrés rouges indiquent les zones de l'image caméra correspondant à la position des marqueurs ; ils sont affichés dans un but de visualisation uniquement.

Calibration des APCD et des lunettes

La plupart des systèmes de capture optique permettent de déclarer des structures rigides : des objets non déformables sur lesquels au moins trois marqueurs non alignés sont attachés de manière non symétrique. Le système peut retrouver la pose de ces structures rigides dans l'espace en reconnaissant leurs positions relatives. Le système fournit alors, pour chaque structure rigide, la position du barycentre des marqueurs et son orientation par rapport à une orientation originale arbitraire. En définissant une structure rigide avec plus de marqueurs, le système est capable d'identifier le motif, même lorsque certains marqueurs sont cachés, augmentant ainsi la robustesse du système en cas d'occultation. La figure A.4 illustre la répartition des marqueurs sur la sphère et sur les lunettes.

Les APCD étant en permanence manipulés, ils sont sujets à de nombreuses occultations : le corps de l'utilisateur bloque tout un angle de capture, les mains se déplacent sur l'appareil cachant parfois certains marqueurs, et l'objet est régulièrement tourné, changeant la face orientée vers le sol et donc cachée des caméras de suivi. Il est donc important de les recouvrir de nombreux marqueurs répartis sur toute leur surface pour assurer un suivi robuste en toute situation. Avec une dizaine de marqueurs, nous n'enregistrons aucune perte du suivi. Le plus souvent, les marqueurs sont placés aléatoirement sur la surface, ainsi leur barycentre ne correspond pas exactement au centre de l'APCD. Pour pouvoir travailler avec un repère calibré, il est nécessaire d'utiliser un algorithme qui détermine la pose géométrique de l'appareil dans le référentiel de ses marqueurs (i.e. le référentiel communiqué par le système de capture). Les auteurs utilisent un algorithme qui part du postulat que tous les marqueurs sont parfaitement alignés sur une surface 3D dont la géométrie est connue, et qui cherche la pose minimisant la distance des observations à la surface théorique.

Pour les lunettes, la position des yeux doit être calibrée par rapport au référentiel de leur structure rigide. Les auteurs utilisent un marqueur volant supplémentaire, qu'ils placent au niveau du verre droit des lunettes afin de connaître la position de ce verre dans le référentiel des marqueurs. Ils font de même avec le verre gauche puis avec un point très éloigné dans la direction du regard. Ce dernier point permet d'appliquer une correction afin d'approximer la position des yeux par rapport à la position des verres : les auteurs ont mesuré que les verres sont placés à 3 cm devant les yeux lorsque les lunettes sont portées. Enfin, la distance entre les deux yeux est homogénéisée à 65 mm, soit la distance moyenne chez les adultes.

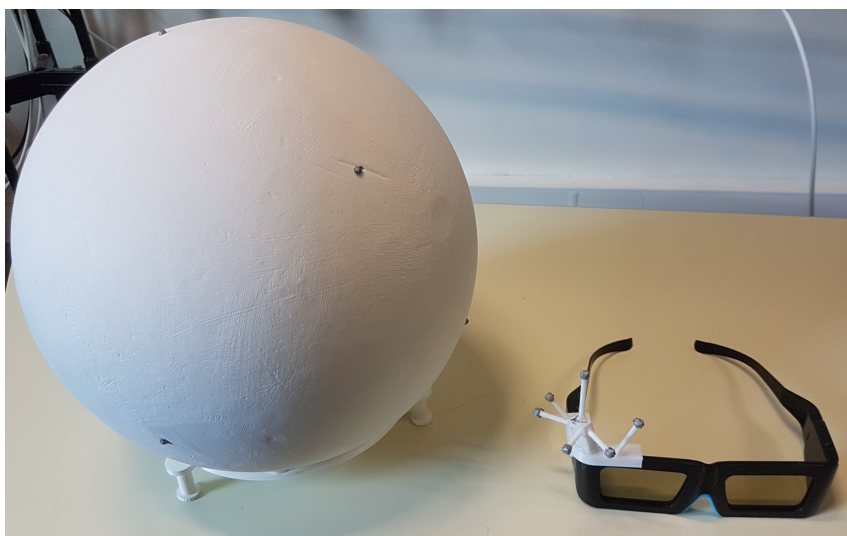


FIGURE A.4: Répartition des marqueurs rétro-réflexifs sur l'APCD et sur les lunettes à obturateurs.

Une fois toutes ces calibrations faites, le système retourne en temps réel la position des yeux et la pose de l'APCD dans un référentiel dans lequel les paramètres intrinsèques et extrinsèques du projecteur sont connus.

A.2 Implémentation logicielle

Bérard et Louis ont développé un logiciel C++ permettant de calculer les images que le projecteur doit afficher afin de visualiser une scène virtuelle au travers de l'APCD, en fonction des positions respectives des yeux, de l'appareil, et du projecteur. Cela permet par la suite de travailler directement avec une scène 3D OpenGL sans se préoccuper de la gestion du rendu. Le rendu s'effectue en trois étapes et est illustré dans la figure A.5.

La première étape consiste à calculer l'image en perspective de la scène virtuelle qu'un œil verrait sur un plan simulé situé entre l'œil et l'appareil. Cette image est obtenue en utilisant une caméra OpenGL placée dans la scène virtuelle au niveau de l'œil et visant le centre de l'APCD. Cette image pourrait être directement utilisée en l'affichant sur un écran plat positionné devant l'APCD, faisant toujours face à l'utilisateur ; mais il ne ferait aucun sens de la projeter telle quelle depuis le point de vue du projecteur.

La seconde étape consiste à mapper l'image précédente sur un maillage 3D représentant la surface de l'APCD. Cette étape permet de texturer un modèle virtuel correspondant exactement à l'appareil physique, de sorte que, vue depuis la position de l'utilisateur, la texture forme exactement l'image obtenue à la première étape. Il s'agit d'une sorte d'illusion d'optique, et le maillage 3D texturé ne semble pas former d'image cohérente s'il est observé depuis un autre point de vue. Si chaque facette de ce maillage correspondait à un écran positionné sur l'APCD (comme il serait envisageable avec un maillage simple de cube tel que celui qu'utilisent Stavness et al. [142]), il serait possible d'afficher directement les textures de chaque facette sur leur écran respectif.

La dernière étape consiste à calculer l'image en perspective que le projecteur verrait d'une scène contenant le maillage 3D précédent positionné à l'emplacement de l'APCD. Cette image est obtenue en utilisant une caméra OpenGL paramétrée avec les paramètres

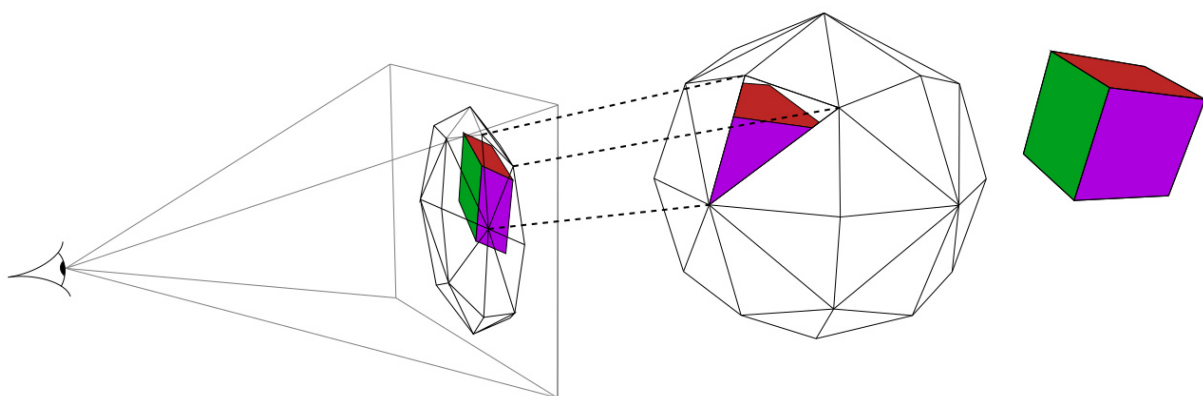


FIGURE A.5: Affichage à perspective corrigée avec mappage de texture. Étape 1 : l'image de texture (base de la pyramide) est créée à partir d'une caméra virtuelle placée à la position de l'œil et visant le centre de l'APCD. Étape 2 : un maillage représentant la forme de l'APCD est texturé en projetant les sommets de chaque triangle dans le plan de projection de la caméra. À des fins d'illustration, un seul triangle d'un maillage sphérique grossier est texturé. La scène virtuelle (le cube multicolore) est affichée derrière l'APCD, mais l'approche se généralise à n'importe quelle position, par exemple à l'intérieur de l'appareil (source : [25]).

intrinsèques et extrinsèques du projecteur. Cette image peut enfin être projetée. Elle illuminera l'APCD de sorte que chacune de ses facettes affiche la texture calculée à la seconde étape et reconstruit ainsi l'image de la scène virtuelle quand l'APCD est vu depuis la position de l'utilisateur.

Le rendu est effectué indépendamment pour chaque œil pour permettre la vision stéréoscopique. Les données de suivi les plus récentes sont toujours utilisées pour chaque rendu. Le système de capture est également synchronisé avec le projecteur pour réduire la latence.

Deux paramètres influent fortement sur le temps de calcul total du rendu (mise à jour de la géométrie et des coordonnées de texture, rasterisation). Ils doivent être optimisés pour maximiser la qualité du rendu, tout en limitant la latence ajoutée. Les deux paramètres sont : la taille de l'image calculée lors de la première étape et utilisée en texture pour la seconde étape, et la taille du maillage utilisé lors de la seconde et de la troisième étape. En ce qui concerne la taille d'image texturée, il n'est pas utile qu'elle soit plus grande que la résolution finale d'affichage du projecteur qui sera limitante au final. Avec leur système, les auteurs ont constaté qu'utiliser une texture de résolution 2048x2048 produisait une même qualité de rendu final que l'utilisation d'une texture de résolution 4096x4096. La taille du maillage est cruciale quand il s'agit d'utiliser des APCD de formes arrondies, qui demanderaient un maillage de taille infinie pour coller parfaitement à la forme de l'appareil. Il faut s'assurer que le maillage approximatif utilisé permette un rendu rapide sans pour autant produire de discontinuité dans l'affichage. Avec un APCD sphérique de 30 cm de diamètre, l'utilisation d'une texture de résolution 2048x2048 avec un maillage constitué de 1536 triangles permet un excellent rendu en moins de 4 ms pour des scènes virtuelles de 100 000 triangles environ, avec ce prototype. Pour afficher des scènes de plus grande complexité, la résolution de la texture peut être réduite à 1024x1024 afin de rester sous les 4 ms de temps de calcul. La réduction de qualité d'image finale dans cette configuration n'est pas visible pour un regard non expert. En revanche, réduire la texture à une résolution de 512x512 a pour résultat une dégradation très perceptible.

Bibliographie

- [1] Johnny ACCOT et Shumin ZHAI. “More than dotting the i’s—foundations for crossing-based interfaces”. In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2002, p. 73-80.
- [2] Valentin AMRHEIN, Sander GREENLAND et Blake MCSHANE. “Scientists rise up against statistical significance”. In : *Nature* 567 (2019), p. 305-307. URL : <https://www.nature.com/articles/d41586-019-00857-9>.
- [3] Georg APITZ et François GUIMBRETIERE. “CrossY : a crossing-based drawing application”. In : *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 2004, p. 3-12.
- [4] Kevin W. ARTHUR, Kellogg S. BOOTH et Colin WARE. “Evaluating 3d task performance for fish tank virtual worlds”. In : *ACM Transactions on Information Systems (TOIS)* 11.3 (1993), p. 239-265.
- [5] Ronald AZUMA et al. “Recent advances in augmented reality”. In : *IEEE computer graphics and applications* 21.6 (2001), p. 34-47.
- [6] Jeremy N. BAILENSON et al. “The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction”. In : *Presence : Teleoperators and Virtual Environments* 15.4 (2006), p. 359-372.
- [7] Domagoj BARIČEVIĆ et al. “A hand-held AR magic lens with user-perspective rendering”. In : *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2012, p. 197-206.
- [8] Lawrence W. BARSALOU. “Grounded cognition”. In : *Annu. Rev. Psychol.* 59 (2008), p. 617-645.
- [9] O. R. BELLOC et al. “OrbeVR : A Handheld Convex Spherical Virtual Reality Display”. In : *ACM SIGGRAPH 2017 Emerging Technologies*. Los Angeles, California : ACM, 2017, 19 :1-19 :2. ISBN : 978-1-4503-5012-9. DOI : 10.1145/3084822.3091104. URL : <http://doi.acm.org/ins2i.bib.cnrs.fr/10.1145/3084822.3091104>.
- [10] Hrvoje BENKO, Andrew D. WILSON et Ravin BALAKRISHNAN. “Sphere : multi-touch interactions on a spherical display”. In : *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 2008, p. 77-86.
- [11] François BÉRARD et Amélie ROCHET-CAPELLAN. “Measuring the linear and rotational user precision in touch pointing”. In : *ACM International Conference on Interactive Tabletops and Surfaces (ITS)*. Cambridge, Massachusetts, USA : ACM, 2012, p. 183-192. ISBN : 978-1-4503-1209-7. DOI : 10.1145/2396636.2396664. URL : <http://doi.acm.org/10.1145/2396636.2396664>.
- [12] Ilias BERGSTRÖM et al. “The plausibility of a string quartet performance in virtual reality”. In : *IEEE transactions on visualization and computer graphics* 23.4 (2017), p. 1352-1359.

-
- [13] Sandra BERNEY. “How Spatial Abilities and Dynamic Visualizations Interplay When Learning Functional Anatomy With 3D Anatomical Models”. In : 2015.
- [14] Lonni BESANÇON et Pierre DRAGICEVIC. “The Continued Prevalence of Dichotomous Inferences at CHI”. In : *ACM Conference on Human Factors in Computing Systems (CHI) Extended Abstracts*. ACM, 2019, alt14 :1-alt14 :11. DOI : 10.1145/3290607.3310432. URL : <http://doi.acm.org/10.1145/3290607.3310432>.
- [15] Lonni BESANÇON et al. “Mouse, tactile, and tangible input for 3D manipulation”. In : *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, p. 4727-4740. URL : <http://dl.acm.org/citation.cfm?id=3025863> (visité le 23/06/2017).
- [16] 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*. 1993, p. 73-80.
- [17] Oliver BIMBER et Ramesh RASKAR. *Spatial augmented reality : merging real and virtual worlds*. CRC press, 2005. URL : <https://books.google.fr/books?hl=fr&lr=&id=JEu3BgAAQBAJ&oi=fnd&pg=PP1&dq=Spatial+Augmented+Reality+Merging+Real+and+Virtual+Worlds&ots=NefqadG1S3&sig=5-MoaKAssk99EZvsRhzlBt4RgB4> (visité le 13/01/2017).
- [18] Oliver BIMBER et al. “The virtual showcase”. In : *ACM SIGGRAPH 2006 Courses*. 2001, 9-es.
- [19] John BOLTON, Kibum KIM et Roel VERTEGAAL. “Snowglobe : a spherical fish-tank vr display”. In : *CHI’11 Extended Abstracts on Human Factors in Computing Systems*. 2011, p. 1159-1164.
- [20] Doug A. BOWMAN et Larry F. HODGES. “An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments”. In : *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997, 35-ff.
- [21] Doug A. BOWMAN et Ryan P. MCMAHAN. “Virtual reality : how much immersion is enough ?” In : *Computer* 40.7 (2007), p. 36-43.
- [22] Doug A BOWMAN et al. *3D user interfaces : Theory and practice*. Redwood City. 2004.
- [23] Frederick P. BROOKS. “What’s real about virtual reality ?” In : *IEEE Computer graphics and applications* 19.6 (1999), p. 16-27.
- [24] Gerd BRUDER et al. “Cave size matters : Effects of screen distance and parallax on distance estimation in large immersive display setups”. In : *Presence : Teleoperators and Virtual Environments* 25.1 (2016), p. 1-16.
- [25] François BÉRARD et Thibault LOUIS. *The Object Inside : Assessing 3D Examination with a Spherical Handheld Perspective-Corrected Display*. 2017.
- [26] François BÉRARD et Thibault LOUIS. “A low-latency, high-precision handheld perspective corrected display”. In : *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2018, p. 397-398.
- [27] François BÉRARD et Amélie ROCHET-CAPELLAN. “The transfer of learning as hci similarity : Towards an objective assessment of the sensory-motor basis of naturalness”. In : *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, p. 1315-1324.

-
- [28] François BÉRARD et al. “Did “Minority Report” get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task”. In : *IFIP Conference on Human-Computer Interaction*. Springer Berlin Heidelberg, 2009, p. 400-414. URL : http://link.springer.com/chapter/10.1007/978-3-642-03658-3_45 (visité le 13/01/2017).
- [29] Xiang CAO et Ravin BALAKRISHNAN. “Interacting with dynamically defined information spaces using a handheld projector and a pen”. In : *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM, 2006, p. 225-234.
- [30] Elie CATTAN et al. “Reducing latency with a continuous prediction : Effects on users’ performance in direct-touch target acquisitions”. In : *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. ACM, 2015, p. 205-214.
- [31] Elie CATTAN et al. “Does Practice Make Perfect ?” In : *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI*. Denver, CO : ACM Press, 2017, p. 5619-5629. DOI : 10.1145/3025453.3025585. URL : <https://doi.org/10.1145%2F3025453.3025585>.
- [32] Kuo-Jen CHAO et al. “Embodied play to learn : Exploring Kinect-facilitated memory performance”. In : *British Journal of Educational Technology* 44.5 (2013), E151-E155. DOI : 10.1111/bjet.12018. URL : <https://doi.org/10.1111%2Fbjet.12018>.
- [33] Long CHEN et al. “Recent developments and future challenges in medical mixed reality”. In : *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2017, p. 123-135.
- [34] Yi-Chen CHEN et al. “Use of tangible and augmented reality models in engineering graphics courses”. In : *Journal of Professional Issues in Engineering Education & Practice* 137.4 (2011), p. 267-276.
- [35] Yu-Chien CHEN, Sheng-Jo WANG et Ya-Ling CHIANG. “Exploring the Effect of Presence in an AR-based Learning Environment”. In : *13th Global Chinese Conference on Computers in Education, Taipei*. Citeseer, 2009.
- [36] Méline CHERDIEU et al. “Make Gestures to Learn : Reproducing Gestures Improves the Learning of Anatomical Knowledge More than Just Seeing Gestures”. In : *Frontiers in Psychology* 8 (2017), p. 1689. DOI : 10.3389/fpsyg.2017.01689. URL : <https://doi.org/10.3389%2Fpsyg.2017.01689>.
- [37] Mingyuan CHU et Sotaro KITA. “Spontaneous gestures during mental rotation tasks : Insights into the microdevelopment of the motor strategy.” In : *Journal of Experimental Psychology : General* 137.4 (2008), p. 706-723. DOI : 10.1037/a0013157. URL : <https://doi.org/10.1037%2Fa0013157>.
- [38] Andy CLARK. *Being there : Putting brain, body, and world together again*. MIT press, 1998.
- [39] Manuel CONTERO et al. “Development of an augmented reality based remedial course to improve the spatial ability of engineering students”. In : *Frontiers in Education Conference (FIE), 2012*. IEEE, 2012, p. 1-5.
- [40] Susan Wagner COOK, Zachary MITCHELL et Susan GOLDIN-MEADOW. “Gesturing makes learning last”. In : *Cognition* 106.2 (2008), p. 1047-1058. DOI : 10.1016/j.cognition.2007.04.010. URL : <https://doi.org/10.1016%2Fj.cognition.2007.04.010>.

-
- [41] Carolina CRUZ-NEIRA et al. “The CAVE : audio visual experience automatic virtual environment”. In : *Communications of the ACM* 35.6 (1992), p. 64-73.
- [42] Mihaly CSIKSZENTMIHALYI et Isabella CSIKSZENTMIHALYI. *Beyond boredom and anxiety*. T. 721. Jossey-Bass San Francisco, 1975.
- [43] James J. CUMMINGS et Jeremy N. BAILENSEN. “How immersive is enough? A meta-analysis of the effect of immersive technology on user presence”. In : *Media Psychology* 19.2 (2016), p. 272-309.
- [44] James DAVIS, Yi-Hsuan HSIEH et Hung-Chi LEE. “Humans perceive flicker artifacts at 500 Hz”. In : *Scientific reports* 5 (2015), p. 7861.
- [45] Michael DEERING. “High resolution virtual reality”. In : *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*. 1992, p. 195-202.
- [46] Susanne DIEKELMANN et Jan BORN. “The memory function of sleep”. In : *Nature Reviews Neuroscience* 11.2 (2010), p. 114-126. DOI : 10.1038/nrn2762. URL : <https://doi.org/10.1038%2Fnrn2762>.
- [47] Julia DIEMER et al. “The impact of perception and presence on emotional reactions : a review of research in virtual reality”. In : *Frontiers in psychology* 6 (2015), p. 26.
- [48] Pierre DRAGICEVIC et al. “Temporal distortion for animated transitions”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, p. 2009-2018.
- [49] Andreas DÜNSER et al. “Virtual and augmented reality as spatial ability training tools”. In : *Proceedings of the 7th ACM SIGCHI New Zealand chapter’s international conference on Computer-human interaction : design centered HCI*. ACM, 2006, p. 125-132.
- [50] David ENGLMEIER et al. “Feel the Globe : Enhancing the Perception of Immersive Spherical Visualizations with Tangible Proxies”. In : *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, p. 1693-1698.
- [51] Dylan FAFARD et al. “FTVR in VR : Evaluation of 3D Perception With a Simulated Volumetric Fish-Tank Virtual Reality Display”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, p. 1-12.
- [52] F. FERREIRA et al. “Spheree : A 3D Perspective-corrected Interactive Spherical Scalable Display”. In : *ACM SIGGRAPH 2014 Posters*. SIGGRAPH ’14. Vancouver, Canada : ACM, 2014, 86 :1-86 :1. ISBN : 978-1-4503-2958-3. DOI : 10.1145/2614217.2630585. URL : <http://doi.acm.org.gate6.inist.fr/10.1145/2614217.2630585>.
- [53] 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. 381-391.
- [54] Bernd FROELICH et al. “The globefish and the globemouse : two new six degree of freedom input devices for graphics applications”. In : *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, p. 191-199. URL : <http://dl.acm.org/citation.cfm?id=1124802> (visité le 13/01/2017).
- [55] Henry FUCHS et al. “Towards performing ultrasound-guided needle biopsies from within a head-mounted display”. In : *International Conference on Visualization in Biomedical Computing*. Springer. 1996, p. 591-600.

-
- [56] Siegfried Johannes GERATHEWOHL. *Fidelity of simulation and transfer of training : a review of the problem*. US Department of Transportation, Federal Aviation Administration, Office of . . . , 1969.
- [57] David GLESSER, François BÉRARD et Jeremy R. COOPERSTOCK. “Overcoming limitations of the trackpad for 3d docking operations”. In : *CHI’13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2013, p. 1239-1244. URL : <http://dl.acm.org/citation.cfm?id=2468578> (visité le 18/01/2017).
- [58] J. GROSJEAN et al. “Evaluation of the Command and Control Cube”. In : *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*. 2002, p. 473-478.
- [59] T. GROSSMAN, D. WIGDOR et R. BALAKRISHNAN. “Multi-finger gestural interaction with 3d volumetric displays.” In : *ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2004, p. 61-70.
- [60] Tovi GROSSMAN et Ravin BALAKRISHNAN. “An evaluation of depth perception on volumetric displays”. In : *Proceedings of the working conference on Advanced visual interfaces*. 2006, p. 193-200.
- [61] Tovi GROSSMAN et Ravin BALAKRISHNAN. “The design and evaluation of selection techniques for 3D volumetric displays”. In : *Proceedings of the 19th annual ACM symposium on User interface software and technology*. 2006, p. 3-12.
- [62] Yves GUIARD. “Asymmetric Division of Labor in Human Skilled Bimanual Action”. In : *Journal of Motor Behavior* 19.4 (1987), p. 486-517. DOI : 10.1080/00222895.1987.10735426. URL : <https://doi.org/10.1080%2F00222895.1987.10735426>.
- [63] Aymeric GUILLOT. “Relationship Between Spatial Abilities, Mental Rotation and Functional Anatomy Learning”. In : 2007.
- [64] Aymeric GUILLOT et al. “Relationship Between Spatial Abilities, Mental Rotation and Functional Anatomy Learning”. In : *Advances in Health Sciences Education* 12.4 (2006), p. 491-507. DOI : 10.1007/s10459-006-9021-7. URL : <https://doi.org/10.1007%2Fs10459-006-9021-7>.
- [65] Mark HANCOCK, Sheelagh CARPENDALE et Andy COCKBURN. “Shallow-depth 3d interaction : design and evaluation of one-, two-and three-touch techniques”. In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007, p. 1147-1156. URL : <http://dl.acm.org/citation.cfm?id=1240798> (visité le 13/01/2017).
- [66] Karin L. HARMAN, G. Keith HUMPHREY et Melvyn A. GOODALE. “Active manual control of object views facilitates visual recognition”. In : *Current Biology* 9.22 (1999), p. 1315-1318. DOI : 10.1016/S0960-9822(00)80053-6. URL : [http://dx.doi.org/10.1016/S0960-9822\(00\)80053-6](http://dx.doi.org/10.1016/S0960-9822(00)80053-6).
- [67] Jay HENDERSON et al. “Investigating the Necessity of Delay in Marking Menu Invocation”. In : *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI’20)*. 2020, p. 13.
- [68] Otmar HILLIGES et al. “HoloDesk : direct 3d interactions with a situated see-through display”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, p. 2421-2430.
- [69] Juan David HINCAPIÉ-RAMOS et al. “Consumed endurance : a metric to quantify arm fatigue of mid-air interactions”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, p. 1063-1072. URL : <http://dl.acm.org/citation.cfm?id=2557130> (visité le 13/01/2017).

-
- [70] Ken HINCKLEY et al. “Usability analysis of 3D rotation techniques”. In : *ACM Symposium on User Interface Software and Technology (UIST)*. Banff, Alberta, Canada : ACM, 1997, p. 1-10. ISBN : 0-89791-881-9.
- [71] Nady HOYEK. “Effectiveness of Three-Dimensional Digital Animation in Teaching Human Anatomy in an Authentic Classroom Context”. In : 2014.
- [72] Nady HOYEK et al. “Enhancement of mental rotation abilities and its effect on anatomy learning”. In : *Teaching and learning in medicine* 21.3 (2009), p. 201-206.
- [73] Tobias HÖLLERER et al. “Exploring MARS : developing indoor and outdoor user interfaces to a mobile augmented reality system”. In : *Computers & Graphics* 23.6 (1999), p. 779-785.
- [74] Paul ISSARTEL et al. “A tangible volume for portable 3d interaction”. In : *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 2016, p. 215-220.
- [75] Robert J. K. JACOB et al. “Integrality and separability of input devices”. In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 1.1 (1994), p. 3-26. ISSN : 1073-0516.
- [76] Robert JK JACOB et al. “Reality-based interaction : a framework for post-WIMP interfaces”. In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2008, p. 201-210.
- [77] Susan JANG et al. “Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment”. In : *Computers & Education* 106 (2017), p. 150-165.
- [78] Brett JONES et al. “RoomAlive : Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units”. In : (2014).
- [79] Brett R. JONES et al. “Build your world and play in it : Interacting with surface particles on complex objects”. In : *2010 IEEE international symposium on mixed and augmented reality*. IEEE, 2010, p. 165-174.
- [80] Brett R JONES et al. “IllumiRoom : peripheral projected illusions for interactive experiences”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013, p. 869-878.
- [81] Kangsoo KIM et al. “Revisiting trends in augmented reality research : a review of the 2nd decade of ISMAR (2008–2017)”. In : *IEEE transactions on visualization and computer graphics* 24.11 (2018), p. 2947-2962.
- [82] Kenrick KIN et al. “Eden : a professional multitouch tool for constructing virtual organic environments”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, p. 1343-1352. URL : <http://dl.acm.org/citation.cfm?id=1979141> (visité le 13/01/2017).
- [83] Max KRICHENBAUER et al. “Augmented reality versus virtual reality for 3d object manipulation”. In : *IEEE transactions on visualization and computer graphics* 24.2 (2017), p. 1038-1048.
- [84] Ernst KRUIJFF, J. Edward SWAN et Steven FEINER. “Perceptual issues in augmented reality revisited”. In : *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*. IEEE, 2010, p. 3-12.
- [85] Gordon KURTENBACH et William BUXTON. “The limits of expert performance using hierarchic marking menus”. In : *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 1993, p. 482-487.

-
- [86] Mikko KYTÖ et al. “Pinpointing : Precise head-and eye-based target selection for augmented reality”. In : *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, p. 1-14.
- [87] Elinda Ai-Lim LEE et Kok Wai WONG. “Learning with desktop virtual reality : Low spatial ability learners are more positively affected”. In : *Computers & Education* 79 (2014), p. 49-58.
- [88] Elinda Ai-Lim LEE, Kok Wai WONG et Chun Che FUNG. “Educational values of virtual reality : The case of spatial ability”. In : *Proceedings of World Academy of Science, Engineering and Technology. Paris, France* (2009).
- [89] Kwan Min LEE. “Presence, explicated”. In : *Communication theory* 14.1 (2004), p. 27-50.
- [90] Jane LESSITER et al. “A cross-media presence questionnaire : The ITC-Sense of Presence Inventory”. In : *Presence : Teleoperators & Virtual Environments* 10.3 (2001), p. 282-297.
- [91] Jingbo LIU et al. “Two-Finger Gestures for 6DOF Manipulation of 3D Objects”. In : *Computer Graphics Forum*. T. 31. Wiley Online Library, 2012, p. 2047-2055. URL : <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2012.03197.x/full> (visité le 19/01/2017).
- [92] Julian LOOSER, Raphael GRASSET et Mark BILLINGHURST. “A 3D flexible and tangible magic lens in augmented reality”. In : *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2007, p. 51-54.
- [93] Thibault LOUIS et Francois BERARD. “Superiority of a Handheld Perspective-Coupled Display in Isomorphic Docking Performances”. In : *Proceedings of the Interactive Surfaces and Spaces - ISS*. Brighton, United Kingdom : ACM Press, 2017, p. 72-81. URL : <https://doi.org/10.1145/2F3132272.3134124>.
- [94] Thibault LOUIS et al. “Is it real? measuring the effect of resolution, latency, frame rate and jitter on the presence of virtual entities”. In : *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. 2019, p. 5-16.
- [95] Thibault LOUIS et al. “GyroSuite : General-Purpose Interactions for Handheld Perspective Corrected Displays”. In : *Proceedings of the 2020 ACM Symposium on User Interface Software and Technology*. 2020.
- [96] Thibault LOUIS et al. “When High Fidelity Matters : AR and VR Improve the Learning of a 3D Object”. In : *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*. 2020.
- [97] P. LUBOS, G. BRUDER et F. STEINICKE. “Analysis of direct selection in head-mounted display environments”. In : *IEEE Symposium on 3D User Interfaces (3DUI)*. 2014, p. 11-18.
- [98] Andrew MAIMONE et Henry FUCHS. “Computational augmented reality eyeglasses”. In : *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2013, p. 29-38.
- [99] Guido MAKRANSKY, Thomas S. TERKILDSEN et Richard E. MAYER. “Adding immersive virtual reality to a science lab simulation causes more presence but less learning”. In : *Learning and Instruction* 60 (2019), p. 225 -236.
- [100] Maurice R. MASLIAH et Paul MILGRAM. “Measuring the allocation of control in a 6 degree-of-freedom docking experiment”. In : *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2000, p. 25-32. URL : <http://dl.acm.org/citation.cfm?id=332403> (visité le 19/01/2017).

-
- [101] Stefan MATEEFF, Joachim HOHNSBEIN et Thomas NOACK. “Dynamic visual capture : apparent auditory motion induced by a moving visual target”. In : *Perception* 14.6 (1985), p. 721-727.
- [102] Frank MEIJER et Egon L. van den BROEK. “Representing 3D virtual objects : Interaction between visuo-spatial ability and type of exploration”. In : *Vision research* 50.6 (2010), p. 630-635. URL : <http://www.sciencedirect.com/science/article/pii/S0042698910000362> (visité le 13/10/2017).
- [103] Mark R. MINE. “Virtual environment interaction techniques”. In : *UNC Chapel Hill CS Dept* (1995).
- [104] Mark R. MINE, Frederick P. BROOKS JR et Carlo H. SEQUIN. “Moving objects in space : exploiting proprioception in virtual-environment interaction”. In : *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, p. 19-26. URL : <http://dl.acm.org/citation.cfm?id=258747> (visité le 03/07/2017).
- [105] Shio MIYAFUJI et al. “Qoom : An interactive omnidirectional ball display”. In : *Proceedings of the 30th annual acm symposium on user interface software and technology*. 2017, p. 599-609.
- [106] Christian MÜLLER-TOMFELDE. “Dwell-based pointing in applications of human computer interaction”. In : *IFIP Conference on Human-Computer Interaction*. Springer, 2007, p. 560-573.
- [107] Albert NG et al. “Designing for low-latency direct-touch input”. In : *ACM Symposium on User Interface Software and Technology (UIST)*. Cambridge, Massachusetts, USA : ACM, 2012, p. 453-464. ISBN : 978-1-4503-1580-7. DOI : 10.1145/2380116.2380174.
- [108] Jocelyn PARONG et Richard E. MAYER. “Learning science in immersive virtual reality.” In : *Journal of Educational Psychology* 110.6 (2018), p. 785.
- [109] Thomas D. PARSONS et Albert A. RIZZO. “Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias : A meta-analysis”. In : *Journal of behavior therapy and experimental psychiatry* 39.3 (2008), p. 250-261.
- [110] David PASSIG et Sigal EDEN. “Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children”. In : *CyberPsychology & Behavior* 4.6 (2001), p. 681-686.
- [111] Emilee PATRICK et al. “Using a large projection screen as an alternative to head-mounted displays for virtual environments”. In : *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2000, p. 478-485. URL : <http://dl.acm.org/citation.cfm?id=332479> (visité le 23/06/2017).
- [112] Andriy PAVLOVYCH et Wolfgang STUERZLINGER. “The tradeoff between spatial jitter and latency in pointing tasks”. In : *ACM SIGCHI symposium on Engineering Interactive Computing Systems (EICS)*. Pittsburgh, PA, USA : ACM, 2009, p. 187-196. URL : <http://doi.acm.org/10.1145/1570433.1570469>.
- [113] Philippe PEIGNEUX et al. “Are Spatial Memories Strengthened in the Human Hippocampus during Slow Wave Sleep?” In : *Neuron* 44.3 (2004), p. 535-545. DOI : 10.1016/j.neuron.2004.10.007. URL : <https://doi.org/10.1016%2Fj.neuron.2004.10.007>.
- [114] Charles PONTONNIER et al. “Designing and evaluating a workstation in real and virtual environment : toward virtual reality based ergonomic design sessions”. In : *Journal on Multimodal User Interfaces* 8.2 (2014), p. 199-208.

-
- [115] Ivan POUPYREV, Suzanne WEGHORST et Sidney FELS. “Non-isomorphic 3D rotational techniques”. In : *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2000, p. 540-547. URL : <http://dl.acm.org/citation.cfm?id=332497> (visité le 19/01/2017).
- [116] Ivan POUPYREV et al. “The go-go interaction technique : non-linear mapping for direct manipulation in VR”. In : *ACM Symposium on User Interface Software and Technology (UIST)*. Seattle, Washington, United States : ACM, 1996, p. 79-80. ISBN : 0-89791-798-7.
- [117] Majid POURMEMAR et Charalambos POULLIS. “Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality”. In : *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. 2019, p. 1-9.
- [118] Matthew PRICE et Page ANDERSON. “The role of presence in virtual reality exposure therapy”. In : *Journal of anxiety disorders* 21.5 (2007), p. 742-751.
- [119] Wen QI et al. “A comparison of immersive HMD, fish tank VR and fish tank with haptics displays for volume visualization”. In : *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*. 2006, p. 51-58.
- [120] Jing QIAN et al. “Portal-ble : Intuitive Free-hand Manipulation in Unbounded Smartphone-based Augmented Reality”. In : *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 2019, p. 133-145.
- [121] Ramesh RASKAR et al. “Shader lamps : Animating real objects with image-based illumination”. In : *Rendering Techniques 2001*. Springer, 2001, p. 89-102.
- [122] Rebekka S. RENNER, Boris M. VELICHKOVSKY et Jens R. HELMERT. “The Perception of Egocentric Distances in Virtual Environments - A Review”. In : *ACM Comput. Surv.* 46.2 (déc. 2013). ISSN : 0360-0300. DOI : 10.1145/2543581.2543590. URL : <https://doi-org.ins2i.bib.cnrs.fr/10.1145/2543581.2543590>.
- [123] Kevin ROCHFORD. “Spatial learning disabilities and underachievement among university anatomy students”. In : *Medical Education* 19.1 (1985), p. 13-26. DOI : 10.1111/j.1365-2923.1985.tb01134.x. URL : <https://doi.org/10.1111/2Fj.1365-2923.1985.tb01134.x>.
- [124] Shirin SADRI et al. “Manipulating 3D Anatomic Models in Augmented Reality : Comparing a Hands-Free Approach and a Manual Approach”. In : *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2019, p. 93-102.
- [125] Krystian SAMP et Stefan DECKER. “Supporting menu design with radial layouts”. In : *Proceedings of the International Conference on Advanced Visual Interfaces*. 2010, p. 155-162.
- [126] Marc A. SCHNABEL et Thomas KVAN. “Spatial understanding in immersive virtual environments”. In : *International Journal of Architectural Computing* 1.4 (2003), p. 435-448.
- [127] Ralph SCHROEDER. “Social interaction in virtual environments : Key issues, common themes, and a framework for research”. In : *The social life of avatars*. Springer, 2002, p. 1-18.
- [128] Martijn J. SCHUEMIE et al. “Research on presence in virtual reality : A survey”. In : *CyberPsychology & Behavior* 4.2 (2001), p. 183-201.

-
- [129] Sarah SHARPLES et al. “Virtual reality induced symptoms and effects (VRISE) : Comparison of head mounted display (HMD), desktop and projection display systems”. In : *Displays* 29.2 (2007), p. 58-69. URL : <http://www.sciencedirect.com/science/article/pii/S014193820700100X> (visité le 23/06/2017).
- [130] Thomas B. SHERIDAN. “Musings on telepresence and virtual presence”. In : *Presence : Teleoperators & Virtual Environments* 1.1 (1992), p. 120-126.
- [131] Do Hyoung SHIN, Phillip S. DUNSTON et Xiangyu WANG. “View changes in augmented reality computer-aided-drawing”. In : *ACM Transactions on Applied Perception (TAP)* 2.1 (2005), p. 1-14.
- [132] Richard SKARBEZ, Frederick P. BROOKS JR et Mary C. WHITTON. “A survey of presence and related concepts”. In : *ACM Computing Surveys (CSUR)* 50.6 (2018), p. 96.
- [133] Mel SLATER. “A note on presence terminology”. In : *Presence connect* 3.3 (2003), p. 1-5. URL : http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/ConsortiumPublications/ucl_cs_papers/presence-terminology.htm (visité le 13/01/2017).
- [134] Mel SLATER. “How colorful was your day ? Why questionnaires cannot assess presence in virtual environments”. In : *Presence* 13.4 (2004), p. 484-493. URL : http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6789102 (visité le 13/01/2017).
- [135] Mel SLATER. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”. In : *Philosophical Transactions of the Royal Society B : Biological Sciences* 364.1535 (2009), p. 3549-3557. URL : <http://rstb.royalsocietypublishing.org/content/364/1535/3549.short> (visité le 13/01/2017).
- [136] Mel SLATER, Martin USOH et Yiorgos CHRYSANTHOU. “The influence of dynamic shadows on presence in immersive virtual environments”. In : *Virtual environments '95*. Springer, 1995, p. 8-21.
- [137] Mel SLATER, Martin USOH et Anthony STEED. “Depth of presence in virtual environments”. In : *Presence : Teleoperators & Virtual Environments* 3.2 (1994), p. 130-144.
- [138] Mel SLATER, Martin USOH et Anthony STEED. “Taking steps : the influence of a walking technique on presence in virtual reality”. In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3 (1995), p. 201-219.
- [139] Francesca SORGINI et al. “Design and preliminary evaluation of haptic devices for upper limb stimulation and integration within a virtual reality cave”. In : *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2016, p. 464-469.
- [140] 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 : *International journal of human-computer studies* 61.6 (2004), p. 751-789. URL : <http://www.sciencedirect.com/science/article/pii/S1071581904001016> (visité le 19/01/2017).
- [141] Martin SPINDLER, Wolfgang BÜSCHEL et Raimund DACHSELT. “Use your head : tangible windows for 3D information spaces in a tabletop environment”. In : *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*. 2012, p. 245-254.

-
- [142] Ian STAVNESS, Billy LAM et Sidney FELS. “pCubee : a perspective-corrected hand-held cubic display”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, p. 1381-1390. URL : <http://dl.acm.org/citation.cfm?id=1753535> (visité le 07/09/2017).
- [143] Anthony STEED et al. “The impact of a self-avatar on cognitive load in immersive virtual reality”. In : *2016 IEEE Virtual Reality (VR)*. IEEE. 2016, p. 67-76.
- [144] Martin F. STOELEN et David L. AKIN. “Assessment of Fitts’ Law for quantifying combined rotational and translational movements”. In : *Human Factors : The Journal of the Human Factors and Ergonomics Society* 52.1 (2010), p. 63-77. URL : <http://hfs.sagepub.com/content/52/1/63.short> (visité le 16/03/2017).
- [145] Ivan E. SUTHERLAND. “The Ultimate Display”. In : *Proceedings of the IFIP Congress*. 1965, p. 506-508.
- [146] John SWELLER. “CHAPTER TWO - Cognitive Load Theory”. In : sous la dir. de Jose P. MESTRE et Brian H. ROSS. T. 55. *Psychology of Learning and Motivation*. Academic Press, 2011, p. 37 -76. DOI : <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>.
- [147] Jari TAKATALO et al. “Presence, Involvement, and Flow in Digital Games”. In : (2010).
- [148] Yichen TANG, Ian STAVNESS et Sidney S. FELS. “The new pCubee : Multi-touch perspective-corrected cubic display”. In : *CHI’14 Extended Abstracts on Human Factors in Computing Systems*. 2014, p. 419-422.
- [149] Robert J. TEATHER et al. “Effects of tracking technology, latency, and spatial jitter on object movement”. In : *2009 IEEE symposium on 3D user interfaces*. IEEE, 2009, p. 43-50.
- [150] Michael TSANG et al. “Boom chameleon : simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display”. In : *ACM Symposium on User Interface Software and Technology (UIST)*. Paris, France : ACM, 2002, p. 111-120. ISBN : 1-58113-488-6. DOI : <http://doi.acm.org/10.1145/571985.572001>. URL : <http://doi.acm.org/10.1145/571985.572001>.
- [151] Huawei TU et al. “Crossing-based selection with virtual reality head-mounted displays”. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, p. 1-14.
- [152] Dimitar VALKOV et al. “2d touching of 3d stereoscopic objects”. In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, p. 1353-1362.
- [153] D. W. F. VAN KREVELEN et Ronald POELMAN. “A survey of augmented reality technologies, applications and limitations”. In : *International journal of virtual reality* 9.2 (2010), p. 1-20.
- [154] Steven G. VANDENBERG et Allan R. KUSE. “Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization”. In : *Perceptual and Motor Skills* 47.2 (1978), p. 599-604. DOI : 10.2466/pms.1978.47.2.599. URL : <https://doi.org/10.2466%2Fpms.1978.47.2.599>.
- [155] Thomas VINCENT, Laurence NIGAY et Takeshi KURATA. “Precise pointing techniques for handheld augmented reality”. In : *IFIP Conference on Human-Computer Interaction*. Springer, 2013, p. 122-139.

-
- [156] Jeenal VORA et al. "Using virtual reality technology for aircraft visual inspection training : presence and comparison studies". In : *Applied ergonomics* 33.6 (2002), p. 559-570.
- [157] Vanessa VUIBERT, Wolfgang STUERZLINGER et Jeremy R. COOPERSTOCK. "Evaluation of Docking Task Performance Using Mid-air Interaction Techniques". In : *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. ACM, 2015, p. 44-52. URL : <http://dl.acm.org/citation.cfm?id=2788950> (visité le 13/01/2017).
- [158] Jonathan WAI, David LUBINSKI et Camilla P. BENBOW. "Spatial ability for STEM domains : Aligning over 50 years of cumulative psychological knowledge solidifies its importance." In : *Journal of Educational Psychology* 101.4 (2009), p. 817.
- [159] Robert WANG, Sylvain PARIS et Jovan POPOVIĆ. "6D hands : markerless hand-tracking for computer aided design". In : *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011, p. 549-558. URL : <http://dl.acm.org/citation.cfm?id=2047269> (visité le 06/04/2017).
- [160] Colin WARE. "Using hand position for virtual object placement". In : *The Visual Computer* 6.5 (1990), p. 245-253.
- [161] Colin WARE et Jeff ROSE. "Rotating virtual objects with real handles". In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 6.2 (1999), p. 162-180. URL : <http://dl.acm.org/citation.cfm?id=319102> (visité le 07/04/2017).
- [162] Pierre WELLNER. "Interacting with paper on the DigitalDesk". In : *Communications of the ACM* 36.7 (1993), p. 87-96.
- [163] Bob G. WITMER et Michael J. SINGER. "Measuring presence in virtual environments : A presence questionnaire". In : *Presence : Teleoperators and virtual environments* 7.3 (1998), p. 225-240.
- [164] Maja WRZESIEN et al. "How technology influences the therapeutic process : a comparative field evaluation of augmented reality and in vivo exposure therapy for phobia of small animals". In : *IFIP Conference on Human-Computer Interaction*. Springer, 2011, p. 523-540.
- [165] Shuntaro YAMAZAKI, Masaaki MOCHIMARU et Takeo KANADE. "Simultaneous self-calibration of a projector and a camera using structured light". In : *CVPR 2011 WORKSHOPS*. IEEE, 2011, p. 60-67.
- [166] Ka-Ping YEE. "Peephole displays : pen interaction on spatially aware handheld computers". In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, p. 1-8. URL : <http://dl.acm.org/citation.cfm?id=642613> (visité le 03/05/2017).
- [167] Shumin ZHAI. "User performance in relation to 3D input device design". In : *ACM Siggraph Computer Graphics* 32.4 (1998), p. 50-54. URL : <http://dl.acm.org/citation.cfm?id=307728> (visité le 19/01/2017).
- [168] Shumin ZHAI, Paul MILGRAM et William BUXTON. "The influence of muscle groups on performance of multiple degree-of-freedom input". In : *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1996, p. 308-315.
- [169] Feng ZHOU, Henry Been-Lirn DUH et Mark BILLINGHURST. "Trends in augmented reality tracking, interaction and display : A review of ten years of ISMAR". In : *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2008, p. 193-202.

-
- [170] Qian ZHOU et al. “Coglobe : a co-located multi-person FTVR experience”. In : *ACM SIGGRAPH 2018 Emerging Technologies*. 2018, p. 1-2.
- [171] David J. ZIELINSKI, Derek NANKIVIL et Regis KOPPER. “Specimen Box : A tangible interaction technique for world-fixed virtual reality displays”. In : *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2017, p. 50-58.