



Approche pour une interaction intuitive et sans contact en bloc opératoire

Alexandre Hurstel

► To cite this version:

Alexandre Hurstel. Approche pour une interaction intuitive et sans contact en bloc opératoire. Chirurgie. Université de Strasbourg, 2019. Français. NNT : 2019STRAD032 . tel-02518854

HAL Id: tel-02518854

<https://theses.hal.science/tel-02518854>

Submitted on 25 Mar 2020

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

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

ÉCOLE DOCTORALE MSII

ED 269

THÈSE présentée par :

Alexandre HURSTEL

soutenue le : **30 Septembre 2019**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Informatique

**Approche pour une interaction intuitive
et sans contact en bloc opératoire**

THÈSE dirigée par :

Mme BECHMANN Dominique

Professeur des Universités, Université de Strasbourg

RAPPORTEURS :

M. BOURDOT Patrick

M. PALJIC Alexis

Directeur de Recherches, Université Paris-Sud

Chargé de Recherches, PSL Research University

AUTRES MEMBRES DU JURY :

M. BASTIEN Christian

Mme DRESP-LANGLEY Birgitta

Professeur des Universités, Université de Lorraine

Directeure de Recherches, Université de Strasbourg

Introduction : problématique et contextualisation

En chirurgie contemporaine, la visualisation des données médicales joue un rôle crucial. Cette visualisation consiste le plus souvent en une représentation des données sous forme d'images 2D ou 3D et sont une aide importante pour la planification des opérations (pré-opératoires), ainsi que pour les procédures (per-opératoires) en tant que support consultatif pour les chirurgiens. C'est ce dernier cas qui a retenu toute notre attention, la consultation des données lors d'une opération demeurant problématique.

Lors d'une intervention chirurgicale en cours, la stérilité doit être préservée à tout instant. Le simple fait de consulter les données au cours d'une opération chirurgicale en cours devient un processus notablement chronophage du simple fait de la dé-stérilisation induite par l'utilisation d'une souris et d'un clavier. Un système d'interaction sans-contact, faisant usage de gestes en guise de contrôles, apparait comme une solution intéressante pour améliorer et fluidifier le déroulement d'une opération chirurgicale faisant usage de consultation d'images.

Ainsi, les travaux présentés dans cette thèse détaillent notre approche méthodologique pour concevoir un tel système, visant un équilibre pertinent entre efficacité et intuitivité.

De l'interaction sans contact à son application en salle opératoire

Après avoir revu le cheminement et développement dans la recherche (principalement en Interface Homme-Machine (IHM) et Réalité Virtuelle (RV)) des systèmes d'interactions gestuelles depuis leur apparition jusqu'à leurs occurrences actuelles, nous nous sommes penchés sur les cas d'utilisation liés aux blocs opératoires. Ce concept a fait ses preuves lors de plusieurs études de cas, nous avons alors décidé d'aborder un problème alors encore inhérent à ces divers travaux : celui de l'usabilité. Des concepts empruntés aux Conceptions Centrées Utilisateurs de la littérature IHM nous ont permis d'élaborer les contours de notre méthodes appliquées au cas d'usage per-opératoire.

Méthodologie et démarche exploratoire

Le cas d’usage pratique pris en référence par notre travail est une opération de laparoscopie^[W] via un exemple de logiciel utilisé dans le bloc opératoire.

Nous avons alors isolé et classifié, en élaborant une taxonomie d’interaction, les différentes interactions fondamentales à la tâche de consultation. L’objectif étant de conserver la même structure fonctionnelle de ce système interactif en substituant les dites interactions fondamentales classiques, par des interactions gestuelles. Cela couplé à une analyse du cas d’usage nous a ensuite permis d’établir différents critères impactant la conception de notre système d’interaction : les critères dits “ contextuels ”, ceux dits “ conceptuels ”, et enfin les critères dits “ techniques ”.

Cette première phase exploratoire nous a alors permis d’écarter une série de gestes rendus inadaptés par ces critères, puis d’élaborer un concept d’interaction bien adapté. Ce dernier repose sur des postures d’enclenchement permettant de réaliser les tâches clés via un système d’interaction sans contact par des mouvements d’amplitude minimale, qui constitue alors un vocabulaire gestuel générique.

Protocole expérimentaux

L’idée clé est d’observer et recueillir les gestes que l’utilisateur produit instinctivement lorsqu’il est confronté à la tâche de produire une interaction donnée. Dès lors, un tel recueil de gestes peut permettre d’identifier les tendances interactives communes aux intuitions collectées. Concrètement, nous employons une méthodologie basée sur des expériences de type “ Magicien d’Oz (Wizard of Oz) ”.

Nous avons divisé notre approche en deux étapes : La première étape de notre démarche consiste à extraire des intuitions gestuelles auprès d’utilisateurs “ tout-venants ” afin d’obtenir un vocabulaire gestuel intuitif générique non spécialisé. La seconde étape de notre démarche a consisté, d’une part, à enrichir et compléter ce vocabulaire par des retours et évaluation provenant cette fois d’utilisateurs cibles, les chirurgiens. Ceci nous permet alors d’exclure ou corriger les gestes non pertinents à l’usage en bloc opératoire. Nous avons alors implémenté ces deux étapes en deux protocoles expérimentaux.

ABSTRACT (ENGLISH)

Introduction : problematic and contextualization

Within modern surgery, medical data visualization plays an essential role. It usually consists of a representation of the data in the form of 2D or 3D images and is an important aid for the planning of operations(pre-op), as well as an advisory support for surgeons during procedures(per-op). For our work, we focused on this latter case.

Sterility must be preserved at all times during an ongoing surgical procedure. Viewing the data during an ongoing surgical operation becomes a significantly time-consuming process simply because of the de-sterilization induced by the use of a mouse and keyboard. A touchless interaction system, using gestures as means of control, appears to be an interesting solution to improve and fluidify the progress of a surgical operation using image consultation.

Thus, the work presented in this thesis details our methodological approach to design such a system, aiming at a relevant balance between efficiency and intuitiveness.

Touchless interaction within the operating room

After reviewing related research work of gesture interaction systems, from their appearance to their current occurrences, we looked at use cases related to operating theatres. This concept has proved its worth in several case studies, and we decided to tackle a problem that was still inherent to this work : usability. Concepts borrowed from the User Centred Design of the HMI literature allowed us to elaborate the counts of our methods applied to the intra-operative use case.

Approach & process

The practical use case referred to by our work is a laparoscopic operation via an example of software used in the operating theatre.

We then isolated and classified, by elaborating an interaction taxonomy, the different user interactions necessary to the consultation task. The goal was to preserve the functional structure of this interactive system by substituting the classical fundamental interactions with gestural interactions. This coupled with an analysis of the case of use then allowed us to establish different criteria impacting the design of our

interaction system : the "contextual" criteria, the "conceptual" criteria, and finally the "technical" criteria.

This first exploratory phase then allowed us to discard a series of gestures made unsuitable by these criteria, and then to develop a well-adapted interaction concept. The latter is based on engagement postures that allow key tasks to be performed via a touchless interaction system using movements of minimal amplitude, which then constitutes a generic gestural vocabulary.

Experiments

The key idea is to observe and collect the gestures that the user instinctively produces when confronted with the task of producing a given interaction. Such a collection of gestures can help to identify interactive trends common to the collected intuitions. Concretely, we use a methodology based on "Wizard of Oz" type experiences.

We have divided our approach into two steps : The first step of our approach consists in extracting gestural intuitions from "all-coming" users in order to obtain a generic non-specialized intuitive gestural vocabulary. The second step of our approach consists, on the one hand, in enriching and completing this vocabulary with feedback and evaluation from the target users, the surgeons. This then allows us to exclude or correct gestures that are not relevant for use in the operating theatre. We then implemented these two steps in two experimental protocols.

TABLE DES MATIÈRES

Introduction	11
Contexte, problèmes et contributions	13
Plan du manuscrit	15
1 Bibliographie	17
1.1 Gestes et interaction	17
Gestes	17
1.1.1 Définition du geste :	17
1.1.2 « Interaction sans-contact »	17
1.1.2.1 Interaction physique :	17
1.1.2.2 Interactions « en vol » :	18
1.1.2.3 Interactions via périphériques externes :	18
1.1.2.4 Interactions via périphériques mettables (wearable) :	18
1.1.2.5 Interaction « nues » :	18
1.1.2.6 Interaction à ciblage spécifique d'une partie du corps :	19
1.1.3 Interaction gestuelle en vol sans-contact à mains nues :	19
1.2 Interaction sans contact à mains nues	19
1.3 Évolution des technologies	20
1.4 La Conception Centrée Utilisateur & l'UX	22
1.4.1 Qu'est-ce que l'UX?	23
1.4.2 La recherche et l'UX	23
1.4.3 Comment faire de l'UX?	24
1.4.4 Se préparer pour l'UX	25
1.5 Approche et conception	26
1.5.1 Conclusion et orientation	29
2 Méthodologie et Démarche	31
2.1 Observation de cas d'usage	32
2.1.1 Environnement d'utilisation	32
2.1.2 Logiciel & interactions	33
2.1.2.1 Cas d'usage " 2D " :	33
2.1.2.2 Cas d'usage " 3D " :	35
2.1.3 Analyse du cas d'usage	36
2.1.3.1 Conditions & contraintes d'environnement	36

2.1.3.2	Choix du capteur	37
2.2	Démarche exploratoire	38
2.2.1	Exploration de gestes	38
2.2.1.1	Translation	39
2.2.1.2	Zoom	40
2.2.1.3	Navigation	41
2.2.1.4	Réglage de la luminosité et du contraste	42
2.2.2	Critères	42
2.2.2.1	Critères contextuels	43
2.2.2.2	Critères conceptuels	44
2.2.2.3	Critères techniques	45
2.2.3	Panel de gestes	46
2.2.3.1	Parenthèse sur les Menus	46
2.2.3.2	Gestes uniques par tâche	47
2.3	Proposition finale	49
2.3.1	Mécanismes interactifs de notre proposition	49
2.3.1.1	“ Posture d’enclenchement ”	49
2.3.1.2	“ Vecteur d’influence ”	49
2.3.2	Vocabulaire conceptuel	53
2.4	Intégrer “ l’intuitivité ”	57
3	Expérience 2D	59
3.1	Approche	60
3.1.1	“ Élicitation ” de gestes	60
3.1.2	Suggestions utilisateur pour un vocabulaire gestuel d’intuition	61
3.1.3	Approche en 2 étapes : raffinement du vocabulaire utilisateur	62
3.2	Protocole - générique	63
3.2.0.1	Dispositif expérimental	64
3.3	Protocole « tout-venants »	65
3.3.1	Procédure	65
3.3.2	Cas d’usage artificiel	66
3.3.3	Vocabulaire gestuel pré-défini	67
3.3.4	Résultats	68
3.3.4.1	Observations & analyses	68
3.3.4.2	Vocabulaire intermédiaire	68
3.4	Protocole « experts »	71
3.4.1	Cas d’usage médical	71
3.4.2	Résultats	71
3.5	Limites & difficultés techniques	74

3.6	Comparaison et discussion concernant une étude similaire	74
4	Conception & Développement	77
4.1	Introduction	77
4.2	Démonstrateur	77
4.2.1	Dépendances et spécifications	77
4.2.2	Compilation et construction	79
4.2.3	Interprétation de gestes	81
4.3	Interface	82
4.4	Conclusion et récapitulatif des fonctionnalités	85
5	Conclusion	87
5.1	Bilan	87
5.2	Limites & perspectives	88
A	Taxonomies	91
B	Protocole : éléments additionnels	93
B.1	Questionnaires	93
B.1.1	Profil utilisateur	93
B.1.2	Expérience d'élicitation	94
B.1.3	Test & évaluation de vocabulaire	95
	Bibliographie	97

INTRODUCTION

Affiliation au projet 3D-Surg

Cette thèse a pour origine le projet *3D-Surg* 2015-2019, porté par Luc Soler. Il s'agit d'un projet issu du " Programme d'Investissement Avenir — PIA " du gouvernement français et bénéficiant du financement de la Banque Publique d'Investissement (BPI). L'objectif principal du projet *3D-Surg* est de « *positionner la France comme leader mondial dans une gamme de technologies 3D appliquées à la chirurgie, à travers le développement de la modélisation 3D des patients, des systèmes sécurisés de transfert de données patient, des nouveaux écrans de visualisation 3D, des interfaces sans contact et des logiciels et lunettes de réalité augmentée* ».

Le projet *3D-Surg* vise à développer un ensemble complet de produits pour permettre la percée d'un nouveau type de chirurgie guidée par l'image exploitant la modélisation des patients à partir de leurs données d'imageries médicale et la visualisation tridimensionnelle combinant environnement réel et virtuel. Grâce à *3D-Surg*, la nouvelle visualisation des données des patients, avant et pendant une opération, permettra une amélioration significative des interventions chirurgicales. Au-delà de ce bénéfice clinique attendu, *3D-Surg* propose de positionner la France comme leader mondial des technologies 3D appliquées à la chirurgie, ce qui représente un enjeu économique majeur pour un marché estimé à plus de 2 milliards € en 2016, avec une croissance annuelle de 12,7%. Pour atteindre ce résultat, *3D-Surg* propose de réunir des acteurs français disposant de technologies complémentaires de pointe couvrant l'ensemble des besoins du marché, de la production d'images 3D (images virtuelles ou réelles) à leur sortie sur tablette, ordinateur portable ou écran 3D, en intégrant des outils de stockage de données normalisées et essentielles pour le domaine médical ainsi que des outils d'interaction et d'exploitation des données. Par ailleurs, ce projet vise la production industrielle française d'équipements dédiés à la salle d'opération. *3D-Surg* propose donc de créer un véritable secteur 3D pour la chirurgie, regroupant des petites et grandes entreprises ainsi que des équipes de recherche technologique et clinique, publiques ou privées, connues et reconnues dans divers domaines.

3D-Surg prend la forme d'un consortium comprenant trois start-ups (Visible Patient, Therapixel et Optinvent), 2 PME (Alioscopia et e-Media), 1 groupe (Altran) et

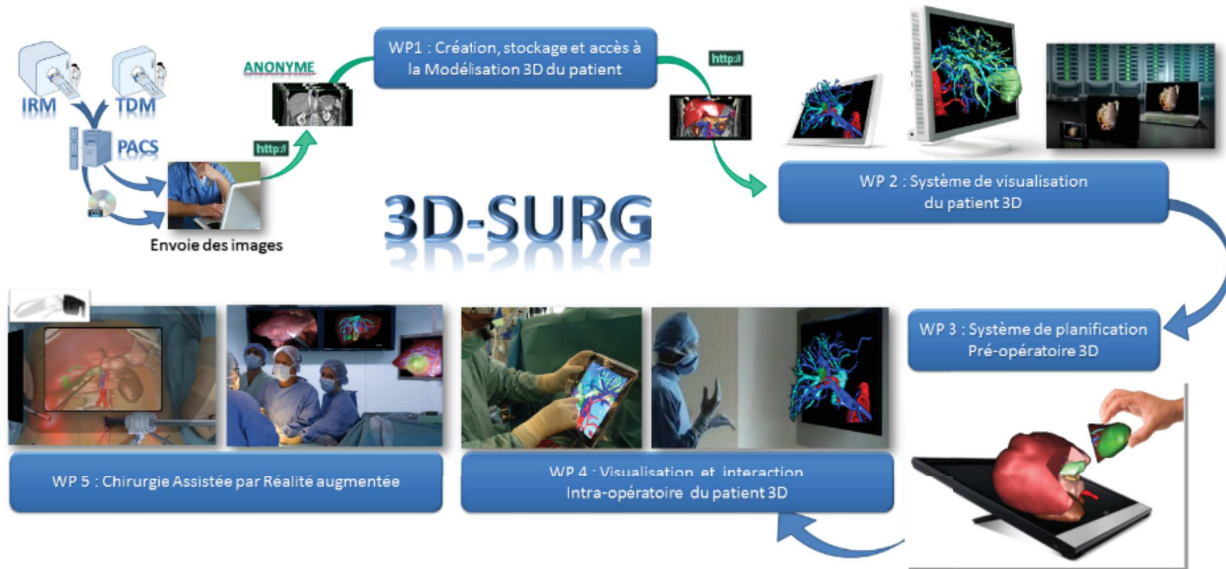


FIGURE 1 – Image tirée du Site web du projet 3D-Surg: <http://www.3d-surg.eu/>

deux instituts privés (IRCAD et IHU Strasbourg). Les partenaires associés sont Alsace Biovalley, l'Ecole Centrale de Nantes et Satt-Connectus représentant 3 équipes de recherche du laboratoire ICube de l'Université de Strasbourg, dont l'équipe IGG.

Le projet fut organisé et découpé en plusieurs “groupes de travail (work-packages)”, axés sur des développement et recherches thématiques :

1. Créations, stockage et accès à la modélisation du patient.
2. Système de visualisation du patient
3. Système de planification Pré-opératoire
4. Visualisation et interaction Per-opératoire du patient
5. Chirurgie assistée par Réalité Augmentée

Chaque groupe de travail est structuré en un découpage fonctionnel plus fin, organisé en sous-tâches. Ainsi, cette thèse réalisée au sein de l'équipe IGG, financée via l'IHU (Institut Hospitalo-Universitaire) de Strasbourg, s'inscrit dans la tâche n° 3.4.1 « Nouveaux modèles d'interaction 3D sans contact ». L'équipe IGG devait également travailler à la tâche n° 3.4.2 « Optimisation et prototypage en vue d'une interaction temps réel ».

En 2018, le projet a été interrompu par la BPI en 2018, et fut définitivement arrêté en 2019. Par conséquent, bien que le contrat doctoral ait pu être mené à son terme, celui-ci n'a pas pu être prolongé par un contrat d'ingénieur, comme initialement prévu.

Contexte, problèmes et contributions

Contexte

En chirurgie contemporaine, la visualisation des données médicales joue un rôle crucial. Cette visualisation consiste le plus souvent en une représentation des données sous forme d'images 2D ou 3D. Ces formes de représentation sont une aide importante pour la planification des opérations (pré-opératoires), ainsi que pour les procédures (per-opératoires) en tant que support consultatif pour les chirurgiens. C'est ce dernier cas qui a retenu toute notre attention dans le présent document, puisque la consultation des données lors d'une opération demeure problématique.

Actuellement, dans les salles d'opération, cette consultation des données par opération se fait à l'aide d'un ordinateur de bureau classique : un ordinateur isolé utilisant la souris et les claviers classiques comme moyen d'interaction avec le système de visualisation des données. Le problème principal et le plus important apporté par ce dispositif est celui de la perte de stérilité. Il apparaît évident que, lors d'une intervention chirurgicale, la stérilité doit être préservée en tout temps. Cependant, l'utilisation d'une souris et d'un clavier ne répond pas à cette nécessité, car ils ne sont pas stériles et leur stérilisation est difficile à réaliser. Cela signifie que l'utilisation de l'ordinateur de bureau oblige le chirurgien (ou tout autre membre de l'équipe chirurgicale) à se dé-stériliser chaque fois qu'il a un besoin consultatif pour l'opération en cours. Cela signifie également que cette personne, une fois la consultation des données terminée, doit être re-stérilisée pour poursuivre sa tâche dans la chirurgie. Par conséquent, le simple fait de consulter les données au cours d'une opération chirurgicale en cours devient un processus notablement chronophage et qui, dans un tel contexte, peut devenir une contrainte critique.

Un moyen de contourner ce problème est d'utiliser une stratégie de délégation des tâches : le chirurgien chef de file délègue la tâche de consultation à l'un de ses assistants, se préservant ainsi du processus de dé-stérilisation/re-stérilisation. Cependant, bien qu'il s'agisse d'une solution acceptable, nous considérons qu'elle n'est " insuffisante " puisque cette solution peut entraîner une perte d'expertise à la consultation puisque l'assistant peut ne pas avoir le même œil expert que le chirurgien en chef. Nous sommes d'avis que, si la contrainte de non-stérilisation/re-stérilisation devait être levée, la tâche de consultation des données deviendrait un outil plus " abordable ", en termes de contraintes, et donc plus fiable et contribuerait à améliorer le déroulement global du travail de l'opération.

Objectif

L'objectif de la thèse consiste à offrir une interaction à mains nues aux praticiens dans le bloc opératoire. Interaction qui devra être possible avec les nombreuses informations ou données à sa disposition : image médicales, visualisation des organes reconstruits, outils de planification de l'opération. Cette thèse se concentrera sur la définition d'un langage gestuel répondant au mieux au besoin d'interaction à identifier auprès des praticiens pour qu'ils puissent utiliser les différents logiciels d'imagerie au cours de leurs pratiques opératoires. Dans ce cadre nous nous sommes concentrés sur le type de logiciel le plus utilisé, qui consiste à consulter des images médicales en navigant image par image à la recherche de ce que le clinicien évaluera être des zones d'intérêts. L'enjeu de ce travail se situe dans la qualité de la définition d'un langage d'interaction, le cas échéant un *vocabulaire gestuel*, permettant un usage adapté aux activités métier des praticiens.

Approche & Contributions

D'une façon globale, notre principale contribution est la mise en place d'une méthode de conception de vocabulaire gestuel adapté et efficace au contexte donné. Au fil de nos travaux, notre démarche a évolué de sorte à intégrer des aspects aussi bien conceptuels que centrés utilisateur au sein de notre méthodologie. Si la mise en œuvre de cette démarche exploratoire s'est faite sur un logiciel particulier utilisé dans le bloc opératoire, nous estimons cette démarche comme suffisamment générique pour être réutilisée afin de concevoir une interaction nouvelle pour d'autres usages.

Notre démarche débute par l'observation d'un cas d'usage : une opération de laparoscopie. Cela nous permet alors d'analyser le contexte d'usage et d'en isoler certaines conditions et contraintes. Dans le même temps nous nous sommes familiarisés avec le logiciel de consultation des images médicales utilisé. Notre démarche a alors consisté à isoler et classifier, via une taxonomie, les différentes interactions fondamentales (pour la tâche) extraites du logiciel. Cette classification a pour objectif d'analyser et décomposer les différentes tâches clés rattachées à ces interactions, dans le cas du logiciel considéré. La même structure fonctionnelle de cette taxonomie sera conservée, l'objectif, d'un point de vue conceptuel, étant de substituer les interactions usant de périphériques classiques, par des interactions gestuelles. Il s'agit alors, en fonction de cette analyse, de pouvoir déterminer le meilleur vocabulaire gestuel pour l'utilisation de ce système de visualisation.

Dans notre démarche, suite à cette analyse, découle l'identification d'une série de critères, classifiés en trois types, impactant le choix des gestes constituant le vocabulaire gestuel. Premièrement, les critères dit " contextuels " qui sont tous les critères issus des besoins et de l'environnement, obtenus par observation du cas d'usage et inter-

views d'utilisateurs. Deuxièmement, les critères dit " conceptuels " qui sont inhérents aux concepts rencontrés dans la littérature de conception d'IHM, tel que l'usabilité, la gestion de la fatigue, etc. Enfin, les critères dit " techniques " qui sont dépendants des outils utilisées, tel que le capteur choisit.

À la lumière de cette première phase exploratoire et des critères établis il est nécessaire, pour chacune des tâches clés, d'écarter les gestes potentiellement inadaptés. À ce stade de notre démarche, il devient alors possible de déterminer un vocabulaire gestuel conceptuel tenant compte d'un maximum de critères.

De la première esquisse de vocabulaire établi, il manque cependant certaines composantes propres à *l'Expérience Utilisateur (UX)*, comme la " naturalité " ou " l'intuition ". Dès lors notre démarche passe par une phase *d'élicitation de gestes*. L'idée clé est d'observer les gestes que l'utilisateur produit instinctivement, face aux tâches clés, pour élaborer un nouveau vocabulaire gestuel centré sur l'intuition. À cet effet nous nous basons sur une méthodologie d'élaboration de protocoles expérimentaux de type " Magicien d'Oz (Wizard of Oz) ". En effet, chaque sujet peut fournir son propre vocabulaire gestuel construit autour de ses intuitions sur la façon d'accomplir ces tâches à l'aide de gestes de la main. Il s'agira ensuite de synthétiser les différents vocabulaires en un recueil de gestes populaires et de tendances interactives récurrentes. Ce recueil étant néanmoins disjoint de nos considérations de conception mis en avant par la plupart de nos critères établis tantôt, il s'agit en réalité de faire évoluer notre vocabulaire conceptuel, de sorte à intégrer cette dimension utilisateur.

Cette démarche s'implémente en deux temps, via deux protocoles : le premier faisant intervenir des utilisateurs « tout-venants », puis le second des utilisateurs finaux ou experts, c'est-à-dire des chirurgiens ou assistants-chirurgiens. Cela nous permet en pratique, à partir du vocabulaire conceptuel initial, d'améliorer ce dernier en un vocabulaire intermédiaire intégrant une dimension " d'intuition ", avant de finaliser et valider ce vocabulaire auprès des utilisateurs experts.

Plan du manuscrit

Ce manuscrit est divisé en 4 chapitres organisés comme suit : Le premier chapitre 1 présente les concepts essentiels et références sur lesquelles se basent nos travaux. Le second chapitre 2 notre cheminement méthodologique. Celui-ci va de pair avec le chapitre 3 suivant qui présente l'implémentation de notre méthodologie en un protocole expérimental ainsi que nos résultats et notre interprétation. Enfin, le dernier chapitre 4 présente le développement informatique nécessaire à nos la réalisation de nos travaux. Naturellement, ce manuscrit se termine par une conclusion et une mise en perspective dans un chapitre 5 de clôture.

RÉFÉRENCES POUR L'INTERACTION

GESTUELLE

1.1 Gestes et interaction

L'interaction gestuelle, à l'épicentre des travaux de cette thèse car celle-ci est le cœur de cible du projet. Naturellement, « l'interaction gestuelle » désigne une interaction contrôlée par les gestes, pour laquelle la réalisation de certaines tâches dépend de la réalisation de certains gestes dans un contexte particulier.

1.1.1 Définition du geste :

Mouvement du corps, principalement de la main, des bras, de la tête, porteur ou non de signification : Faire des gestes en parlant.

Manière de mouvoir le corps, les membres et, en particulier, manière de mouvoir les mains dans un but de préhension, de manipulation : Métier qui demande une précision dans le geste.

-- Définitions tirées du siteweb *larousse.fr*.

1.1.2 « Interaction sans-contact »

Dans cette thèse nous évoquerons à plusieurs reprises « l'interaction sans-contact » ou « l'interaction gestuelle ». Dans le cadre de nos travaux ces termes feront office de synecdoques que nous expliciterons en fin de section (1.1.3). En effet, l'interaction gestuelle peut adopter plusieurs modalités dont les plus courantes sont les suivantes :

1.1.2.1 Interaction physique :

L'interaction gestuelle sur une surface physique interactive, comme l'écran tactile a pris une place dominante dans les usages informatiques personnels de par l'émergence suivie de la conquête par le grand public des smartphones et des tablettes au cours de la décennie passée. Beaucoup de monde en effet est désormais familier avec des gestes tel que le balayage

(“ *swipe* ”) pour naviguer dans une collection d’images, ou encore le pincement (“ *pinch* ”) pour zoomer sur une carte. Ce type d’interaction, basée sur le geste nécessite un contact physique avec le système cible.

1.1.2.2 Interactions « en vol » :

Les interactions gestuelles dites « en vol » (“ *mid-air gesture interactions* ”), à l’opposé des interactions gestuelles sur un support physique, sont induites à partir de gestes produits par un utilisateur dans un espace en 3 dimensions, sans contact avec une surface quelle qu’elle soit, tel un mur immersif ou un écran 3D.

Lorsque les mouvements de la main, positionnée sur un objet 3D contrôle directement les déplacements de cet objet l’interaction est dite directe.

1.1.2.3 Interactions via périphériques externes :

Il s’agit d’utiliser un appareil ou périphérique dont les mouvements dans l’espace (réel) sont trackés. Ainsi la trajectoire de cet appareil, soumis aux actions de l’utilisateur, est utilisée pour extraire et reconnaître les gestes associés aux interactions. D’un certain point de vue, une souris peut donc être considérée comme un périphérique à interface gestuelle. Le geste principal étant de déplacer la souris, dans une optique générale de sélection, auquel s’ajoute des gestes de validation et de contextualisation matérialisés par les cliques souris. Bien entendu, dans cet exemple l’interaction gestuelle est contrainte à un plan 2D matérialisé par la surface sur laquelle est posée la souris. Naturellement, le concept d’interaction à base de gestes évoque bien souvent la possibilité d’une expression gestuelle en 3D. Dès lors, un tel exemple d’interface gestuelle utilisant un périphérique tiers bien connu, est celui de la *WiiRemote*, accessoire principal de la console de jeux *Nintendo Wii* sortie en 2007. Cette console de jeux constitue alors un exemple notable, son catalogue de jeux proposant un nombre important d’exemples vidéo-ludiques offrant des exemples variés et divers d’interaction gestuelle utilisant cet accessoire.

1.1.2.4 Interactions via périphériques mettables (*wearable*) :

L’interaction gestuelle via tracking d’un périphérique wearable consiste à mettre, ou porter (au même sens qu’un vêtement), un accessoire ou périphérique, sur une ou plusieurs parties du corps de l’utilisateur, ces dernières étant alors trackées dans l’espace. Il peut s’agir par exemple de ce qu’on appelle communément des “ gants de données ” qui sont, par exemple, des gants sur lesquels sont disposés des marqueurs réfléchissants pour pouvoir suivre les mouvements de l’utilisateur via des caméras infrarouges. Dans certains cas il peut s’agir d’un capteur à part entière faisant office de dispositif wearable.

1.1.2.5 Interaction « nues » :

Il s’agit d’une interaction se réalisant sans l’usage actif par l’utilisateur de périphériques de quelque nature que ce soit, a contrario des deux interactions gestuelles précédentes (usant

de périphériques). À titre d'exemple le Microsoft Kinect™ est un capteur trackant le corps de l'utilisateur, sans l'aide de marqueurs ou d'autres périphériques permettant une lecture et une interprétation de ses gestes.

1.1.2.6 Interaction à ciblage spécifique d'une partie du corps :

Un geste peut être exprimé de façons diverses via certaines voir toutes les parties du corps d'un utilisateur. Certains systèmes de tracking pour la lecture de gestes ciblent de façon exclusive certaines parties du corps, telles que les mains, ou les bras. C'est le cas notamment du LeapMotion™ qui est un périphérique dédié au tracking et la reconnaissance de la position des mains et des doigts.

1.1.3 Interaction gestuelle en vol sans-contact à mains nues :

Pour des raisons contextuelles que nous avons présentées dans la section précédente et que nous détaillerons au long de cette thèse, l'interaction gestuelle sur laquelle se centrera cette thèse, est une *interaction gestuelle en vol, sans-contact et à mains nues*. Dès lors et **dans la suite de cette thèse**, par souci de concision, lorsque nous parlerons « d'interaction sans-contact » en référence à notre travail, nous parlerons bien sûr d'une telle « interaction gestuelle en vol, sans-contact à mains nues ».

1.2 Interaction sans contact à mains nues

Compte tenu des précédents dans le domaine de l'interaction, et plus spécifiquement l'interaction sans contact, il apparaît que, l'idée de son utilisation dans le contexte per-opérateur a déjà été explorée. D'autant plus que, au cours de la dernière décennie, en raison du regain de popularité des dispositifs d'interaction sans contact, et de leur accessibilité sur le marché multimédia grand public, recherche et études en interaction gestuelle sans contact ont intégré divers domaines d'application, dont le scénario de consultation des données médicales.

Les exemples explorés, de par la nature de nos besoins, et leurs dominances dans la littérature, sur lesquels nous revenons ci-dessous sont des exemples faisant usages de systèmes de détection gestuelle basée sur la vision (“ *Vision-Based Gesture Systems* ”). Un des travaux pionnier dans ce domaine est celui de Krueger et al. [17], qui, dès 1985, posèrent les premières pierres de l'interaction sans-contact grâce à leur prototype “ *VIDEOPLACE* ». Utilisant une projection de la silhouette d'un utilisateur sur un écran, le système pouvait alors permettre d'interagir avec des objets virtuels affichés sur ce même écran, créant ainsi ce que les auteurs appelaient alors une « réalité artificielle ».

Une autre solution technique pour implémenter des interfaces gestuelles serait les gants de données. Il s'agit de périphériques sous forme de gants, auxquels sont intégrés ou greffés divers dispositifs électro-mécaniques permettant de détecter divers positions et mouvements de la main. Ce type de périphériques est apparu lui aussi dès les années 80, comme l'atteste les travaux de Zimmerman et al. [43]. Dès 1987, ils élaborèrent deux modèles de gants données,

les “ Z-Glove ” et les “ DataGlove ”, permettant d’interagir avec des objets d’un environnement virtuel. Les possibilités de tels gants de données sont encore explorées à l’heure actuelle pour divers sujets et notamment l’interaction gestuelle. À titre d’exemple, Kumar et al. [18] proposent un système de reconnaissance gestuelle basé sur un tel périphérique. Dans ces 2 exemples, les gants de données offrent néanmoins une possibilité supplémentaire : le retour tactile. Cependant, de part leur nature mettable, et a fortiori de gants, ces périphériques réservent l’usage des mains à la tâche dédiée d’interaction. Le problème évident des gants est que l’alternance entre l’interaction et d’autres tâches requiert alors de mettre et d’enlever ces gants systématiquement ce qui est un processus encombrant et peu pratique. De plus, ces périphériques semblent difficilement stérilisables. De fait, dans notre contexte du bloc opératoire, l’usage de gants de données au sein du bloc opératoire est inenvisageable pour le moment.

1.3 Évolution des technologies pour l’interaction sans-contact pour le domaine médical

Appliquée au domaine médical, tout comme dans les autres champs d’application plus communs, l’interaction gestuelle « en vol » fut sujette à de nombreuses expérimentations et approches pour en améliorer divers aspects, notamment l’utilisabilité.

En 2004, Graetzel et al. [12] ont présenté un aperçu des avantages offerts par des systèmes d’interaction sans contact pour la consultation de données médicales en substituant le contrôle d’un curseur via une souris par un système utilisant la vision par ordinateur pour suivre les mains des utilisateurs. En effet, ce système n’était pas facile à utiliser, était peu efficace, avait une faible convivialité et provoquait de la fatigue chez les utilisateurs. Ceci est imputable à l’absence d’une conception gestuelle spécifique, et au fait qu’un mapping fonctionnel efficace ne saurait être fourni par une simple substitution des interactions souris.

Dès 2008, Wachs et al. [41] proposent un système d’interaction gestuel en vol plus avancé avec un suivi des mains de l’utilisateur plus élaboré, nécessitant un outil pour le suivi. Le mode d’interaction proposé se base sur le principe suivant : l’utilisateur présente sa main dans le champ de détection, face aux capteurs, pour signifier son intention d’interagir. Puis, le mouvement de la main, suivant son déplacement et ses rotations, va induire une fonction associée : gauche/droit/haut/bas pour naviguer dans une base de données d’images, et avant/arrière pour zoomer sur l’image en question. Des rotations sur une image, induites directement à partir de la rotation de la main, sont également possibles en utilisant l’instrument droit devant le capteur.

Dès son apparition sur le marché fin 2010, le capteur KinectTM [W] de Microsoft a rapidement suscité l’intérêt pour son potentiel en recherche et développement d’interactions gestuelles en vol. Très vite apparaissent nombre de travaux profitant de ses capacités. Un premier exemple est celui de Gallo et al. [11] qui proposent d’utiliser le KinectTM pour manipuler des images médicales via des fonctions de zoom, translation, rotation, etc. Employant un système de machine à état, permettant notamment de discriminer des postures “ limite ” d’activation de la détection de gestes, les fonctions d’un logiciel PACS de visualisation d’images médicales,

MITO, sont directement mappées à des gestes. Ces travaux décrivent ainsi l'implémentation technique d'un contrôle via une interaction gestuelle sans contact en vol d'un logiciel de visualisation d'images médicales existant. De façon similaire, suite à l'apparition du capteur de Microsoft, Ruppert et al. [32] proposent une approche pour explorer et manipuler des données médicales sous forme de modèles 3D. On retrouvera de telles démarches, par exemple dans les travaux de Tan et al. [38] qui ont intégré les fonctionnalités interactives offertes par le Microsoft KinectTM sur le logiciel DiCOM TRICS. Ici, il s'agit de mapper certains gestes avec des événements souris pour pouvoir contrôler le logiciel de visualisation " InVesalius ". D'autres travaux employant le KinectTM pour l'usage en visualisation d'images médicales ont continué à apparaître régulièrement, dont Strickland et al. [37] et Ebert et al. [10]. Ces deux études ont directement mappé les fonctionnalités consultatives de leurs logiciels DiCOM respectifs avec l'ensemble des gestes reconnus par l'appareil. Chacune de ces approches a donné des résultats encourageants, car ces installations ayant été mises à l'épreuve au cours d'opérations réelles ont reçu un retour positif, démontrant ainsi que l'interaction sans contact à des fins consultatives dans le contexte de la chirurgie répond aux attentes enthousiastes des chirurgiens. Cependant, encore une fois, la simple mise en correspondance d'une fonctionnalité avec le geste présente une lacune en utilisabilité qui aurait pu être atténuée si le système interactif avait été développé à l'aide d'un vocabulaire gestuel spécialement adapté, conjointement à des fonctionnalités conçues pour le contrôle gestuel. De plus, Silva et al. [35] ont par exemple élaboré un système interactif spécialement conçu pour la consultation d'images médicales, motivés par la possibilité de faciliter l'interaction avec des logiciels de manipulation d'images médicales dans le bloc opératoire et de surtout maintenir la barrière de stérilité lors d'interactions avec ces logiciels. Le système proposé, basé sur le tracking des joints des squelettes acquis par KinectTM, y est qualifié de " simple et efficace ", car utilisant des gestes simples en observant, au sein de l'espace de détection mappé en 9 régions d'intérêts, les déplacements par rapport aux axes au sein d'une des régions. Fort d'une évaluation utilisateur positive, le système est estimé comme étant " prometteur " reconnaissant toute fois que des améliorations peuvent être apportées au niveau de l'intuitivité de la solution.

Un autre périphérique qui a rapidement suscité de l'intérêt est le LeapMotionTM, dont les premiers modèles fut mis sur le marché en 2013 (premiers modèles distribués dès 2012). Les premières applications de ce capteur pour la consultation d'images médicales, apparaissent dès 2014 avec, par exemple, les travaux de Manolova [20]. À l'instar des premiers travaux intégrant le KinectTM, il s'agissait de mapper directement les fonctionnalités consultatives de logiciels PACS avec un ensemble de gestes reconnus par l'appareil, comme se fut le cas pour Bizzotto et al. [6] qui ont adopté une telle approche en utilisant le LeapMotionTM pour logiciel OsiriX. Les résultats se sont montrés encourageants car ce dispositif, testé en direct a reçu des retours positifs, démontrant à nouveau qu'une interaction sans contact en chirurgie était possible et bien accueillis. Là encore divers travaux, comme ceux de Cornelio et al. [8], présentant des compléments méthodologiques ont continués d'apparaître régulièrement.

En résumé, comme le notent Mewes et al. [21] dans son tour d'horizon « *Avant 2013, les approches pour concevoir interface utilisateur naturelle dans le domaine médical en utilisant des*

caméras RVB, des caméras Stéréo, une caméra temps de vol, des capteurs inertiels mettable ou la reconnaissance vocale étaient très diversifiées. [...] La sortie de la console grand public et de la caméra de jeux pour PC Microsoft KinectTM 1 en 2010 et du SDK Kinect en 2012 a évidemment eu un impact important sur la communauté des interactions homme-machine. Avec ce capteur de profondeur, il a été facile de créer des interfaces sans contact à un prix abordable. [...] Par rapport aux années précédentes, on constate une tendance à la hausse du nombre de publications et donc de l'intérêt pour le domaine de la recherche. ».

Au fur et à mesure que l'intérêt pour l'interaction sans contact en chirurgie augmentait, le besoin d'une méthodologie claire pour concevoir de telles interfaces est devenu plus évident. O'Hara et al. [28] ont apporté une contribution significative à ce sujet en soulignant les enjeux et les défis posés par une telle conception de l'interaction dans ce contexte. Tout d'abord, l'une des préoccupations de ce travail est la notion de "richesse de l'expressivité (expressive richness)", la capacité de mapper un grand nombre de gestes avec un grand nombre de fonctionnalités. Deuxièmement, la question de l'utilisation de gestes à une ou deux mains et de leur impact sur la convivialité de l'interface se pose dans cette étude. Troisièmement, les préoccupations relatives aux "gestes mondains (mundane gestures)", gestes produits sans intention d'interaction, et les interférences entre les transitions gestuelles sont également mises en lumière. Enfin, la question des gestes d'engagement et de désengagement est également remise en question, car il s'agit d'un moyen d'éviter les gestes involontaires en dehors de toute phase d'interaction et d'assurer le suivi de l'utilisateur actif dans un environnement multi-utilisateurs. Pour étudier et relever ces défis, O'Hara et al. ont conçu un système utilisant un capteur Microsoft KinectTM, exploitant ses capacités de détection pour assurer cette "richesse de l'expressivité" du dit système. Ce système interactif mis en place reposait sur des gestes bi-manuels ou mono-manuels, selon l'interaction. Il exploitait également la commande vocale comme moyen d'aborder la question de l'engagement/désengagement et utilise une fonction de verrouillage basée sur le temps dans laquelle il est nécessaire de maintenir une posture des mains pendant une durée fixe pour activer une fonctionnalité spécifique. De ce travail, nous remarquons que, premièrement, la plupart des préoccupations dépendent, au moins en partie, du capteur; et, deuxièmement, l'ensemble des gestes qui contrôlent ce système, même s'ils sont basés sur des entretiens avec les utilisateurs finaux, sont apparemment choisis directement par les développeurs.

1.4 La Conception Centrée Utilisateur & l'UX

Dans cette thèse, et comme nous le verrons dans les prochains chapitres, nous nous sommes intéressés à la question du "ressenti" de l'utilisateur. Cela est dû au fait, qu'il est accepté, que cette composante occupe, de toute évidence, une place très importante quant à l'acceptation et l'appréciation de l'utilisation d'un système d'interaction ce qui se répercutera à son tour sur les performances lors de son usage. C'est pourquoi, l'évaluation et l'optimisation du ressenti, a occupé et occupe une place grandissante au sein de la recherche et de la conception d'interaction. Pour répondre à cet état de fait, sont nées les méthodes dites de "Conception Centrée

Utilisateur ” et ensuite plus spécifiquement de “ UX ”. Plus loin dans nos travaux (chapitres 2 et 3), la recherche et l’exploration vis-a-vis de ces méthodes nous sera utile pour intégrer des aspects clés de notre conception, notamment la composante “ d’intuitivité ”.

1.4.1 Qu’est-ce que l’UX ?

Dans le domaine de l’Interaction Homme-Machine (IHM), la nécessité et la pertinence d’une approche intégrant l’utilisateur aux processus de conception pour mieux répondre au besoin de celui-ci n’est plus à démontrer. C’est dans cette optique qu’a été développé la Conception Centrée Utilisateur (CCU), dont le processus a été adopté et normalisé (en 1999, ISO 13407, puis révisé en 2010, ISO 9241-210).

Les CCUs prennent pour principe une compréhension forte à la fois des utilisateurs, des tâches ou objectifs à accomplir et de l’environnement d’interaction. Les utilisateurs sont sollicités tout le long d’un processus itératif alternant entre phases de conception et phases d’évaluation centrées utilisateur. La notion d’*usabilité* occupe une place centrale au sein des CCUs ; conceptuellement il s’agit du degré auquel un produit peut être utilisé dans un contexte spécifique pour permettre à un utilisateur, au profil type défini, d’accomplir une tâche ou un objectif de façon efficace et performante.

L’UX (User Experience) est une approche qui trouve ses fondements dans l’approche des CCUs. Là où l’usabilité, mise au centre des CCUs, se concentre sur des aspects objectifs et quantifiable de l’interaction, l’UX complète cette approche en s’intéressant également aux aspects du vécu humain offert par une interaction. Autrement dit, au sein de l’UX, l’usabilité n’est plus le cœur de l’approche mais un simple aspect pragmatique. En réalité l’UX s’étend à tous les aspects qui influent ou découlent de l’interaction offerte par le produit concerné. Le ressenti offert par le produit, la compréhension de ce dernier, la satisfaction éprouvée lors de son utilisation sont autant de facteurs à considérer. En somme l’UX se focalise sur toutes les émotions positives que le produit peut optimalement offrir ainsi que celles qui peuvent améliorer son utilisation.

1.4.2 La recherche et l’UX

Bien que l’UX soit un concept reconnu et utilisé depuis les années 2000, et bien qu’il bénéficie d’une bonne appréciation et reconnaissance générale, il reste à ce jour un domaine de recherche encore peu exploré. Cela peut être expliqué notamment par le fait que le concept ne bénéficie pas de consensus reconnu quant à sa méthodologie ou au traitement des résultats produits. Cela est d’autant plus notable sachant que jusqu’à aujourd’hui le domaine de l’UX est encore le théâtre du débat “ quantitatif contre qualitatif ” que nous nous allons discuter maintenant.

Comme évoqué précédemment, l’UX est une approche se concentrant principalement sur des aspects positifs hédoniques de l’interaction dit “ *non-instrumentals* ”, comme les émotions

positives, l'esthétique visuelle, le confort, le bien-être, etc. Mais de par son héritage, l'UX repose aussi sur des qualités " *instrumentales* " mesurables. Ce dernier état de fait est à l'origine de la controverse au sein du domaine : l'UX peut-elle et doit-elle muter en une approche rigoureuse basée sur des mesures strictes ? Ainsi deux points de vue s'opposent : d'une part certains considèrent qu'une approche de l'UX basée sur des mesures strictes est ce qu'il manque à l'UX pour lui permettre d'avancer et de s'imposer comme méthodologie forte et incontournable de conception IHM moderne, alors que, d'autre part, beaucoup considèrent que l'UX est trop hédonique de nature pour permettre à une vision quantitative de diriger tout type de design UX. À l'heure actuelle c'est bien ce dernier point de vue, celui de l'approche quantitative qui domine le terrain de l'UX. [19][4]

1.4.3 Comment faire de l'UX ?

L'UX étant une approche de conception héritée des approches CCUs, elle en reprend donc les lignes directrices du processus de conception. Globalement, il s'agit d'un processus itératif basé sur un degré élevé de compréhension de l'utilisateur, de son environnement et du contexte d'utilisation mis à l'épreuve au travers de phases d'évaluation centrées sur les utilisateurs. Le processus de conception se découpe en 4 phases : les phases d'exploration, d'idéation, de génération et d'évaluation. Le processus implique une transition séquentielle stricte mais avec possibilité, lors de n'importe quelle phase du processus de retourner à une phase antérieure (voir figure 1.1).

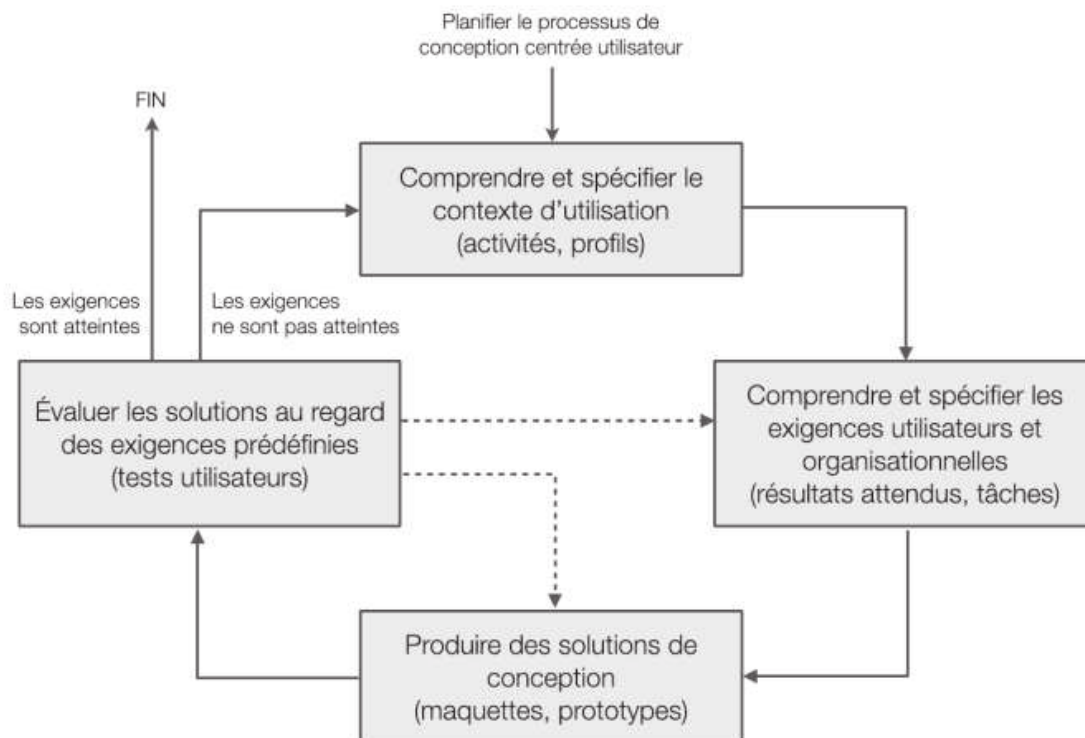


FIGURE 1.1 – Processus itératif de conception des CCUs, tiré de [13]

Ainsi l'approche UX se distingue de l'approche "classique" du processus des CCUs non pas par une refonte structurelle de celui-ci, mais par l'usage de méthode qualitative hédonique et fortement empathique des utilisateurs. Concevoir selon l'UX implique avant tout une compréhension de comment l'expérience utilisateur par rapport à un produit donné se manifeste et évolue dans le temps. Au fil du temps plusieurs modèles théoriques modélisant les interactions clés des approches théoriques ont émergé, de succédant, chacun apportant de nouvelles dimensions clés à l'appropriation de l'UX. Aujourd'hui les modèles les plus récents (cf. modèle de Karapanos, voir figure 1.2) modélisent l'UX par une approche multidimensionnelle intégrant l'interaction entre l'utilisateur, le système/produit, le contexte, et, intégrée plus récemment, la temporalité (avant, pendant et après expérience).

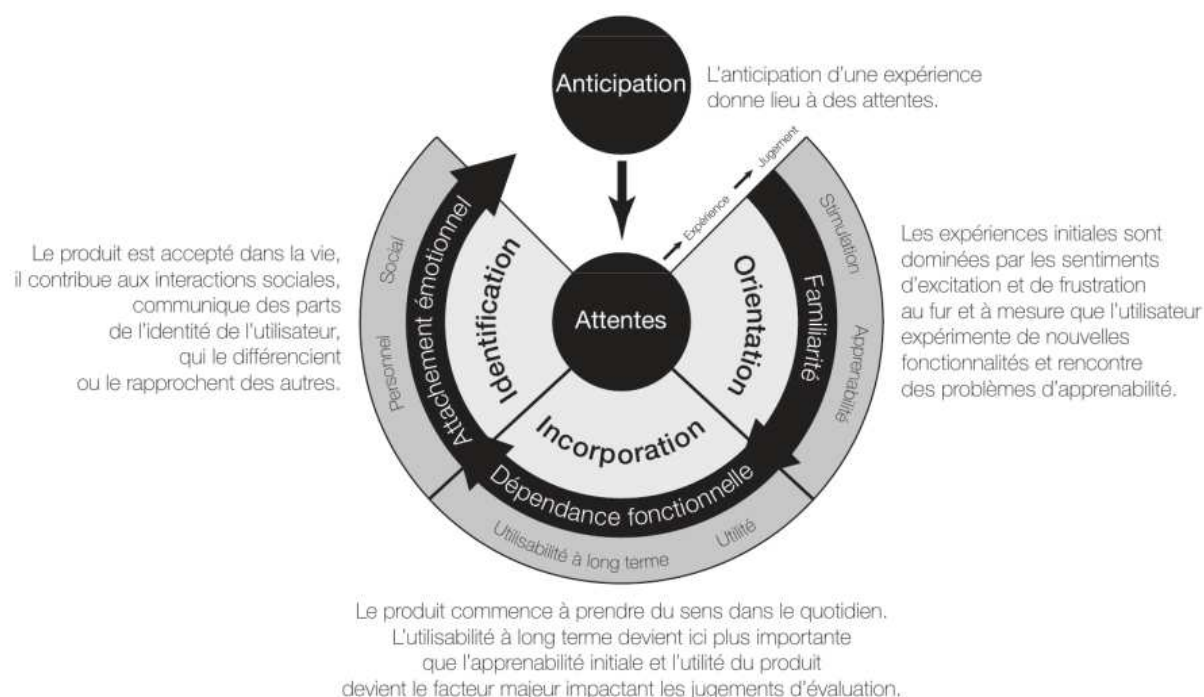


FIGURE 1.2 – Modèle de temporalité de l'expérience (traduit et adapté de Karapanos), tiré de [13]

1.4.4 Se préparer pour l'UX

Bien que le processus de design consiste à transiter de manière itérative entre des phases de conception (exploration, idéation, génération et évaluation), il est une étape antérieure, n'entrant pas dans le processus itératif : la phase de préparation. Cette étape est essentielle car c'est elle qui définit les grandes lignes du projet, définit les contextes et environnements de l'expérience, cible les utilisateurs et encadre le déroulement théorique de toutes les autres phases de conception. Posé de façon plus terre à terre il s'agit de préparer le terrain sur lequel va s'enraciner tout le reste du projet.

Dans la pratique, la démarche UX affecte aussi pleinement l'étape de planification. Il s'agit dans un premier temps de bien définir le projet, en adéquation avec l'approche UX, en organi-

sant une ou plusieurs rencontres préliminaires au projet entre tous les acteurs du projet dans le but d'établir, dans une optique de conception partagée, le contexte, les objectifs, les besoins et les cibles du projet. Dans un second temps, la préparation du projet UX passe par une étape de recherche secondaire portée aussi bien sur des études de marchés, que des revues de littérature scientifique. Enfin un des points clé de la préparation consiste à préparer le terrain pour la recherche utilisateur.

1.5 Approche et conception

Avant les récents aboutissements technologiques, notamment dans les années 90, une partie de la recherche est consacrée à améliorer la détection des gestes, et donc, en ce qui nous concerne, les mains. Au-delà d'aspects purement technologiques liés aux capteurs, ce processus nécessitait alors de mener des études sur la nature des gestes et des appendices concernés. C'est notamment le cas de Quek [30] dont les travaux visant à élaborer une reconnaissance efficace des gestes des mains, passaient aussi bien par une étude anthropologique des gestes que par des analyses et décompositions de leur réalisation. Dès lors, Quek décompose les gestes de la main via une taxonomie (voir fig. 1.3) les divisant en deux catégories : les actes et les symboles.

Les gestes symboliques sont basés sur des mouvements qui exigent un certain niveau de connaissance pour être interprétés, étant alors considérés comme faisant référence à un objet ou à un concept ; par exemple mimer une arme à feu sur la tempe pour indiquer qu'on a fait une erreur. Ces gestes peuvent aussi être modalisant ; par exemple, porter la main à un certain niveau pour indiquer la taille d'une personne. La deuxième catégorie, les gestes d'actes se décomposent en gestes mimétiques et gestes déictiques. Respectivement ils imitent ou miment une tâche à accomplir, et découlent du pointage. Ces gestes d'actes peuvent être décomposés en gestes spécifiques, génériques et métonymiques qui sont chacun physiquement exécutés de la même façon, mais interprétés différemment en fonction du contexte.

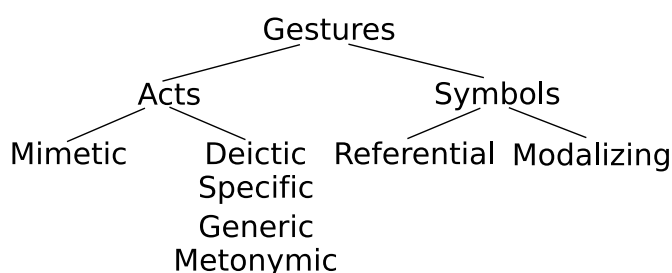


FIGURE 1.3 – Taxonomie des gestes de Quek

Plus tard, Pavlovic et al. [29] reprendront cette taxonomie et l'étendront pour décomposer la segmentation des gestes en un ensemble de règles gouvernant les interfaces gestuelles à mains nues :

1. L'intervalle de temps d'exécution des gestes se compose de trois phases : préparation, course et rétraction.
2. La pose de la main pendant la course suit une trajectoire classifiable dans un espace des paramètres.
3. Les gestes sont confinés à un volume spatial défini.
4. Les mouvements répétitifs des mains sont à considérer comme des gestes
5. Les gestes de manipulation ont un intervalle de temps d'exécution plus long que les gestes de communication.

Quand il s'agit de conception d'interfaces se pose la question des interactions standardisées contre les interactions personnalisées. Norman et Donald [27] ont montré la limite des interactions gestuelles standardisées en soulignant que les conventions ne peuvent pas prendre en compte la dimension inter-individuelle ou les différences culturelles (par exemple, agiter la main signifie « bonjour » dans la culture occidentale, mais signifie un rejet ou une désapprobation pour les utilisateurs indiens) :

« Most gestures are neither natural nor easy to learn or remember. Few are innate or readily predisposed to rapid and easy learning. Even the simple head-shake is puzzling when cultures intermix. »

Comme Nielsen et al. [26] l'ont souligné, le développement d'un système interactif basé sur le geste n'est ni trivial, ni négligeable. Les auteurs ont fait remarquer, et beaucoup d'autres par la suite, par exemple Jegou [15], que le vocabulaire gestuel (c'est-à-dire l'ensemble des gestes auxquels sont associés les actions disponibles du système) ne peut être considéré comme universel et est, en fait, propre au contexte. Concernant les interactions " adaptatives ", Jegou suggère que les vocabulaires gestuels standardisés et prédéfinis pour une utilisation inter-applications ne sont pas optimaux puisque la signification des gestes et expressions gestuelles est, d'après lui, sensible au contexte en arguant que « *une interface standardisée risque de rendre l'interaction non accessible de prime abord, car décontextualisée.* » Il fait également remarquer que le vocabulaire gestuel (c'est-à-dire l'ensemble des gestes auxquels sont associés les actions disponibles du système) ne peut être considéré comme universel et est, en fait, propre au contexte. Il est important de noter cependant que ces travaux portent sur des interactions gestuelles portant sur les domaines de " la thérapie à domicile par la réalité virtuelle et les arts scéniques numériques " qui sont des domaines à la fois hédoniques et relevant du ressenti ou de l'expression personnelle.

Revenant à un point de vue plus générique, le travail de Bobillier-Chaumon et al. [7] offre un point de vue multi-disciplinaire autour de la question de la standardisation des IHM. Dans leurs travaux, ils notent comment les références culturelles communes ont rendu les interactions standardisées accessibles, faciles à mettre en œuvre et réutilisables dans diverses appli-

cations et cas d'utilisation. À l'opposé, se trouvent les interactions personnalisées ou " adaptatives " (à un utilisateur spécifique), qui offrent leurs avantages propres : une flexibilité et une souplesse de l'interface pour un confort et une efficacité bien adaptée à l'utilisateur. Cependant, ces avantages sont contrebalancés par un isolement de l'utilisateur dans la tâche et un risque de contre-productivité en cas de mauvaise adaptation du système.

Concernant à nouveau les interactions standardisées, les inconvénients que Bobillier-Chaumon et al. énumèrent, à savoir une décontextualisation vis-à-vis de la tâche, la rigidité de l'interface, et la non prise en compte des différences inter-individuelles, semblent amoindris en considération de notre contexte d'utilisation (le bloc opératoire). En effet, dans ce cas, les effets bénéfiques de la standardisation, sont notamment la simplicité de la mise en œuvre et la " sûreté " et le contrôle du processus d'interaction.

Ainsi, ces travaux nous ont orienté vers une approche standardisée pour les utilisateurs du groupe cible, à savoir les chirurgiens, mais prenant en compte les spécificités de ce groupe. Cela va a contrario de travaux, par exemple ceux de Jégo [15], portant sur l'adaptation des interactions individuelles. Ce choix est motivé par les avantages offerts par la standardisation de l'interface, qui nous semblaient plus pertinents vis-à-vis de notre contexte d'utilisation spécifique et que l'on considère comme " forts " pour les utilisateurs cibles (et donc pour leur contextualisation). Il s'agit donc d'une adaptation à granularité non pas au niveau de l'individu, mais du groupe qu'il compose en fonction du contexte d'utilisation.

Exemples d'approches vis-a-vis du contexte opératoire

Strickland et al. [37] emploient et recommandent une méthode de conception pour le bloc opératoire suivant ces lignes directrices :

1. Le système doit permettre le contrôle sans contact d'un ordinateur PACS, le système informatique le plus susceptible d'influencer la prise de décision peropératoire.
2. Pour ignorer les commandes involontaires, le système de reconnaissance gestuelle doit être inactif jusqu'à ce qu'une action distinctive l'interpelle. Le système doit être verrouillé à l'aide d'une autre action distinctive.
3. Le système doit utiliser les informations provenant des membres supérieurs et du torse de l'opérateur pour mettre en œuvre les fonctionnalités de base d'un appareil semblable à une souris.
4. La reconnaissance des gestes doit être robuste et fiable.
5. L'interface utilisateur doit nécessiter un minimum d'équipement et tenir compte de la fatigue et de la non-intentionnalité de l'utilisateur tout en optimisant l'intuitivité, l'interaction en temps réel et la facilité d'apprentissage.
6. Tous les gestes doivent respecter les contraintes imposées par les règles de la salle d'opération pour la technique aseptique et le travail en étroite collaboration avec les assistants.
7. Le système devrait être facile à intégrer dans les salles d'opération existantes avec un minimum de distraction, de formation ou de ressources humaines.

Néanmoins, et de façon similaire aux travaux de Ebert et al. [10], qui conçoivent également une interface gestuelle à mains nues pour consultation d'images radiologiques, la conception du vocabulaire gestuel est réalisée sans participation ou consultation des utilisateurs concernés. A la lumière des travaux cités tantôt, nous serons quant à nous incités à centrer notre approche autour du vocabulaire gestuel centré autour des utilisateurs.

Vaquero et al. [39] abordent la question de la visualisation des données médicales, énumèrent et analysent des techniques actuelles de visualisation et les systèmes d'interactions qui leur sont associés pour proposer ensuite des stratégies pour répondre aux besoins actuels. Cet article rappelle que la visualisation des données scientifiques est complexe du fait de la variété des sources, de la haute dimensionnalité des données ou encore de la grande quantité de données à visualiser sur plusieurs échelles. Il déplore que la considération envers les approches centrées utilisateurs soit insuffisante alors que celle-ci faciliterait considérablement, in fine, la compréhension de larges jeux de données par les utilisateurs. En vue d'améliorer l'expérience utilisateur, l'article préconise de fortes considérations sur les facteurs de perceptions (théorie des couleurs, perception des formes, évaluation de combinaison des éléments visuels, etc.), la pertinence des représentations et métaphores utilisées ou encore une évaluation fréquente du système durant la phase de développement. Un défi considérable serait également de pouvoir donner du sens aux données tout en réduisant le temps d'apprentissage en offrant par exemple une intégration centralisée des outils appuyés par une interopérabilité des données et une navigation suivie. Mais le plus grand défi pointé par cet article serait d'intégrer l'aspect multi-échelle des données, notant que l'interprétation des données scientifiques peut exiger de recouper des données allant à la fois de l'ordre du microscopique jusqu'au macroscopique, tout autant que l'aspect multi-scalaire apporté par les natures très différentes que peuvent prendre les données. L'article énumère plusieurs techniques pouvant être utilisées pour des telles visualisations tout en recommandant d'être particulièrement méticuleux quant à leur choix en fonction de certains facteurs sur la nature des données (type de données, visualisation usuelle, échelle, interactions usuelles). Enfin l'article, rappelant les différentes catégorisations possibles des IHM, présente les tendances actuelles et modernes des IHM évoluées ("*Intelligent HCI, Adaptive HCI, Ubiquitous computing, Ambient Intelligence* ") et propose des grandes lignes stratégiques (incorporer l'interface au sein même de la représentation des données, modification automatique des modes de navigation en fonction des échelles de visualisation, considération des possibilités offertes par les perceptions 2D ou 3D, etc.). Pour finir, cet article insiste sur la nécessité de disposer de connaissances et ressources multidisciplinaires tout en rappelant que l'unique et premier but est la compréhension constante et à tout instant de l'interface et de ses fonctions par l'utilisateur.

1.5.1 Conclusion et orientation

Au regard de cet état de l'art, nous avons décidé d'orienter notre démarche comme suit :

- De par les objectifs de la thèse et les contraintes imposées par le contexte, nous allons utiliser un système de détection des gestes basé sur la vision. Plus spécifiquement, pour des raisons exposées dans le chapitre 2 à venir (2.2.2.3) nous ferons le choix de porter

notre développement autour du capteur LeapMotion™.

- Comme déjà mentionné tantôt, nous avons décidé d'adopter une approche de conception d'interactions standardisées pour les utilisateurs cibles. En effet, le contrôle sur le processus d'interaction, l'accessibilité, et la sûreté à moyen terme, apportés par la stabilité et l'invariabilité des interactions standardisées nous semblent le plus adapté pour remplir nos objectifs étant donné le contexte du bloc opératoire. De plus, nous anticipons que l'usage des logiciels PACS et DICOM contribuent à créer des expériences interactives et des attentes similaires chez les chirurgiens, renforcé par une culture professionnelle et un contexte d'utilisation partagés.
- Néanmoins sensibles aux avantages des interactions adaptées et centrées autour des utilisateurs, nous adoptons une méthode “ *d'incitation aux gestes* ” en vue de construire un vocabulaire gestuel basé sur l'intuition des utilisateurs cibles.

Notre méthode “ *d'éllicitation de gestes* ” (“ *gesture elicitation* ”), qui sera présentée au cours des deux prochains chapitres, appartient à un type de méthodologie ayant gagné du terrain dans divers domaines de la recherche interactive au cours des dernières années. Paru en 2018 (cette thèse ayant été lancée en 2015), l'enquête de Vogiatzidakis et al. [40] nous a permis de constater que le bloc opératoire était encore sous représenté parmi les domaines d'application de ces méthodes, appuyant ainsi la légitimité de notre approche.

MÉTHODOLOGIE ET DÉMARCHE

Ce chapitre a pour but de décrire et de rendre compte de notre démarche méthodologique ainsi que de relater, dans une certaine mesure, de son évolution au fil de l'avancement de nos travaux. Pour ce faire nous commençons par y présenter et détailler l'analyse contextuelle de notre problème initial via l'étude d'un cas d'usage particulier. Nous prenons en compte aussi bien l'environnement d'utilisation (la salle d'opération) et le contexte logiciel. Forts de cette analyse qui nous permet alors de plus clairement contextualiser notre interaction, nous sommes alors en mesure d'identifier les contraintes d'utilisation et environnementale. De même, cela nous permet dès lors d'isoler les tâches clés auxquelles l'interaction sans-contact devra être en mesure de proposer de nouvelles interactions se substituant aux interactions classiques. Cet ensemble de nouvelles interactions, matérialisées de fait par un ensemble de gestes, constitue alors le cœur de cible de notre recherche, à savoir le vocabulaire gestuel le mieux adapté à notre contexte d'utilisation. De cette optique, nous établissons alors, en fonction de nos observations du cas d'usages, de notions de la littérature et de nos choix techniques, un ensemble de critères, répartis en trois catégories, qui servira de boussole dans notre recherche de vocabulaire gestuel. Nous nous servons alors de cet outil comme appui pour concevoir un premier vocabulaire gestuel pour répondre aux besoins identifiés tantôt.

Le vocabulaire gestuel conceptuel, élaboré alors, fait encore défaut d'une composante de " naturalité " ou " d'intuitivité ". En d'autres termes la dimension " utilisateur " n'est alors que peu prise en compte. Face à cet état de fait, nous nous tournons vers des méthodes dites " *centrées utilisateurs* " et plus spécifiquement, des méthodes dites " *d'éllicitation* ", en l'occurrence, d'éllicitation de gestes. Nous prenons alors en référence un type de méthodes nommées « *Magicien d'Oz* » (« Wizard of Oz »). Nous intégrons alors cette méthodologie au sein d'un protocole utilisateur scindé en deux phases, chacune portée sur une population différente. La première phase est portée sur des utilisateurs génériques, dit « tout-venant » à partir de laquelle nous extrayons des gestes issus de leurs intuitions interactives. À partir de ce corpus de gestes " d'intuition ", nous raffinons notre vocabulaire conceptuel élaboré précédemment pour y intégrer, dans la mesure du respect des critères établis précédemment, la composante " intuitive " manquante. Dès lors, la deuxième phase, portant elle sur une population d'utilisateurs " experts ", nous sert conceptuellement à réitérer ce raffinement de vocabulaire et concrètement à valider le vocabulaire raffiné lors de la première phase.

2.1 Observation de cas d'usage

Dans cette section nous présentons le cas d'usage étudié à partir duquel nous basons notre conception de système d'interaction gestuelle. Nous y présentons ensuite notre analyse de ce cas d'usage : les contraintes d'environnement et de contexte ainsi que les interactions cibles identifiées. Nous explicitons alors comment cette analyse a influencé notre choix de travailler avec le capteur LeapMotion™.

2.1.1 Environnement d'utilisation

Intégrant le projet *3D-Surg*, nous avons été mis en contact avec certains chirurgiens et “fellows” chirurgiens partenaires. Ces contacts nous ont alors présentés, à leur contexte d'utilisation issue de leur spécialité. Ainsi donc, circonstancielllement au projet, le cas d'usage pratique particulier pris en référence par notre travail est une opération de laparoscopie ^[W] (ou cœlioscopie). Telle qu'observée, celle-ci fut réalisée par une équipe de cinq personnes. La figure 2.1 présente un schéma d'organisation de l'opération telle qu'observée : L'équipe se compose comme suit :

- chirurgien
- premier assistant
- second assistant
- instrumentiste/infirmier de bloc
- anesthésiste

Durant l'opération observée, 3 écrans étaient en utilisation :

- Le moniteur de visualisation interne qui affiche les images récoltées par la caméra optique introduite dans la cavité abdominale du patient. Cet affichage est essentiel à l'opération et ne dépend pas d'un système informatisé nécessitant une interaction particulière. Cet écran est monté sur un bras articulé permettant de placer l'écran en fonction de l'opération réalisée ; ici il est placé face au patient, surélevé pour être niveau du chirurgien, au-dessus des jambes du patient.
- Le moniteur de télémétrie qui affiche les informations du système de surveillance télémétrique du patient permettant de surveiller les constantes de celui-ci (rythme cardiaque, respiration, etc.). Se situant derrière le patient et séparé de celui-ci et de la table d'opération par un drap matérialisant une barrière de stérilité. La surveillance constante se fait de façon continue par un membre de l'équipe opératoire non-stérilisé. L'environnement étant non stérile et la fonction de l'écran et du système associé étant uniquement observationnelle, il n'y a, a priori aucune raison d'intervenir sur l'interaction proposée par ce dispositif.
- Le moniteur de consultation des images acquises par scans tomographiques (CT-scans). Il permet de naviguer dans le logiciel de consultation des CT-scans, déjà préparés en pré-op, pour une seconde consultation. L'interface de ce système étant l'interface classique clavier-souris elle constitue un environnement non stérile. Ce dispositif constitue donc un problème plusieurs fois soulevé dans la littérature inhérente aux Interactions

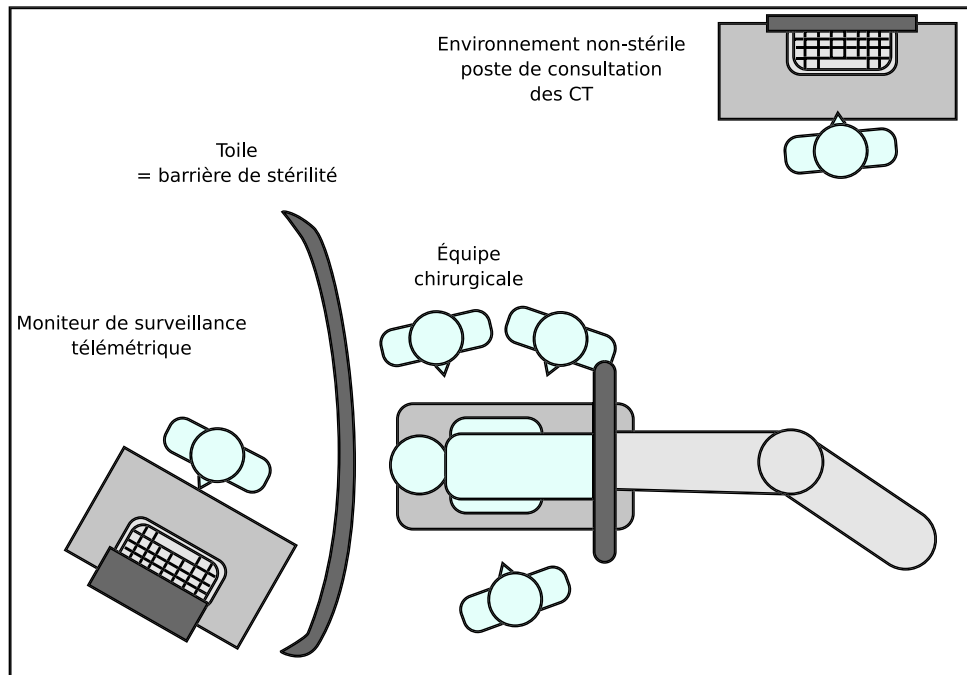


FIGURE 2.1 – Schéma de la salle d'opération

Homme-Machine appliquées au domaine médical (e.g.), du fait que toute utilisation par un membre de l'équipe opératoire, entraîne une dé-stérilisation pénalisant la personne et pouvant potentiellement encombrer l'opération. Ainsi, ce dispositif est donc la cible évidente pour une introduction à de nouveaux modes d'interaction garantissant la stérilité tels que des techniques d'interaction sans contact.

2.1.2 Logiciel & interactions

De la même façon que pour la salle d'opération, des membres de l'opération chirurgicale nous ont présenté le logiciel d'interaction employé au cours de l'opération. Nous avons ainsi pu analyser le contexte d'utilisation, du point de vue de l'environnement logiciel.

Note : *Durant cette présentation, nous ont été présentés deux cas d'usage : celui que nous nommons " 2D " et celui que nous nommons " 3D ". Dans les sections qui suivent nous concentrons notre démarche de conception sur le premier cas d'usage. La raison est que le cas d'usage " 3D " était principalement utilisé en phase de planification d'opération ce qui a eu pour effet de baisser la priorité de traitement de ce cas d'usage. De fil en aiguille le cas d'usage " 3D " est passé au second plan, bien que notre démarche fut adaptée et pensée avec la possibilité de l'appliquer à ce cas précis.*

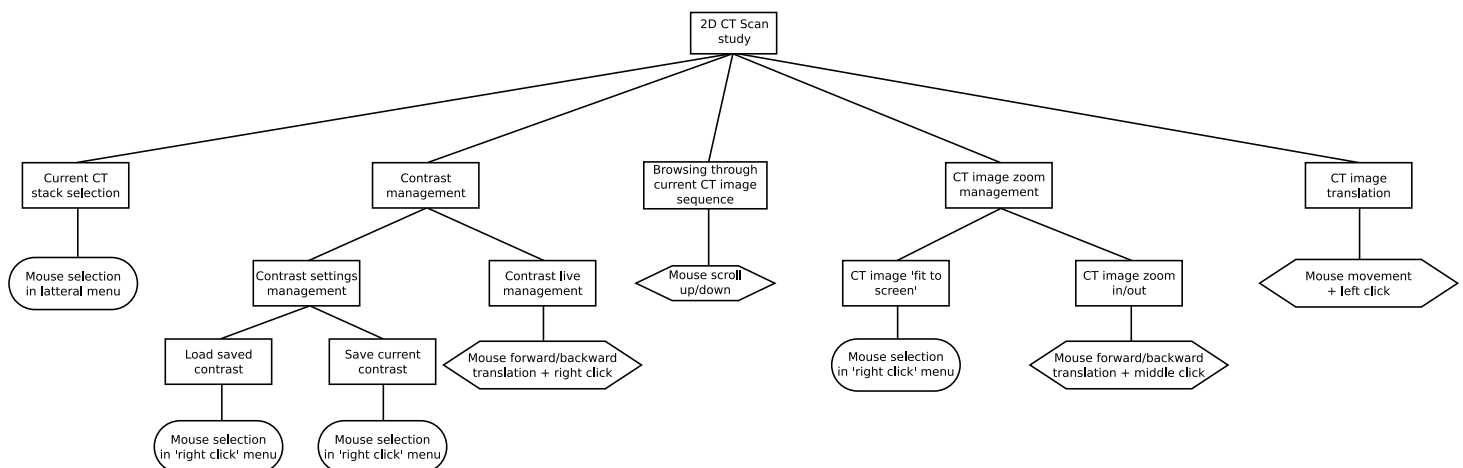
2.1.2.1 Cas d'usage " 2D " :

Le logiciel utilisé dans le bloc opératoire permet au chirurgien (en pré-op comme en per-op) de naviguer image par image dans une ou plusieurs piles de CT-scan à la recherche de ce que le clinicien évaluera être des zones d'intérêt (" *Region of Interest* (*ROI*) "). Parmi les interactions proposées pour des piles d'images tomographiques, celles utilisées dans le contexte

d'une laparoscopie (pré-op ou per-op) sont les suivantes :

Changer la pile courante	Sélection avec la souris, de la pile à explorer, dans le menu latéral.
Naviguer dans la pile d'images et descendant/remontant la pile	Molette de la souris vers l'avant / l'arrière
Zoom avant / arrière sur l'image courante de la pile	Maintien du clic molette de la souris + mouvement en avant/arrière de la souris
Translation sur l'image courante de la pile	Maintien t du clic gauche de la souris + mouvement de la souris
Recentrage et dézoom de l'image courante de la pile	Navigation dans le " menu clic droit " sur l'image
Augmenter/diminuer le contraste sur les images affichées durant la navigation dans la pile	Maintien du clic droit de la souris + mouvement en avant/arrière de la souris
Enregistrer l'état courant du contraste ou charger un état de contraste sauvegardé	Navigation dans le " menu clic droit " sur l'image

À partir de cette liste d'interactions, on propose la taxonomie suivante (voir aussi A) :



NDLA : À noter que, ici ces taxonomies ne sont pas à confondre à celles mentionnées dans le chapitre précédent (1.5), qui fonctionnellement s'apparentaient à un découpage sémantique des gestes. Ici, il s'agit comme expliqué d'un découpage fonctionnel des tâches à remplir. En ce sens, ces taxonomies se rapprochent des résultats produit par la procédure de « Hierarchical Task Analysis (HTA) » [3] dans le sens où, comme ces dernières, elles reposent sur une analyse fonctionnelle résultant en une hiérarchie sous forme de diagramme de description des tâches. Néanmoins, les HTA ont pour but de générer un diagramme descriptif des tâches et de leurs mécanismes ayant pour but la compréhension en vue d'une optimisation et réorganisation éventuelle de l'organisation fonctionnelle

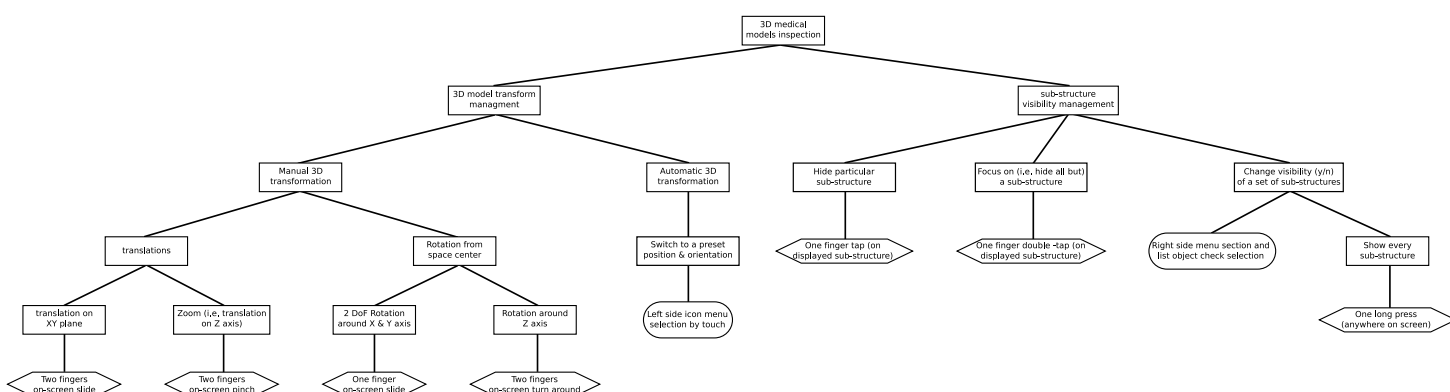
en fonction de l'interaction ou des besoins ciblées. Nos taxonomies quant à elle, ont pour but spécifique de réduire la description de nos fonctionnalités en un ensemble de sous-tâches élémentaires, qui seront donc les “ feuilles ” de nos diagrammes en forme d'arbres. Comme nous l'expliquerons (2.2.2), notre but sera alors de conserver la structure intégrale du diagramme pour conserver le découpage fonctionnel de l'application à laquelle l'on souhaitera substituer sa modalité d'interaction par une autre.

2.1.2.2 Cas d'usage “ 3D ” :

Ce logiciel, contrairement à celui présenté précédemment, est un logiciel de visualisation d'imagerie médicale en 3D dont les modèles 3D visualisés en question sont spécifiques au patient. Cette application est ainsi associée à un autre mode de visualisation permettant la détection, par des spécialistes, d'anomalies médicales. De plus, ce logiciel étant une application pour équipements mobiles tactiles (tablette, smartphone), les modes d'interactions sont sensiblement différents de ceux basés sur le modèle clavier/souris. En pratique, d'après le spécialiste consulté, l'équipement mobile, et donc le mode actuel d'interaction, ne constitue pas un enjeu majeur par rapport à la stérilité en bloc opératoire. En effet, la barrière de stérilité peut être maintenue en enveloppant le périphérique dans un sachet transparent stérile. Nous pouvons néanmoins pointer l'inélégance et le possible encombrement d'une telle solution. Il reste cependant intéressant de lister les interactions proposées par ce système :

Rotation de l'objet autour des axes X & Y de l'espace monde (2 DOF)	Glissement du doigt sur l'écran
Rotation de l'objet autour de l'axe Z de l'espace monde (1 DOF)	Glissement circulaire à 2 doigts sur l'écran
Translation plan XY de l'objet (2 DOF)	Glissement de 2 doigts sur l'écran
Zoom (équivaut à une translation sur l'axe Z) (1 DOF)	Pincement (2 doigts) sur l'écran
Repositionner et réorienter la structure dans une position orthogonale clé (e.g. : vue droite du sommet du modèle, vue de côté du modèle, etc.)	navigation par tap dans un menu latéral
Cacher une structure ou un organe	Tap sur l'organe ou structure affiché à l'écran
Focus sur un organe ou une structure (i.e. : cacher tout sauf la structure/organe en question)	Double tap sur l'organe ou structure affiché sur l'écran
Afficher tous les organes ou structures préalablement cachés	Long press sur l'écran
Afficher ou cacher un ensemble d'organes ou sous-structures	navigation par tap dans un menu latéral puis sélection/désélection de sous-structure à afficher/cacher

À partir de cette liste d'interactions, nous proposons la taxonomie suivante (voir aussi A) :



2.1.3 Analyse du cas d'usage

2.1.3.1 Conditions & contraintes d'environnement

Suite aux observations relatées dans la section 2.1, et appuyées par des interviews d'utilisateurs experts (chirurgiens et " fellow " chirurgiens), nous avons pu identifier des facteurs d'utilisation et d'environnement ayant un impact sur la conception d'un potentiel nouveau



FIGURE 2.2 – Illustration du “ carré de stérilité limitant les mouvements du chirurgien ”.

système à intégrer dans le bloc opératoire.

Premièrement, dans la figure 2.1, on voit que le système de visualisation d'images médicales en per-opératoire dispose d'un espace dédié. Bien que, naturellement, l'emplacement et l'espace dédié sont susceptibles de varier selon les blocs opératoires, il apparaît que l'environnement du bloc propose effectivement un tel espace dédié à cette utilisation bien que relativement isolé. Or, notre objectif est de pouvoir concevoir un système d'interaction s'intégrant dans le bloc qui ait un impact d'intégration minimale. En d'autres termes, son intégration, ou installation ne devrait ni modifier ni l'organisation de l'espace, ni modifier significativement les habitudes des chirurgiens. En effet, une majorité de chirurgiens interviewés ont spontanément émis la réflexion que le succès d'intégration d'un nouveau système d'interaction serait fortement influencé par la minimisation des changements subis par les chirurgiens, aussi bien terme d'organisation que d'habitude d'utilisation.

Deuxièmement, après interviews et observations, il apparaît que les chirurgiens sont limités dans leur mouvement et donc dans leurs gestes. En effet, dans le cas d'usage observé, les chirurgiens limitent leurs mouvements dans une fenêtre qu'on appellera ici “ carré de stérilité ” (voir figure 2.2) en guise de précaution supplémentaire de maintien de la stérilité. Le respect de ce carré de stérilité, impose de maintenir ses mains, pendant la durée de participation à l'opération, au-dessus de la taille et en dessous des épaules. De même il limite l'ouverture horizontale des bras à un champ restreint approximativement à celui du torse. Ainsi, il apparaît donc du fait de cette contrainte, que le système à concevoir ne doive considérer qu'une petite zone de détection face au chirurgien, de facto excluant les gestes et interactions basées sur des mouvements amples ou étendus des bras.

2.1.3.2 Choix du capteur

Au vu du contexte, de l'environnement d'utilisation et de l'analyse que nous en avons fait, pour notre démarche de conception nous nous sommes tournés vers le capteur LeapMotion™.

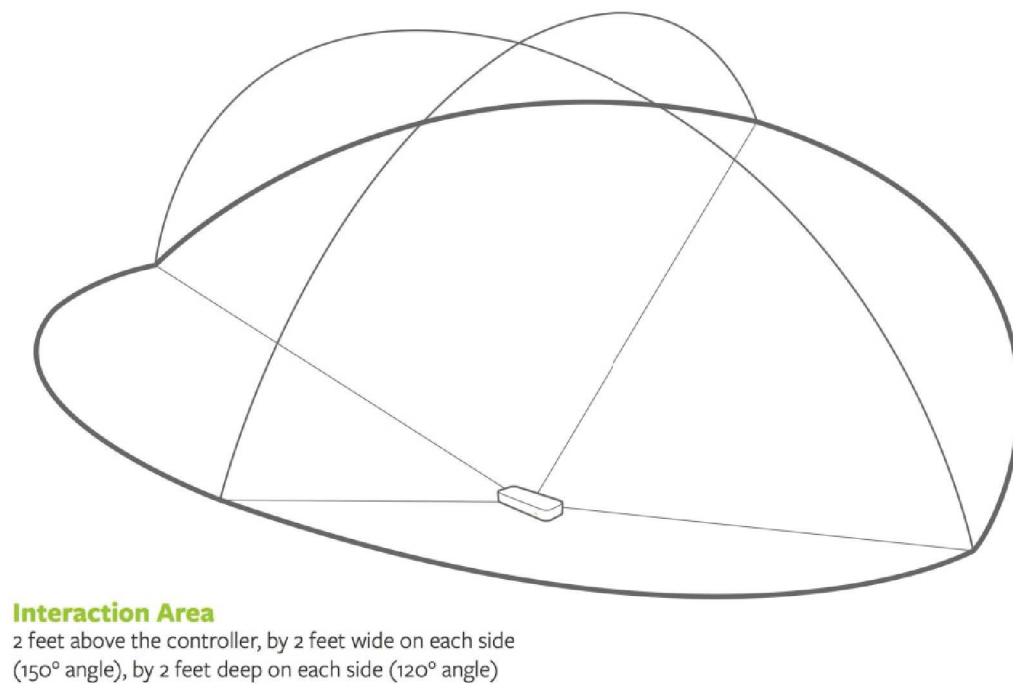


FIGURE 2.3 – Schéma de la zone de détection du capteur . Source : blog d'actu-
alités LeapMotion — <http://blog.leapmotion.com/hardware-to-software-how-does-the-leap-motion-controller-work/>

L'une des premières raisons à cela est la petite taille de ce capteur ($13 \times 30 \times 76$ mm) qui garantit un aménagement logistique minimum pour un impact d'intégration faible. Dans le même ordre d'idée, le champ de détection conique vertical relativement restreint (150° fov, ≈ 60 cm) limite l'espace à dédier à l'interaction et réduit les risques d'interférence avec d'autres personnes, ou éléments de l'environnement. De plus, le LeapMotion™ est compatible " par conception " à notre problème. Celui-ci est un capteur dédié au suivi et à la détection de la main et de ses postures. Or, comme expliqué dans la section 2.1.3.1, les chirurgiens sont grandement limités dans leurs gestes, par le " carré de stérilité ", ne laissant qu'un nombre limité d'options pour s'exprimer par le medium gestuel. Parmi ces options restantes, un grand nombre se trouve dans le spectre de gestes couverts et ciblés par le LeapMotion™ : des mouvements, de faible amplitude, des mains et des avant-bras.

2.2 Démarche exploratoire

Suite à l'analyse décrite dans la section précédente et avant de pouvoir être en mesure de concevoir un vocabulaire gestuel répondant aux contraintes et besoins identifiés, il faut d'une part pouvoir apprécier les rapports déjà existants entre les tâches isolées et des interactions proposées dans d'autres cas d'usages communs. Dans un second temps, nous avons travaillé à établir une série de critères auxquels soumettre tout candidat potentiel de geste pour les interactions explorées. Ces critères, tenant compte de plusieurs facteurs influençant la viabilité des gestes considérés, nous seront utiles pour cadrer notre recherche de vocabulaire et nous

permettre de proposer un vocabulaire gestuel adapté aux contraintes et besoins identifiés.

2.2.1 Exploration de gestes

Ici, nous proposons une observation comparative pour chaque tâche isolée : nous établissons un recueil des idiomes associés, dans d'autres cas d'utilisation, à ces interactions. À ces derniers se succèdent ceux utilisés effectivement dans notre cas d'usage lors de l'interaction classique (clavier/souris, telle qu'observée). Cela nous permet d'anticiper les attentes et influences possibles que porteront d'éventuels utilisateurs lors de la réalisation de ces tâches.

2.2.1.1 Translation

La translation, dans notre contexte d'étude, correspond à l'action de pouvoir déplacer un objet, en l'occurrence une image, contraint ou non, selon un axe déplacement (ou degré de liberté). Parmi les interactions isolées, il s'agit très certainement de la plus répandue. En effet, en plus d'être une action essentielle dans de nombreux logiciels à buts créatifs ou artistiques (CAO, DAO, création d'images, etc.) elle fait également souvent office de métaphore interactive dans beaucoup d'IHM; en effet, des fonctionnalités comme le Glisser-Déposer (“ *Drag’n Drop* ”) ou encore le Défilement (“ *Scroll* ”) peuvent être considérées comme des translations respectivement contraintes sur des axes de déplacement 2D et 1D. Ainsi, cette interaction est sujette à beaucoup de a priori et les “ intuitions ” qui lui sont associés sont largement influencées par l'état courant des IHM existantes.

L'un des idiomes interactifs les plus associés à la réalisation d'une translation sur un objet correspondant à la mimique du mouvement naturel pour effectuer cette action, c'est-à-dire « attraper puis déplacer ». C'est ce principe que l'on retrouve par exemple derrière la fonctionnalité de “ *Drag’n Drop* ” qui est un pilier des IHM graphique bidimensionnelles modernes. On retrouve d'ailleurs cette méthode dans les interfaces des périphériques à interfaces tactiles qui sont omniprésentes dans le paysage numérique de part la seule utilisation du smartphone (en 2018, en France ce périphérique équipe 3 personnes sur 4 [1]). Ainsi, cela permet de supposer que cette fonctionnalité est sujette à beaucoup de a priori qui influenceront sur la perception de l'intuitivité, par l'utilisateur, de toute interaction que l'on pourra lui associer.

Le cas d'usage 2D observé utilise une interaction similaire aux principes du “ *Drag’n Drop* ” : la mimique « d'attraper » l'objet est réalisée via l'appui et le maintien d'un clic gauche souris, l'on déplace alors l'objet saisi jusqu'à atteindre l'objectif ou résultat souhaité et l'on « dépose » en relâchant l'appui du clic gauche.

Le cas d'usage 3D tactile utilise également une sémantique quasi-similaire pour réaliser cette interaction : la mimique « d'attraper » l'objet y est réalisée en maintenant le contact (“ *press and hold* ”) de deux doigts sur l'écran, et la translation de l'objet s'effectue alors en glissant les deux doigts utilisés sur l'écran dans la direction souhaitée. Lorsque les deux doigts sont dissociés de l'écran (“ *release* ”), l'interaction s'arrête, ce qui équivaut à « déposer » l'ob-

jet. Il faut noter que la translation, dans ce cas, reste contrainte sur le plan parallèle à l'écran contenant le centre (défini) de l'objet.

On remarque que dans ces deux cas d'usages une différence au regard de la cohérence entre plan d'interaction et le plan d'action : dans le cas d'usage tactile tridimensionnel, le plan d'interaction (l'écran tactile) et d'action (plan virtuel de déplacement de l'objet lors de la translation) sont parallèles ; a contrario, le cas d'usage bidimensionnel à la souris requiert un plan d'interaction (le support de la souris) et un plan d'action (plan virtuel parallèle à l'écran) indépendants. Notre cas d'usage cible quant à lui, repose sur une interaction utilisateur qui n'est pas contrainte sur un plan, mais use de gestes devant être exprimés dans un espace 3D sans support.

À la lumière de cette analyse il apparaît que la relation entre l'espace d'interaction et l'espace des actions à réaliser présente un aspect inédit relativement aux deux cas d'usage. Les questions qui se présentent alors sont les suivantes : Doit-on contraindre (sans support physique) le mouvement à un plan ? Doit-on restreindre les degrés de liberté sur la translation ? Si l'on choisit de se baser sur l'interaction " *Drag'n Drop* ", quel geste pour modéliser les différentes composantes de cette interaction ?

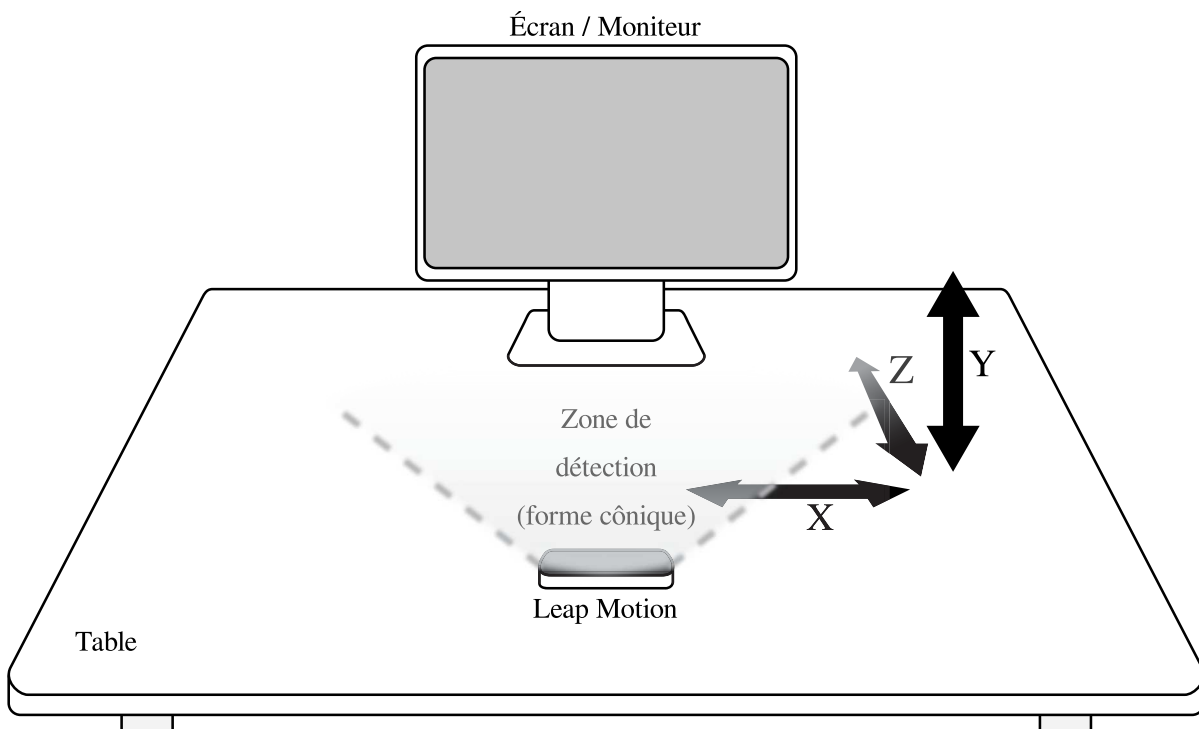


FIGURE 2.4 – Schéma d'un système d'interaction gestuel usant du LeapMotion™ ; nos 3 axes de références de l'espace de détection, X (gauche/droite), Y (haut/bas), Z (vers ou fuyant l'écran) y sont représentés.

2.2.1.2 Zoom

L'interaction de zoom (terme qui regroupe ici et dans la suite les notions de zoom (" *zoom-in* ") et dé-zoom (" *zoom-out* ")), dans ce contexte applicatif, consiste à agrandir la visualisation d'une zone d'intérêt d'une image. À l'instar de la translation, il s'agit d'une interaction

répandue et familière pour bon nombre d'utilisateurs. En effet l'on retrouve cette interaction dans des logiciels aussi bien grand public (visualiseurs d'images, documents) que dans des programmes spécialisés.

Là encore, c'est une interaction que l'on retrouve au quotidien sur des périphériques à interface tactile de type smartphone. Par conséquent cette interaction est elle aussi sujette à un idiome interactif très répandue qui associe cette fonctionnalité à geste spécifique : le pincement. Sémantiquement, le geste consiste à rapprocher deux doigts, comme si l'exécutant voulait pincer quelque chose, pour zoomer (" *zoom-in* "). Inversement le sens inverse du mouvement, un écartement des deux doigts induit une interaction inverse, donc un dé-zoom (" *zoom-out* "). Encore une fois, cela laisse à penser que cette interaction est sujette à beaucoup de a priori qui influenceront sur la perception de son intuitivité par de potentiels utilisateurs.

Le cas d'usage 2D observé l'interaction utilisée emploie le maintien de l'appui sur un bouton, en l'occurrence le bouton molette de la souris, associé au déplacement de la souris vers l'avant ou l'arrière (suivant le sens du zoom). L'on remarque que cette interaction est quelque peu non standard, là où habituellement un tel zoom est induit par un défilement molette, ou l'appui de bouton clavier tels que +/-.

Le cas d'usage 3D tactile quant à lui fait usage de l'idiome interactif mentionné tantôt, celui du pincement, se reposant sur l'usage commun.

Dans ces deux cas observés, bien que l'exécution soit différente, l'on remarque que le zoom est lié à un geste ou à un déplacement dont le sens peut être inversé pour influencer la direction du zoom. Dès lors se posent également les questions suivantes : Quels gestes employer dès lors que l'on se base sur des gestes sans supports physiques ? Un idiome commun tel que le pincement sont-ils compatibles avec cette modalité d'interaction ?

2.2.1.3 Navigation

Dans notre contexte applicatif, l'interaction dite de " Navigation " correspond à une transition d'état de visualisation sur les données images traitées par les applications qui constituent nos cas d'usage de référence. Cependant, nous isolons cette interaction en particulier, car elle essentielle au cas d'usage 2D, sur lequel nous nous sommes concentrés en priorité.

Le cas d'usage 2D observé l'interaction consiste à naviguer dans une pile d'images. En effet, ici les données sont une pile d'image dans laquelle le chirurgien navigue, une seule image, dite l'image courante, est affichée à la fois. Chaque image est en réalité une coupe d'image qui permet de naviguer dans une acquisition d'un organe par des techniques de type tomographie [W], coupe par coupe. L'interaction utilisée consiste simplement à utiliser le roulement molette de la souris (" *scroll* ") vers l'avant ou l'arrière pour naviguer respectivement vers le haut ou bas de la pile.

Le cas d'usage 3D tactile ne comporte pas de fonctionnalités similaires ou suffisamment proches.

Bien que les exemples offerts par nos deux cas de référence n’offrent pas de piste substantielle, le scroll utilisé pour le cas 2D, nous permettent d’associer à cette interaction la notions de défilement, proche de celle utilisée dans les usages numériques quotidiens (défilement de page web, par exemple). On note aussi, que pour les usages sur des périphériques à écran tactile, cette interaction est substituée par un balayage sur l’écran qui déclenche un défilement pouvant être arrêté par l’utilisateur par un simple touché. La gestuelle et le principe derrière ce dernier exemple seront également à considérer lors de la transposition de l’interaction de navigation à l’interaction gestuelle 3D.

2.2.1.4 Réglage de la luminosité et du contraste

Ici et dans la suite, nous regroupons ces deux interactions pour limiter la redondance étant donnée la nature très similaire de ces deux interactions. Ces interactions sont moins répandues dans les usages communs que les précédentes. Ces tâches consistant par essence à moduler, souvent manuellement, une valeur unidimensionnelle, les logiciels proposant ces interactions (DAO, graphisme, etc.) usant souvent de sliders qui consistent eux-mêmes à moduler une valeur en glissant un curseur sur une ligne représentant un intervalle sur lequel choisir cette valeur.

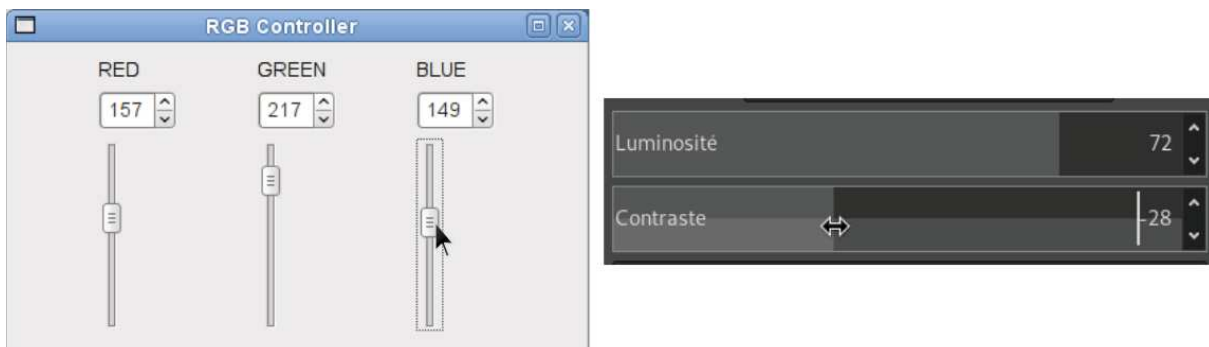


FIGURE 2.5 – Deux exemples de sliders : verticaux classiques à gauche (interface générée par Qt ^[W]); horizontaux modernes à droite (capture d’écran du logiciel The GIMP ^[W]).

Le cas d’usage 2D observé intervenir sur la valeur modulant les variables que sont le contraste ou la luminosité consiste en un mouvement de souris, activé par une pression d’un bouton, dont le sens de déplacement détermine la diminution ou l’augmentation de cette valeur. Ici on retrouve donc une modalité qui, par principe, fait penser à celle d’un slider purement virtuel.

2.2.2 Critères

Après avoir pu explorer les différents mécanismes et idiomes interactifs liant, aussi bien dans d’autres cas d’usage que dans ceux observés, les gestes ou certaines de leurs composantes, aux interactions isolées, nous sommes en mesure de mieux anticiper les attentes de potentiels

utilisateurs quant à la manière de réaliser ces dites interactions. Cela étant, avant de pouvoir conceptualiser un vocabulaire gestuel pour un système d'interaction sans contact ciblant notre cas d'usage, il faut déterminer un ensemble de critères permettant d'encadrer la recherche de gestes de sortes à ne considérer que des gestes qui soient adaptés à notre cas d'usage. Forts des analyses et observation réalisée tantôt, nous sommes en mesure de spécifier et lister de tels critères. Nous divisons ces critères en trois catégories distinctes.

2.2.2.1 Critères contextuels

Pour répondre au problème initial, au regard des cas d'usages qui nous ont été présentés ainsi qu'aux diverses contraintes techniques et environnementales rencontrées par les chirurgiens nous avons choisi de développer une réponse méthodologique qui aura pour objectif de satisfaire les critères contextuels qui suivent.

Interaction mono-manuelle L'objectif visé est d'offrir une interaction utilisant une seule main. Cela s'explique d'une part par le fait qu'un chirurgien, durant une opération, est amené à employer divers outils tels que le scalpel. Par conséquent il serait préférable de disposer d'un système interactif qui ne requiert pas que l'utilisateur soit obligé de poser ses outils, comme cela serait le cas pour un système bimanuel.

Main dominante/non dominante Ce critère est directement lié au précédent; en effet, étant donné que la manipulation potentielle d'outils tiers par le chirurgien est à considérer, il apparaît donc deux possibilités d'interaction mono-manuelle : en utilisant la main dominante (main droite pour les droitiers, gauche pour les gauchers) ou en utilisant la main non-dominante. N'émettant pas d'hypothèse sur la priorisation des tâches tiers, nous considérons comme possible que le chirurgien n'ait pas nécessairement sa main dominante disponible pour interagir avec l'application. En conséquence, l'idéal serait de prévoir la mise au point d'un système interactif permettant d'utiliser indifféremment la main gauche ou droite pour la réalisation des interactions prévues.

Absence de contrôle vocal Une opération chirurgicale se déroule en équipe. Chirurgiens, infirmiers et assistants, communiquent de façons répétées et fréquentes en se parlant. Nous avons donc évacué l'usage d'un système à interaction vocale conjoint à notre système d'interaction gestuelle. Techniquement, une telle option serait envisageable, en implémentant un système de reconnaissance vocale de mots-clés, en filtrant le bruit ambiant, couplé à un système de discrimination d'utilisateur par la voix. Néanmoins, dans cette option a été écartée du projet 3D-Surg pour se concentrer sur l'interaction gestuelle.

Interaction à " mains nues " À nouveau, les contraintes fortes du contexte opératoire imposent que le chirurgien ne soit pas encombré d'outillage supplémentaire. C'est pourquoi nous avons choisi de définir une interaction à mains nues plutôt que nécessitant un périphérique d'interaction portable ou portatif (gants de données, wiimote, etc.).

Respect de la taxonomie Il s'agit de définir un vocabulaire gestuel pour les cas d'usage dans la section 2.1.2. Nous considérons que les utilisateurs garderont l'usage des logiciels de visualisation à interface « classique » en dehors du bloc opératoire. Ceux-ci sont familiers avec le découpage fonctionnel de l'application tel que identifié dans les taxonomies. En conséquence, nous avons décidé de respecter strictement les taxonomies en substituant les tâches des feuilles de l'arbre, correspondant aux actions classiques, par des actions à mains nues, tout en conservant la structure arborescente telle quelle.

2.2.2.2 Critères conceptuels

Ayant isolé et considéré les critères contextuels nous permettant de déterminer comment répondre à notre problème initial, il nous faut alors aussi isoler et déterminer les critères à considérer au mieux pour nous permettre d'apporter une réponse efficace. Nous listons et détaillons ces critères conceptuels ci-dessous :

Utilisabilité & performance Il s'agit, par définition [2], de façon générique, du « *degré selon lequel un système, un produit ou un service peut être utilisé, par des utilisateurs spécifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié* ». Ici, les notions clés pour optimiser ce critère sont, d'une part, l'efficacité définie comme « *la précision et degré d'achèvement avec lesquels l'utilisateur atteint des objectifs spécifiés* » et, d'autre part l'efficience définie comme « *rapport entre les ressources utilisées et les résultats obtenus* ». La notion d'utilisabilité intègre donc la notion de performance qui est le regroupement des notions d'efficacité et d'efficience.

Stress et fatigue Souvent regroupé au sein du terme “ confort ”, affectant l'utilisabilité ainsi que le confort, le stress et la fatigue sont des critères critiques pour tout système d'interaction. La “ fatigue ”, ici, désigne l'inconfort, voire une faible douleur musculaire, résultante d'une utilisation prolongée entraînant généralement une baisse significative des performances de l'utilisateur. Ce que l'on désigne ici par “ stress ”, est l'effort ou la pression subie par les muscles, articulations ou les tendons des parties du corps de l'utilisateur impliquées dans l'interaction. Bien que le stress ne soit pas nécessairement aussi critique à court terme que la fatigue au regard des performances ou de l'utilisabilité, celui-ci doit être minimisé car il demeure inconfortable et est délétère au long terme sur la santé.

Bien qu'il y ait, à notre connaissance, encore peu de travaux détaillant spécifiquement les liens entre la fatigue et le stress musculaire et les interfaces gestuelles sans contact à mains nues, nous pouvons élargir notre champ de vision pour appréhender cette problématique. Par exemple, [5] met en évidence des associations positives entre le syndrome du canal carpien [W] ainsi que la tendinopathie [W] affectant la main et le poignet et la réalisation d'actions répétitives. De même, l'exposition fréquente à des postures “ extrêmes ”, c'est-à-dire exerçant une tension sur les muscles ou tendons des mains sont liées aux troubles sus-mentionnés.

En termes de fatigue, un problème fréquemment rencontré dans l'étude d'interactions gestuelles est celui connu sous la désignation de “ gorilla-arm ”. Des travaux ayant été entrepris

pour définir une métrique de quantification de cet effet [14] ont également dégagé des recommandations pour minimiser cet effet lors de la conception d'interface gestuelle sans contact :

- Considérer des mouvements avec objectifs ou cibles relatifs.
- La position de bras repliés et centrés (i.e. la paume et le coude alignés sur un axe horizontal) est la moins fatigante
- Utiliser des techniques qui minimisent les mouvements demandés. Par exemple dans le cas d'une tâche de sélection, privilégier les sélections avec maintien de positions chronométrées.

Intuition & cohérence d'utilisation partagées Le sens associé aux gestes est variable selon les personnes (cultures, langues, habitudes, etc.) et fortement lié au contexte. En revanche, nous avons également explicité comment certaines habitudes interactives dues aux usages quotidiens influencent de façon similaire les utilisateurs dans leurs expériences interactives, et donc répandent certaines pré-conception et attentes quant au fonctionnement de certaines interactions. De plus, le cas échéant, notre cas d'usage se destine à un ensemble d'utilisateurs dans un même contexte d'utilisation, et se substituant à un cas d'usage supposément bien connu, et donc au fonctionnement mnémonique leur étant familier. Par conséquent si, pour une tâche donnée, les attentes des utilisateurs quant à un geste à effectuer, ou des composantes de ce geste, sont similaires pour plusieurs utilisateurs, alors ce geste, ou les composantes concernées, peuvent être considérés comme une intuition partagée. Nous considérons ces occurrences comme des indicateurs d'une inventivité pertinente de geste.

Dans un même ordre d'idée, si lorsqu'une interaction propose plusieurs gestes pour sa réalisation, dans le cas où l'un de ces gestes est préféré plus fréquemment par plusieurs utilisateurs, et ce de façon cohérente vis-à-vis d'un contexte d'utilisation donné, l'on parlera de cohérence d'utilisation partagée. Cette cohérence partagée sera également à considérer comme un indicateur pour le choix définitif d'un geste à associer à une tâche donnée.

Mémorisation Comme évoqué tantôt, la disposition à se rappeler, ou à pouvoir se remémorer l'utilisation d'une interface est un critère important pour évaluer la pertinence de celle-ci. La " loi de Miller " [23] postule que le nombre d'éléments mémorisables à court terme ne serait que de 7, plus ou moins, selon les médiums concernés. S'agissant d'évaluer des gestes parmi tout un corpus, au regard de ce critère, deux types d'erreurs sont attendues : l'oubli — l'utilisateur ne sait plus quels gestes sont possibles pour réaliser une tâche donnée ; et la confusion — l'utilisateur confond certains gestes associés à une tâche, avec des gestes associés à une autre. Le nombre d'interactions, et donc de gestes à retenir, étant limité, nous nous attendons à rencontrer des erreurs du second type, lors de la recherche auprès d'utilisateurs. Un geste étant souvent confondu, pourra alors être considéré comme un indicateur d'un geste inadapté, au regard du sens et de la naturalité qu'il transmet, par rapport à la tâche d'interaction que l'on voudrait lui associer.

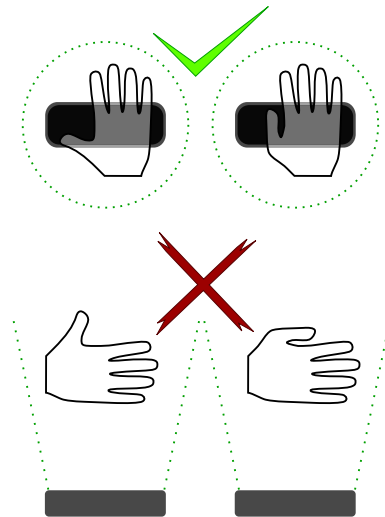


FIGURE 2.6 – Exemple d’occlusion dépendant de la position des mains en utilisant le LeapMotionTM : lorsque la paume est orientée parallèlement au capteur (partie supérieure), celui-ci peut aisément détecter si le pouce est replié ou non ; lorsqu’elle est pivotée de 90°, le pouce devient occulté pour le capteur, qui ne peut donc pas déterminer la posture du pouce.

2.2.2.3 Critères techniques

Une dernière catégorie de critères qu’il nous faut prendre en compte, sont des critères dit “ techniques ”. Étant donné que, pour nos travaux, nous travaillons avec le capteur LeapMotionTM, qui affecte directement et principalement ces critères.

L’espace de détection : Comme déjà évoqué, le LeapMotionTM emploie une zone de détection restreinte dont il faut tenir compte (voir figure 2.3).

Amplitude du geste : De fait, des gestes trop amples ne sont pas adaptés, car il est difficile de les cadrer à l’intérieur de la zone de détection et facile pour l’utilisateur de sortir par mégarde de la zone de détection avant d’avoir complété la réalisation du dit geste.

Détection : Il faut également prendre en considération les capacités et contraintes de détection des gestes imposées par le capteur. Dans le cas du LeapMotionTM, à titre d’exemple, la précision de détection des gestes ou de la posture par le capteur est dépendant de l’orientation de la main, du fait de problèmes d’occlusion (voir figure 2.6).

2.2.3 Panel de gestes

Une fois que nos critères de validation de gestes sont établis, nous sommes alors en mesure de procéder à une recherche de gestes adéquats pour élaborer un vocabulaire gestuel adapté à notre cas d’usage. Dans cette section, nous allons présenter, de manière exhaustive, toutes sortes d’interactions à mains nues pouvant être envisagées, pour chacune des tâches à réaliser. Ces interactions sont issues de propositions d’interactions basées sur des concepts d’interactions connues, adaptés à notre contexte et cas d’usage. Ci-après nous détaillons notre démarche

consistant à construire des interactions en fonction des critères établis dans la section précédente en partant d'interactions communément répandues.

2.2.3.1 Parenthèse sur les Menus

L'usage de menus est un standard du contrôle d'application. Ce mode d'interaction consisterait alors à accéder aux différentes tâches, via une sélection au travers d'un menu contrôlé par un geste unique. La suite de tâches à réaliser nécessiterait donc une suite de sélections des différentes entrées du menu. Une alternance de ces sélections deviendrait rapidement pénalisant pour l'utilisateur au regard de critères conceptuels tel que la fatigue du fait de la répétition du geste et, dans une moindre mesure, du temps d'exécution de par le temps d'alternance entre ces sélections. Nous n'avons par conséquent pas retenu cette possibilité, bien qu'elle serait intéressante à comparer à une solution finale purement gestuelle. L'usage de menus ayant été écartés, nous allons donc considérer l'usage de gestes pour chaque tâche séparément.

2.2.3.2 Gestes uniques par tâche

Translation

La manipulation directe présente des caractéristiques dont les qualités ont été mises en avant par Ben Schneiderman [34]. Il nous a donc paru logique d'explorer en premier lieu cette technique. Conceptuellement cela implique que l'objet manipulé suive *stricto sensu* les mouvements de la main. Tandis que notre cas usage est en 2D, les mouvements de la main sont exprimés dans un espace en 3D lors de la manipulation directe. Ainsi, une translation dans un plan perpendiculaire à l'écran devrait en principe engendrer un équivalent à l'action de zoom (translation sur l'axe Z). À l'évidence ceci ne respecterait pas la taxonomie qui sépare les actions de translation et de zoom. Dès lors, une des composantes du mouvement nécessitera d'être ignorée, filtrée ou adaptée sous peine que celle-ci ne soit rendue gênante ou redondante. La manipulation est facile à appréhender pour des utilisateurs novices, mais son utilisation par des experts peut se révéler contre-indiquée, car non satisfaisante en terme rapidité compte tenu de l'amplitude du geste à réaliser.

Compte tenu de ces considérations, nous présentons les gestes envisagés.

Main plate ouverte : cette posture suivie d'un déplacement sur un plan parallèle à la table peut sembler idéale à première vue, celle-ci étant simple à mémoriser et à réaliser. En effet il est facile d'associer ce mouvement à celui de l'on ferait pour déplacer un objet, telle une feuille de papier, sur cette même table. Cependant, la posture " main pleine ouverte " est l'une des deux postures naturelles de repos, la seconde étant le poing fermé. Le risque d'interprétation erronée de l'intention de l'utilisateur au travers de cette posture par le système est donc élevé.

Attrapage via fermeture du poing : pour des raisons similaires à la main ouverte, cette posture peut sembler intéressante. De même, la mimique de saisir un objet pour ensuite le

déplacer (similaire au principe du drag'n drop) est intéressante. En revanche, comme pour la main ouverte, il s'agit d'une potentielle posture réflexe de repos ou de transition.

Attrapage via pincement de deux doigts : une alternative intéressante serait de substituer le geste d'attrapage par le poing fermé par un attrapage à deux doigts (i.e : un pincement), cette posture étant rarement réalisée de façon inconsciente. Bien que ce geste puisse être adapté en principe, la transition vers la posture de pincement peut aisément entrer en interférence avec d'autres gestes ou postures de la main employant les mêmes doigts que ceux utilisés par le pincement.

Zoom

Comme pour la tâche de translation, la première technique considérée est celle de manipulation directe. Tel qu'évoqué dans la section 2.2.1.2, la tâche de zoom est sujette à certains a priori : l'idée de l'écartement pour agrandir (" *zoom-in* ") ou de rapprochement pour rétrécir (" *zoom-out* "). Ainsi, dans notre contexte applicatif utilisant des gestes de la main, une des premières piste serait de considérer l'écartement et rapprochement des mains pour induire respectivement un zoom ou dé-zoom. Or, comme mentionné précédemment, notre objectif est de pouvoir contrôler l'application de façon mono-manuelle. Dès lors demeure l'option de réaliser un pincement ou écartement en utilisant deux doigts d'une même main. Cependant, dans notre cadre d'interaction gestuelle 3D, le geste de pincement ne semble pas adapté, pour les mêmes raisons que celles évoquées lors de la discussion sur le pincement pour réaliser une translation d'une part. D'autre part, la mimique de pincement/écartement associée au zoom/dé-zoom pose des problèmes notables quant à son interprétation par un système de détection de gestes. Cela est dû au fait qu'il est difficile de reconnaître l'intention de l'utilisateur lors de sa transition vers le geste de pincement : " l'utilisateur rapproche-t-il ses doigts pour effectuer un pincement (zoom) ou pour préparer un écartement (dé-zoom) ? ". À noter que ce problème est endémique à ce mode d'interaction; dans le cadre d'une interaction via écran tactile, ce problème n'est pas présent, car le contact physique avec l'écran fait office d'action caractérisant l'intention de réaliser un zoom ou un dé-zoom.

Navigation

Comme vu dans la section 2.2.1.3, il s'agit d'une tâche pouvant être associée à un scroll. Ce principe de scroll constitue donc un déplacement sur une seule dimension. Une idée qui viendrait donc naturellement serait de fournir un balayage dans une direction pour engendrer un défilement dans un sens ou dans un autre. Mais le geste de balayage, dans un espace d'interaction 3D pose certains problèmes. Par exemple l'absence de support, comme un écran tactile, complique l'arrêt du défilement qui nécessiterait un geste supplémentaire. De plus, le scroll, dans beaucoup d'exemples, comme le défilement d'un texte ou d'une page web, l'utilisateur bénéficie d'une visibilité vers l'avant et vers l'arrière. Cela fait néanmoins défaut, dans notre cas d'usage où l'on navigue dans une pile d'images, une seule image à la fois. Aussi, l'usage

de gestes trop amples potentiellement engendrés par ce balayage, est à proscrire que cela soit relativement au contexte d'usage ou pour des considérations de fatigue et de stress musculaire.

Luminosité & contraste

Comme déjà mentionné (section 2.2.1.4), les interactions de modification de luminosité et de contraste, très similaires de nature, sont des tâches moins répandues dans les usages courants que les tâches précédentes. La métaphore interactive la plus souvent utilisée pour ce type d'interaction est celle du slider. Ne disposant pas d'exemple d'usage courant spécifique à cette tâche nous décidons de prendre appui sur cette dernière. C'est pourquoi nous avons voulu élaborer un geste rappelant celui de déplacer un curseur sur une ligne.

2.3 Proposition finale

2.3.1 Mécanismes interactifs de notre proposition

À la lumière des observations de la section précédente, nous avons fait le choix d'associer chacune de nos interactions à une posture particulière de la main qui aura pour signification l'activation d'un « mode d'interaction » correspondant à cette interaction. Ce mode correspond conceptuellement, pour le système, à l'entrée dans un état où les mouvements/déplacements de la main de l'utilisateur seront dédiés uniquement pour interpréter l'interaction associée à ce mode. Concrètement, cela veut dire que lorsqu'un utilisateur adoptera la posture signifiant, par exemple, « translation », l'application basculera dans le mode « translation ». Notre concept d'interaction repose donc sur une transition d'état, entre modes interactifs, où chaque interaction y est représenté, en plus d'un mode neutre/inactif destiné au repos où à l'interruption d'interaction. Ce principe pose alors la question de la réalisation effective de cette transition.

2.3.1.1 “ Posture d'enclenchement ”

Deux solutions pour réaliser ce “ basculement d'état ” se présentent. La première : une transition en deux temps, requérant une action pour entrer dans cet état puis un autre pour en sortir. La deuxième solution, est une activation continue qui reposerait sur le maintien d'une posture pour l'utilisateur désirant rester dans l'état actuel et dont tout abandon résulterait dans la sortie de cet état.

De prime abord, la première solution pourrait sembler plus simple à réaliser et à implémenter. Cependant, celle-ci présente une surcharge cognitive étant donné que celle-ci requiert l'emploi de deux actions (entrée puis sortie de l'état « translation »). À l'inverse la deuxième solution associe une seule posture à la réalisation d'une action, simplifiant la mémorisation. Nous avons donc choisi d'adopter la deuxième solution. Notant que celle-ci requiert des postures, d'une part, aisément discriminables des deux postures de repos évoquées précédemment et, d'autre part, n'entrant pas en interférence avec d'autres gestes ou postures lors de son adoption. Ne reste alors qu'à choisir la posture de la main que l'on associera à la tâche. Pour que

celle-ci soit facilement reconnaissable par le système et aisément mémorisable par l'utilisateur, nous optons pour des postures à l'expression prononcée. Une posture construite comme une combinaison de doigts tendus répond à l'ensemble de ces contraintes. Les postures possibles sont théoriquement au nombre de $32 (2^5)$. En revanche, certaines postures sont difficiles à réaliser, notamment celles incluant l'annulaire et l'auriculaire conjointement ou individuellement. Les interactions isolées, restent néanmoins inférieures au nombre de postures restantes.

2.3.1.2 “ Vecteur d'influence ”

Bien que concept de posture associée à un mode d'interaction réponde aux critères conceptuels de confort et de mémorisation tout en facilitant la détection, et donc en minimisant les erreurs, il n'est pas, en l'état, en mesure de répondre complètement à d'autres critères que sont l'économie motrice et la restriction de l'espace d'interaction. En effet, le mouvement intervenant après l'activation du mode d'interaction, via la posture, et induisant in fine la translation, doit aussi être sujet à discussion. En pratique, il apparaît rapidement que lier directement le mouvement de translation à celui de la main, à la manière d'un simple tracking, n'est pas satisfaisant. En cause première, l'espace de détection restreint oblige l'utilisateur à segmenter son interaction en une série de gestes. En effet, dès lors que l'utilisateur souhaite réaliser une importante translation en continue celui-ci franchira les limites de l'espace de détection, l'obligeant ainsi à stopper son interaction pour repositionner sa main dans cet espace. Ce problème peut être réduit en fonction du mapping qui est implémenté entre le déplacement de la main et l'amplitude de la translation induite, mais il ne pourra pas être totalement supprimé. Et ce d'autant plus que l'espace de détection de notre capteur est un cône. Dès lors, il apparaît qu'un tracking direct, même aux déplacements pondérés, n'est pas adapté à notre contexte technique. Il en va de même pour notre contexte environnemental, étant donné que le tracking direct incite à des gestes amples, ceux-ci étant proscrits dans notre cas d'utilisation.

Ainsi, pour nous affranchir de ces problèmes, plutôt que d'avoir une réalisation de l'interaction dépendante de l'amplitude des mouvements de l'utilisateur, nous optons pour une approche basée sur leur direction. Pour ce faire, nous utilisons la position de l'espace à laquelle l'utilisateur a adopté sa posture de déclenchement du mode translation. Cette position, P_0 servira de point de référence pour induire notre interaction tant que l'utilisateur sera dans le mode correspondant, c'est-à-dire tant que l'utilisateur maintiendra sa posture. Ensuite, lorsque l'utilisateur déplacera sa main dans l'espace d'interaction, transitant ainsi vers une nouvelle position P_n , toujours en maintenant sa posture de la main, le vecteur différentiel $\vec{V}_n = P_n - P_0$ induira la translation. P_0 demeurant fixe durant le maintien du mode, cela implique que tout déplacement de la main provoque donc un changement de valeur de P_n et de facto de la norme et la direction de $\vec{V}_n$. Ce vecteur, $\vec{V}_n$, n'est pas de nature ponctuelle, mais induit en continue la translation sur l'image. En d'autres termes, c'est l'existence de ce vecteur qui provoque la translation souhaitée, et ce, tant qu'il est défini. Subséquemment, la direction et la vitesse de la translation induite sont contrôlées par la direction et la norme de ce vecteur, constamment

modulées par l'utilisateur de la manière suivante :

$$T_n = T_{n-1} + k \cdot \overrightarrow{V_n}$$

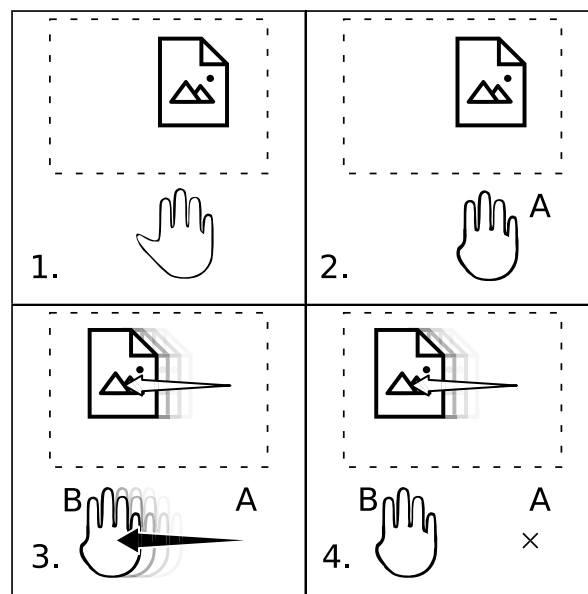
avec :

- T_n la position de l'image à l'instant n
- k une constante

Ce principe, illustrer via la figure 4.2 permet donc à un utilisateur de moduler la “ vitesse ” ou “ l'intensité ” de l'interaction en agissant directement sur la norme du vecteur. De même la direction du vecteur permet de contrôler la “ direction de l'interaction ” ou son “ sens ”, suivant la dimensionalité de l'interaction. En effet, le vecteur est généré de fait dans un espace 3D (l'espace de détection d'interaction du LeapMotionTM), or les interactions sont modulables sur un espace de dimensionalité réduite. Par exemple, la translation est une interaction s'exprimant sur un plan, donc de dimensionalité 2. Le zoom, quant à lui, est une interaction à une seule dimension, car il s'exprime en fonction de l'intensité seule (intensité nulle : aucun zoom ; intensité positive : zoom ; intensité négative : dé-zoom). Parmi nos interactions isolées, la dimensionalité maximale est donc de deux, alors que notre vecteur est en trois dimensions. Nous filtrons donc les composantes du vecteur qui seront actives, autrement dit celles qui ont une influence sur l'interaction, en rendant celles qui n'en ont pas inactives. En d'autres termes, si l'on prend en exemple la translation que l'on induit via un mouvement sur un plan, XY le cas échéant, cela veut concrètement que, seules les composantes X et Y du mouvement détecté sont considérées pour le geste, la composante Z est ignorée. De plus les différentes composantes du mouvement pour induire le vecteur, ne sont pas nécessairement mappées avec leur équivalent en espace virtuel. Toujours en prenant pour exemple la translation, interaction bidimensionnelle sur un plan virtuel X'Y', l'on peut imaginer différent mapping des composantes vectorielles à celle de l'interaction induite :

1. mapper le plan XY (ignorant donc la composante Z du vecteur). Le mouvement considéré sera donc tout mouvement en parallèle au plan du support, ignorant dans la génération d'interaction tout mouvement parallèle à l'écran, i.e. : vers le haut ou le bas.
2. mapper le plan XZ (ignorant donc la composante Y du vecteur). Le mouvement considéré sera donc tout mouvement en perpendiculaire au plan du support, ignorant dans la génération d'interaction tout mouvement perpendiculaire à l'écran, i.e. : vers l'avant ou l'arrière.

Par conséquent, le mapping des composantes du vecteur, aussi bien que le filtrage des composantes non-effectives, est une caractéristique à part entière du vocabulaire basé sur ce principe de « vecteur d'influence ». Il conviendra dès lors de spécifier pour chaque interaction en spécifiant les dimensions dans lesquelles s'inscrivent le mouvement considéré.



1. Posture neutre : aucune interaction induite.
2. Posture associée au mode translation : initie la translation. La posture est adoptée au point de l'espace A.
3. Mouvement de la main en maintenant la posture : le vecteur virtuel $\overrightarrow{AB}$, B étant la position de l'espace courante de la main, induit une translation continue sur l'image selon sa direction et sa norme.
4. La posture de la main est maintenue. Aucun mouvement de la main. La translation selon $\overrightarrow{AB}$ se poursuit tant que la posture est maintenue.

FIGURE 2.7 – Exemple de “ posture déclencheur ” prédéfinie pour le cas d’usage 2D.

2.3.2 Vocabulaire conceptuel

Nous avons donc choisi d'adopter ce principe interactif de couplage « Vecteur d'influence — Modalité via une posture d'enclenchement » et de le mettre en pratique pour chaque interaction isolée. Pour chaque interaction, il demeure alors encore des choix à faire : celui de la posture associée à l'interaction et celui de la composante du vecteur d'influence à considérer pour induire l'interaction.

Translation Vis-à-vis du plan d'interaction, c'est-à-dire le plan de l'espace sur lequel les mouvements sont suivis pour retranscrire l'interaction (voir : 2.3.1.2), deux choix s'offrent à nous : soit un plan horizontal parallèle à la table, soit un plan vertical parallèle à l'écran. Nous avons privilégié le plan horizontal parallèle à la table, par cohérence avec le cas d'usage de référence 2D, employant la souris, qui utilise la table. De plus, un plan vertical parallèle à l'écran nécessiterait des mouvements de levé du bras et du coude engendrant de la fatigue ainsi qu'une posture stressante du poignet, qui s'avère problématique compte tenu de la fréquence d'utilisation de cette tâche. Concernant la posture, nous avons choisi arbitrairement la posture « pouce replié — restes des doigts tendus ». La posture « main pleinement ouverte » est déjà utilisée comme posture par défaut, mais nous voulions employer une posture proche, car celle-ci rappelle le geste de faire glisser un objet sur un support.

Zoom Ayant mis de côté l'interaction paraissant la plus évidente, le pincement ou le rapprochement de deux doigts, et en reprenant les mêmes considérations que pour l'interaction de translation (espace d'interaction restreint, discrimination des gestes, exclusion de la posture de repos, etc.) nous choisissons d'opter pour ce même concept d'interaction pour le zoom. Ce choix a été fait, d'une part, par absence d'autre exemple d'interaction duquel s'inspirer, et, d'autre part, considérant prématurément la possibilité d'étendre le principe de « Vecteur d'influence — posture d'enclenchement » à tout le vocabulaire par souci de cohérence de fonctionnement pour éviter une potentielle surcharge cognitive.

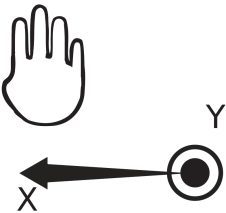
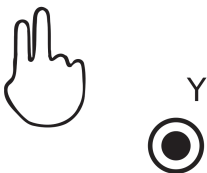
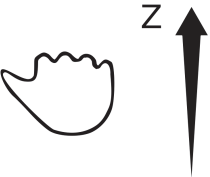
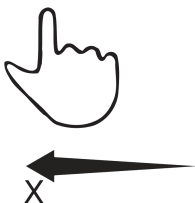
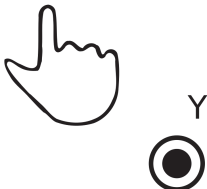
Reste alors la question du plan d'interaction sur lequel l'on choisit de filtrer la détection du zoom. Le zoom est une interaction modulable sur une seule dimension : zoom-avant correspondant à une augmentation de la valeur de zoom, zoom-arrière à une diminution de cette valeur. Ainsi s'impose le choix de la dimension filtrante sur laquelle lire le déplacement du vecteur (avant/arrière = Z, haut/bas = Y, gauche/droite = X). N'ayant pas réellement de point de référence spécifique à cette question, ce choix restera arbitraire. Nous décidons d'utiliser l'axe Z de notre vecteur ponctuel relatif d'interaction, le sens avant/arrière, concrètement “ vers ” ou “ fuyant ” l'écran, en extrapolant l'idée suivante : l'utilisateur “ pousse ” virtuellement, via le sens de son mouvement, vers l'écran, induisant un dé-zoom, ce qui se traduit visuellement pas une image qui semble s'éloigner de l'utilisateur, en cohérence donc avec ce mouvement. Inversement l'utilisateur “ tire ” l'objet vers lui, induisant un zoom et donc l'illusion visuelle que l'image se rapproche de lui.

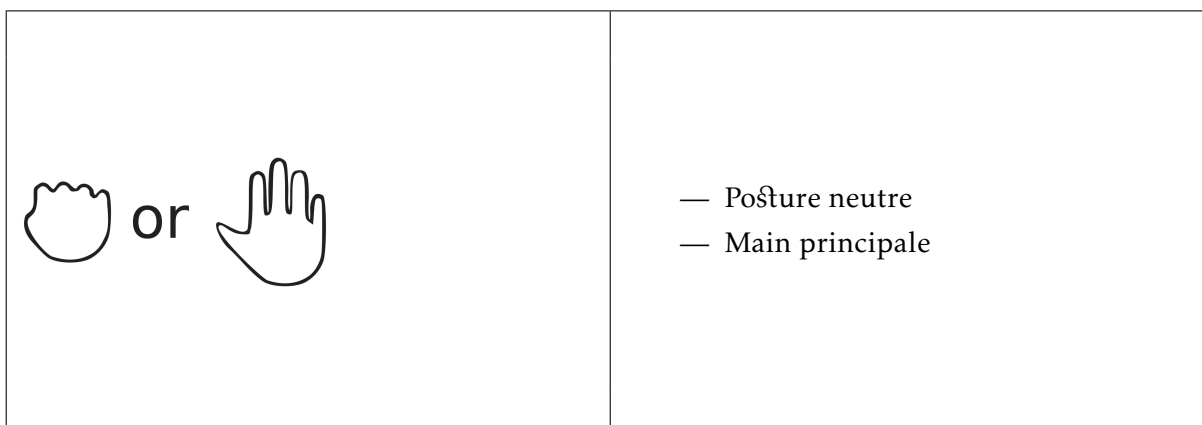
Navigation Du fait des considérations établies dans la section précédente, nous avons à nouveau choisi ici d’y associer le principe interactif couplant « postures d’enclenchement » avec un « vecteur ponctuel relatif d’interaction ». Cette fois encore, il nous faut choisir les dimensions spatiales à filtrer sur le vecteur 3D pour cette interaction. Comme expliqué précédemment nous prenons en référence les interactions de type “ défilement ” qui s’appuient sur la notion de monter/descendre. Dès lors nous choisissons ici de considérer le déplacement du vecteur sur l’axe Y (haut/bas) pour contrôler la navigation dans la pile d’images. Ce faisant nous choisissons de privilégier l’aspect mnémonique par rapport au stress musculaire en se basant sur l’observation d’utilisation : la navigation alterne entre des phases de défilement rapide, dans le but d’atteindre une région d’intérêt global, et une phase de défilement lent pour trouver une image particulière dans la pile. En utilisant notre système de vecteur relatif, alterner entre ces deux phases nécessite peu de mouvement, atténuant la fatigue potentiellement engendrée par un mouvement vertical. Il ne reste alors qu’à déterminer une posture discriminante vis-à-vis de celles utilisées par les autres interactions.

Luminosité & Contraste Nous avons traduit l’idée de déplacement d’un curseur (mentionnée dans la section précédente) par un déplacement de la main sur un axe d’un plan parallèle à la table (plan XZ). Bien qu’ici le fait qu’il s’agisse de sélectionner une valeur dans un intervalle restreint puisse s’accommoder de l’espace d’interaction limité, nous décidons de reprendre le concept élaborer plus haut, l’activation par « postures d’enclenchement » dont l’interaction est contrôlée par le « vecteur ponctuel relatif d’interaction ». À ce stade, le principal argument d’utilisation de cette technique est la cohérence avec les autres interactions qui en font usage. Comme nous avons donc deux interactions de natures très proches, nous souhaitons d’une part discriminer leur activation de façon efficace mais également les regrouper au sein d’une gestuelle sémantiquement proche. De fait, notre idée première est de discriminer ces deux interactions via l’axe du déplacement employé sur un même plan, ici, celui de la table. Concrètement, nous avons choisi que le déplacement le long de l’axe X (gauche/droite, parallèle à l’écran) contrôlera la luminosité et le contraste serait contrôlé par un déplacement sur l’axe Y (avant/arrière, approche ou fuite de l’écran). Néanmoins, il apparaît que détecter l’interaction à réaliser simplement en fonction de la direction entraîne des interférences entre les deux interactions. Il apparaît en pratique qu’il est fréquent pour un utilisateur de dévier d’une trajectoire et d’engendrer une interaction filtrée sur un autre axe. On peut répondre à ce problème en utilisant une zone tampon sur chaque axe d’interaction, sur laquelle la modulation du vecteur relatif d’interaction n’engendre pas d’interaction. Bien que cette solution soit fonctionnelle, le fait de disposer d’une telle zone morte proche de l’origine de notre vecteur encourage potentiellement la réalisation des gestes amples. De plus, les limites de notre espace d’interaction ne pouvant être clairement délimitées car la dimension de celui-ci est variable (zone de détection en forme de cône), et l’origine du vecteur d’interaction n’étant pas constante, il apparaît des cas où la création du vecteur résulte en une interaction non fonctionnelle ou prêtant à confusion. Dès lors nous faisons le choix de confirmer la discrimination de ces deux tâches cette fois encore en utilisant des postures de main distinctes.

Postures neutres Le principe de « Modalité via posture d'enclenchement » demande, en plus de définir une posture correspondant à un mode d'interaction associé à chaque tâche, de définir une « posture neutre ». Cette posture sert de posture de repos, ou bien de posture transitoire entre les différentes tâches à remplir. Cela signifie qu'une posture neutre ne fait pas basculer l'environnement dans un mode interactif, et que donc tout mouvement de la main pendant que celle-ci maintient une posture n'induit aucune interaction. En pratique, définir une posture neutre ne correspond pas à l'implémenter spécifiquement dans notre système de détection, mais revient à spécifier formellement cette posture comme étant " réservée " pour le repos/l'inactivité.

Une posture neutre doit être rapidement " accessible " depuis les autres postures et depuis une posture proche des positions naturelles de repos de la main. Nous avons donc choisi comme postures neutres, deux postures : le poing fermé et la main pleinement ouverte (tous les doigts tendus). Ces deux postures sont donc " réservées " par notre vocabulaire, étant donné que ce sont deux positions de repos " naturelles " pour les mains, et facilement accessibles depuis d'autres éventuelles postures : il suffit de tendre ou replier tous les doigts pour stopper la tâche en cours. Il est à noter cependant que, dans notre implémentation, toute posture non formalisée par notre vocabulaire aura l'effet d'une posture neutre, mais seul les deux postures sus-mentionnées seront formalisées et donc réservées en tant que telle.

	— Posture pour le mode translation — Main principale (dans le réglage par défaut, main gauche pour les sujets gauchers, et main droite pour les droitiers) — Mouvement induisant la translation doit être effectué sur le plan XY (plan parallèle à la table)
	— Posture pour le mode zoom — Main principale — Mouvement induisant le zoom doit être effectué sur l'axe Y (éloigné ou vers l'écran)
	— Posture pour le mode de navigation dans la pile d'image — Main principale — Mouvement induisant la navigation dans la pile doit être effectué sur l'axe Z (de haut en bas)
	— Posture pour le mode de changement du contraste — Main principale — Mouvement induisant le changement du contraste sur l'image doit être effectué sur l'axe X (de gauche à droite)
	— Posture pour le mode de changement de luminosité — Main principale — Mouvement induisant le changement de la luminosité sur l'image doit être effectué sur l'axe Y (éloigné ou vers l'écran)



2.4 Intégrer “ l’intuitivité ”

Dans ce chapitre, nous avons retracé notre cheminement méthodologique depuis l’observation du cas d’usage jusqu’à l’élaboration d’un premier concept de vocabulaire gestuel. Pour ce faire nous avons entre-temps explicité notre analyse du cas d’usage, détaillé notre analyse des tâches et des gestes qui leur sont habituellement associées, puis nous avons établi un ensemble de critères régissant notre recherche de geste adaptés. Le vocabulaire gestuel alors proposé est régi selon un principe interactif, nommé un couplage « Vecteur d’influence — Modalité via posture d’enclenchement », dans l’optique d’être en accord avec les critères de conception précédemment présentés. Néanmoins, la satisfaction de la plupart des critères conceptuels n’est pas couverte jusque-là par les solutions proposées. Les critères conceptuels découlant de « l’expérience utilisateur », ceci peuvent difficilement être intégré sans consultation des utilisateurs eux-mêmes. En effet, jusqu’ici nous avons adopté une approche analytique du problème qui, bien que permettant d’aborder plusieurs aspects de la conception de systèmes interactifs, est hermétique aux ressentis utilisateur lors de l’utilisation. Or cette composante « d’expérience utilisateur » impacte fortement l’efficacité et l’appréciation finale du système conçu et est devenu un incontournable de la conception de système interactif (voir 1.4).

Spécifiquement, les critères conceptuels n’ayant pas été couverts ou adressés jusque là sont nommément “ l’utilisabilité & performance ” et “ l’intuition & cohérence d’utilisation ”. Le critère “ d’utilisabilité & de performance ” bien que pouvant être, et ayant été, anticipés lors de la conception, nécessite d’être évalué (au même titre que le critère de “ mémorisation ”) via des évaluations par des utilisateurs de la solution proposée. Quant au critère “ d’intuition & de cohérence d’utilisation ”, il s’agit de s’assurer de disposer d’un vocabulaire dont les gestes qui le composent ont un haut niveau d’intuitivité dans leur réalisation, et que, dans la mesure du possible, cette intuitivité soit perçue ou partagée par une majorité d’utilisateur.

L’étape suivante de notre travail est donc de mettre en place cette contribution utilisateur à la conception de notre vocabulaire. Comme nous le détaillerons dans le prochain chapitre, nous organisons donc des expériences utilisateur ayant un double objectif : d’une part obtenir une composante « d’intuitivité » à intégrer dans notre vocabulaire gestuel, et d’autre part évaluer notre proposition précédente afin de s’assurer qu’aucun des critères conceptuels susmentionnés n’est, en pratique, invalidé. Comme nous le verrons dans la suite, pour remplir

ce premier objectif, nous nous inspirons de méthodologies dites « d'élicitation d'utilisateurs ». Cela nous permet alors d'extraire des intuitions de gestes à partir desquels nous pourrions compléter notre proposition de vocabulaire à l'aide des intuitions de gestes proposés par l'utilisateur.

EXPÉRIENCE 2D

Dans le chapitre précédent, nous avons présenté la première partie de notre travail. Celui-ci présentait le point de départ de notre travail méthodologique en abordant le problème par une démarche analytique. Partant d'une analyse de cas d'usage et de son environnement nous avons isolé les tâches centrales pour formaliser ce dit cas d'usage. De même, cela nous a permis d'identifier et appréhender les conditions et contraintes d'utilisation rencontrées par les chirurgiens lors de la consultation d'images médicales en per-opératoire. Forts de ces analyses nous avons alors été en mesure de faire un choix sur la technologie employée pour le reste de notre approche en choisissant de travailler dès lors avec le capteur LeapMotionTM. De plus, cela nous a permis d'orienter notre démarche exploratoire en nous permettant de cibler les interactions et donc les tendances interactives à observer. Ainsi, grâce aux observations recueillies jusque-là, nous avons été en mesure d'établir un ensemble de critères à remplir pour orienter la recherche des gestes qui seraient à même de fournir une solution d'interfaces gestuelles efficaces.

Néanmoins, malgré notre travail d'analyse et d'étude sur le cas d'usage et les usages communs, il reste un aspect que l'on ne pouvait intégrer à toute proposition de vocabulaire gestuel à partir d'une telle approche analytique : l'intuitivité de notre solution. De plus, bien que pris en compte lors de nos réflexions, plusieurs aspects doivent être confirmés par évaluation et retours d'utilisateurs, tels que la performance, l'induction de fatigue, etc. C'est pourquoi, la seconde partie de notre démarche, décrite dans ce chapitre, consiste en l'implémentation d'expériences utilisateur. Cette expérimentation, formalisée en protocoles utilisateurs expérimentaux a pour but de remplir un double objectif : extraire des intuitions de geste à partir desquels nous pourrions améliorer notre proposition de vocabulaire gestuel et évaluer notre solution par des retours et essais utilisateurs de notre solution.

Concernant l'extraction de gestes intuitifs pour chaque tâche, l'objectif est donc de prendre connaissance des intuitions de gestes généralement associées par les sujets à ces dites tâches. Pour ce faire, nous nous sommes tournés vers des méthodes dites " d'éllicitation de gestes " qui consistent à recueillir des gestes directement auprès d'utilisateurs en demandant à ceux-ci de produire des gestes dans un contexte contrôlé. Plus spécifiquement nous reprenons le concept des expériences dites de « Magicien d'Oz (Wizard of Oz) », lors desquelles l'expérimentateur demande au sujet de réaliser des tâches en proposant des gestes de son propre chef, mais où les effets des interactions devant être réalisées le sont en fait par l'expérimentateur lui-même.

Notre protocole, qui sera détaillé dans ce chapitre intégrera donc une phase d'éllicitation suivie d'une phase d'évaluation de solution. Ainsi à l'issue de la mise en place de ce protocole nous serons en mesure d'apporter des corrections à notre proposition, en utilisant, d'une part

les propositions de gestes “ d’intuition ”, et d’autre part les résultats des évaluations. Or, les méthodes “ d’éllicitation ” desquelles l’on s’inspire, font partie de méthodes dites « centrées utilisateurs » qui trouvent leur efficacité dans un procédé d’amélioration itératif. En somme, ces méthodes sont appliquées plusieurs fois et chaque nouvelle itération améliore ou « raffine » les qualités de la solution issue de la précédente itération. C’est pourquoi, n’étant pas indifférents à ce procédé, nous avons pris la décision d’implémenter notre protocole expérimental en deux itérations : la première destinée à des utilisateurs génériques, dit « tout-venants », et la seconde aux utilisateurs finaux ou experts, c’est-à-dire les chirurgiens. Les utilisateurs « tout-venants » constituent une population accessible que nous considérons comme adaptée pour extraire des intuitions de gestes pour les tâches isolées. En effet, comme on l’a mentionné dans les chapitres précédents, les interactions isolées peuvent être retrouvées, en partie, dans d’autres cas d’usages communs, ce qui nous amène à considérer les intuitions produites lors de la phase d’éllicitation comme pertinente. Ce sont ces gestes, « d’intuitions génériques » que nous utiliserons après cette première itération, pour intégrer les composantes pertinentes de ses propositions à notre vocabulaire. De plus, l’étape d’évaluation à laquelle sont soumis les utilisateurs « tout-venants », lors de cette première itération, permet d’avoir une première évaluation sur notre proposition émise au précédent chapitre. Enfin, la seconde itération, auprès des utilisateurs experts, nous permettra d’isoler des tendances interactives éventuellement manquantes lors de la précédente phase d’éllicitation et de confirmer la viabilité de notre solution lors de la phase d’évaluation.

Dans ce chapitre, nous poserons les principes de notre démarche puis nous détaillerons le protocole expérimental tel qu’implémenter. Ensuite, nous présenterons les résultats récoltés, que nous interpréterons et dont nous présenterons les points clés.

3.1 Approche

Dans cette section, nous revenons sur les principes sous-jacents à notre démarche. Nous y détaillons comment nous nous sommes inspiré du “ Wizard of Oz ” pour construire notre méthodologie. Nous expliquons également comment nous utilisons conjointement les phases d’éllicitation et d’évaluation pour améliorer notre vocabulaire gestuel.

3.1.1 “ Élicitation ” de gestes

Les méthodes dites “ d’éllicitation ” de gestes sont des méthodes qui s’inscrivent dans la lignée de la méthodologie “ centrée utilisateurs », étant une forme de “ design participatif », et sont désormais bien connus et décrites depuis quelque temps déjà [33]. Ces méthodes sont devenues un procédé relativement populaire pour des recherches de gestes s’affranchissant des limites des technologies utilisées et des biais de conception techniques en se recentrant sur les envies de l’utilisateur [25]. Ces méthodes sont d’autant plus employées pour divers cas d’application : écrans tactiles [42], périphériques mobiles [31], et autres systèmes multimodaux pouvant faire usage d’autres modes d’interaction tel que la reconnaissance vocale [24] [22]. Bien que ces méthodes n’aient pas de formalisme précisément défini, leur principe consiste

à isoler des opérations cibles, c'est-à-dire des fonctions interactives ou tâches à remplir, et à consulter les utilisateurs cibles en leur demandant d'exécuter les gestes qui, selon eux, seraient à même de remplir au mieux ces tâches au cas par cas. Le but est donc de collecter les propositions de geste émises par les utilisateurs pour en extraire celles qui semblent les plus adaptées au regard de critères définis. L'intérêt de procéder en utilisant ce type d'approche est que les gestes proposés sont issus d'intuitions d'utilisateurs. On peut dès lors présumer, qu'après observation et analyse des gestes obtenus, l'on puisse extraire des gestes les plus communs, ou des aspects de gestes apparaissant de façon récurrente, l'on puisse en déduire des aspects de la composition de ces gestes ayant une bonne composante " d'intuitivité ". Il apparaît donc qu'adopter une telle approche pour compléter notre démarche soit adaptée étant donné que nous avons effectivement isolé les tâches cibles, effectuer une analyse des besoins et déterminé des critères pour concevoir un vocabulaire gestuel adapté, tel que recommandé par [26].

3.1.2 Suggestions utilisateur pour un vocabulaire gestuel d'intuition

L'idée principale est donc d'observer les gestes que veulent produire naturellement les utilisateurs confrontés à la directive de proposer une interaction pour accomplir une interaction spécifique. En d'autres termes, pour un ensemble donné de tâches, pour lequel un utilisateur doit exécuter une série de tâches spécifiques, nous lui demandons de produire le geste qui lui semble instinctivement le plus approprié ou pertinent. Ainsi, chaque sujet nous fournit alors un vocabulaire gestuel spécifique et personnel, construit autour de leur intuition propre vis-à-vis de l'usage de geste de la main pour les cas d'usage présentés. De toute évidence, un tel vocabulaire ne doit pas être considéré comme universel et varie vraisemblablement d'un sujet de l'autre. En revanche, un nombre suffisant de vocabulaires utilisateurs permettent d'isoler et d'étudier les aspects caractéristiques. À titre d'exemple, nous pouvons répertorier et isoler les types de gestes récurrents et observer si ces gestes sont employés de façon proche ou similaire suivant les usages. Ce faisant, nous pourrions alors utiliser les gestes isolés pour déduire un vocabulaire globalement intuitif (pour les utilisateurs cibles) tout en restant spécialisé au contexte d'utilisation.

Ce type de méthodologie ayant gagné en popularité dans divers domaines de recherche au fil du temps, Vogiatzidakis et al. [40] ont fourni une veille (" *survey* ") vis-à-vis des " méthodes d'incitation de gestes " (" *methods of gestures elicitation* ") pour l'interaction gestuelle sans-contact appliquée à divers domaines d'application (e.g., réalité virtuelle, desktop, affichage public, etc.). Dans cette étude le bloc opératoire (" *operating room* ") était le domaine le moins représenté, une seule étude sur 47 papiers sélectionnés indiquant un manque partiel d'application de telles méthodologies où l'utilité de l'interaction sans contact a pourtant été démontré.

Pour collecter les gestes proposés par les sujets pour chaque tâche à accomplir, notre approche consiste à proposer une expérience basée sur des expériences type " Magicien d'Oz (Wizard of Oz) ". Ces expériences consistent, comme expliqué par Dahlback et al. [9], en " des

études où les sujets se voient expliqués qu'ils interagissent avec un système informatique via une interface à langage naturel (NDLA : utilisant un vocabulaire gestuel), sans que cela ne soit réellement le cas. À la place, l'interaction est arbitrée par un opérateur humain tiers, le "magicien", ce qui a pour conséquence que le sujet se voit offert une liberté d'expression accrue, ou qu'en utilisant directement des interfaces naturelles existantes". Évidemment les interactions ne sont pas réellement induites par le sujet, mais par l'expérimentateur, à l'insu du dit sujet. Ce principe permet aux expérimentateurs d'observer et collecter les interactions, dans notre cas des gestes, qui viennent spontanément à l'esprit de l'utilisateur pour réaliser les tâches demandées en étant confronté à des cas d'usage similaires aux cas d'usage cibles. Cette méthode fournit potentiellement autant de vocabulaires qu'il y a de sujets. Ce faisant, les gestes récoltés sont analysés à posteriori pour identifier les tendances gestuelles qui se dégagent. Bien qu'il ne soit pas attendu que des gestes strictement identiques soient retrouvés pour tous les utilisateurs, il est néanmoins attendu que certaines composantes de ces gestes soient récurrentes (posture de la main, mouvement, métaphore gestuelle, etc.).

3.1.3 Approche en 2 étapes : raffinement du vocabulaire utilisateur

Les interactions sur lesquelles nous nous concentrons, comme listées dans la section 2.2.1, sont soit familières pour une grande variété d'utilisateurs génériques (telle que la translation ou le zoom), soit peu connues car relativement endémiques au contexte d'utilisation (telle que la navigation dans une pile image). De plus, les utilisateurs cibles, les chirurgiens, constituent une population dont la disponibilité est limitée. Dès lors, nous avons divisé notre approche en deux étapes.

1. **Intuitions génériques** : La première étape de notre approche consiste à extraire les intuitions de gestes, par la méthode décrite dans la section 3.1.2, pour les utilisateurs "tout-venants", qui ne sont pas nécessairement experts vis-à-vis du contexte d'utilisation. Les interactions étudiées étant conceptuellement simples et proches de certaines que l'on peut trouver dans d'autres logiciels, et le profil utilisateur des chirurgiens est similaire à celui d'utilisateurs génériques, nous amène à considérer que les retours de tels utilisateurs, dit "tout-venants". De fait, le but de cette première étape est donc d'extraire un vocabulaire gestuel d'intuition générique.
2. **Raffinement du vocabulaire** : La seconde étape de notre approche consiste, premièrement à enrichir et compléter le vocabulaire extrait précédemment en utilisant les intuitions de gestes extraites lors de l'étape précédente grâce à des intuitions cette fois extraites auprès d'utilisateurs cibles. Deuxièmement, ces utilisateurs cibles évalueront le vocabulaire d'intuition, de l'étape précédente, nous permettant ainsi d'utiliser leurs retours pour écarter ou corriger certains gestes. Le but est donc de raffiner le vocabulaire en utilisant un point de vue d'expert.

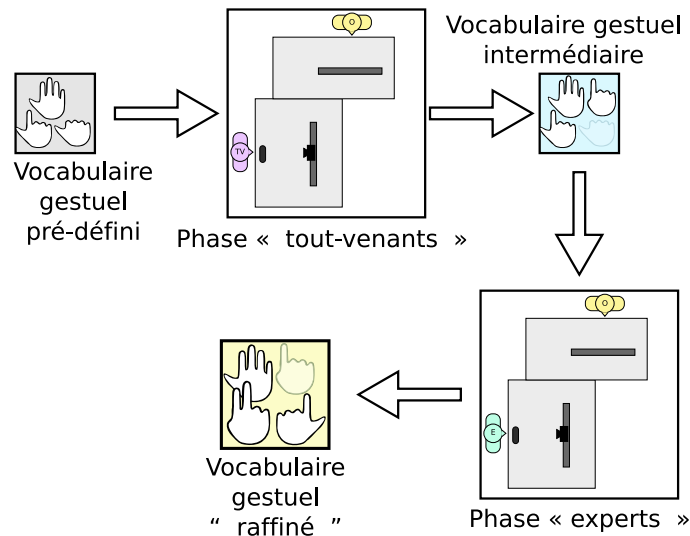


FIGURE 3.1 – Itération de notre démarche protocolaire, découpée en deux étapes.

3.2 Protocole - générique

Une fois notre approche définie, nous l'avons implémentée selon le protocole tel que décrit dans cette section. Comme expliqué dans la section précédente et détaillé ci-après, il s'agit d'un protocole expérimental divisé en deux itérations : la première pour extraire des intuitions génériques auprès d'utilisateurs « tout-venants » en vue d'intégrer une composante d'intuitivité à un vocabulaire gestuel pré-défini ; la seconde nous permet de valider et corriger le vocabulaire intermédiaire obtenu précédemment et de recueillir les retours des utilisateurs cibles, en concordance avec les méthode d'élicitation, pour obtenir un vocabulaire plus adapté intégrant au mieux, selon nos critères, une dimension d'intuitivité. Cette démarche globale est illustrée via la figure 3.1, et résumé via le tableau 3.1.

TABLE 3.1 – Protocole en deux itérations

Sujets « tout-venants »	Sujets experts
1. Pré-questionnaire 2. Présentations & explications 3. Présentations des tâches & interactions 4. Phase d'élicitation de gestes 5. Questionnaire relatif aux gestes & interactions 6. Évaluation d'un <i>vocabulaire gestuel pré-défini</i> 7. Questionnaire de comparaison	1. Pré-questionnaire 2. Présentations & explications 3. Présentations des tâches & interactions 4. Phase d'élicitation de gestes 5. Questionnaire relatif aux gestes & interactions 6. Évaluation du <i>vocabulaire gestuel intermédiaire</i> 7. Questionnaire de comparaison
Cas d'usage <i>artificiel</i>	Cas d'usage : <i>Données médicales</i>

3.2.0.1 Dispositif expérimental

Pour notre expérience, nous avons installé notre dispositif expérimental selon la configuration suivante (voir Figure 3.2) :

- Un PC exécutant une application (*ICube forge* (GitLab) : https://icube-forge.unistra.fr/hurstel/touchlessurg_proto1.git) implémentant toutes les interactions et la visualisation nécessaires à notre cas d'utilisation. Cette implémentation a été écrite en C++ et utilise la bibliothèque Qt utilisant une interface *QML* pour implémenter le *GUI*. Ce programme permet également d'exécuter des séquences d'interactions automatisées prédéfinies permettant ainsi de produire, passer ou réinitialiser ces interactions.
- Deux moniteurs. Un moniteur utilisé pour afficher les images et les retours visuels des interactions pour le sujet. Le second, pour que l'opérateur puisse déclencher l'interaction pré-enregistrée pendant que le sujet proposait ses intuitions gestuelles.
- Un capteur LeapMotion™ installé sur un bureau devant lequel le sujet est sollicité pour réaliser son intuition d'interaction. Bien que, comme expliqué, les gestes effectués n'induisent pas réellement d'interaction, le capteur est utilisé pour enregistrer les données gestuelles telles que formatées par défaut par l'API du LeapMotion™.
- Une webcam placée entre l'écran et le sujet pour filmer les gestes proposés par l'utilisateur, pour une évaluation post-protocole et interprétation des gestes.

Nous avons utilisé une application personnalisée plutôt qu'un logiciel DiCOM existant pour notre protocole pour des raisons de flexibilité : l'interaction sans contact étant un mode interactif non supporté par un tel logiciel, nous voulions pouvoir avoir un contrôle complet de l'affichage et des retours visuels. Pour le cas d'utilisation 2D, nous avons utilisé un petit re-

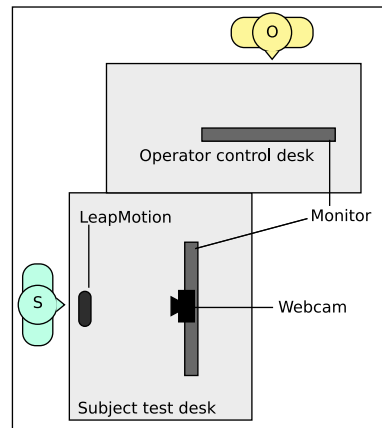


FIGURE 3.2 – Disposition de dispositif expérimental.

père affiché sur l'image montrée afin que l'utilisateur puisse suivre le déplacement de l'image pendant l'interaction telle qu'une traduction et avoir une meilleure attente concernant les résultats des interactions telles que le zoom, puisque ce repère permet de cibler des domaines spécifiques de l'application. De plus, nous avons jugé nécessaire d'ajouter des repères visuels pour aider l'utilisateur à se situer dans la pile d'images lors de la navigation dans la pile.

3.3 Protocole « tout-venants »

3.3.1 Procédure

Pour la première étape de notre approche nous avons appliqué, auprès d'utilisateurs " tout-venants ", la procédure suivante :

1. **Pré-questionnaire** : Le sujet se voit demander de compléter un questionnaire de profil utilisateur via un web-form. Voir annexe B.1.1.
2. **Présentations et explications** : Nous débutons en présentant aux sujets les grandes lignes, intérêts et principes de notre travail et démarche, ainsi que le capteur LeapMotionTM et ses spécificités.
3. **Présentation des interactions** : Nous avons ensuite présenté toutes les interactions (listées dans Section 2.2.1), pour lesquelles l'utilisateur devra donner une intuition.
4. **Propositions d'interactions** : L'utilisateur se voit présenter un cas d'utilisation artificiel (voir Section 3.3.2) pour lequel on lui demande de proposer des gestes, lors de la réalisation d'un ensemble spécifique de tâches séquentielles, à chaque fois qu'il utilise explicitement une certaine interaction. Conformément au principe des expériences " Magicien d'Oz ", chaque interaction était en fait induite par l'opérateur, tandis que le sujet proposait un geste. Aucune interaction ne sera induite, mais seulement la conséquence à l'écran de l'interaction. Par exemple, lorsque nous présentons l'interaction de translation, une translation pré-enregistrée est effectuée à l'écran. La raison étant que nous voulions nous assurer que le sujet comprenne l'interaction demandée et ses conséquences attendues avant de lui demander une proposition gestuelle pour cette interaction.

5. **Questionnaires relatifs aux interactions :** On a demandé au sujet de remplir un formulaire Web, de type qualitatif, demandant de décrire l'interaction qu'il a proposée pour chaque interaction ainsi que son inspiration. Cela nous a permis de mieux comprendre l'intention derrière le geste, de détecter tout écart éventuel entre l'intention et la réalisation, et de détecter tout facteur de mémorisation largement négatif de l'association geste-interaction. Voir annexe B.1.2.
6. **Évaluation d'un vocabulaire gestuel pré-défini :** Le sujet est confronté à un vocabulaire gestuel prédéfini (voir Section 3.3.3). On demande à l'utilisateur de reproduire la même séquence de tâches qu'auparavant, mais cette fois-ci en utilisant une interface gestuelle fonctionnelle sans contact qui met en œuvre notre vocabulaire prédéfini.
7. **Questionnaire de comparaison :** Le sujet est ensuite invité à remplir un formulaire qui, pour chaque interaction, lui demande de comparer son intuition, donnée précédemment, avec l'interaction imposée par notre vocabulaire. Ceci a été fait pour suggérer implicitement au sujet de corriger son intuition, si le sujet le juge approprié, en tenant compte de facteurs tels que la minimisation de la réalisation motrice, l'utilisation à moyen et long terme, le confort postural, ou la fatigue. Le but est de recueillir le retour des sujets sur la comparaison pour observer si le sujet corrige son intuition en tenant compte de ces facteurs, sans influencer l'intuition initiale. Voir annexe B.1.3.

3.3.2 Cas d'usage artificiel

Puisque les sujets “ tout-venants ” n'étaient pas des utilisateurs cibles, et n'ont donc pas d'expérience en matière de consultation de données d'images médicales, nous avons estimé qu'il serait déroutant et donc contre-productif d'effectuer le protocole avec les mêmes données que dans un cas cible d'utilisation. C'est pourquoi nous avons choisi de mener cette expérience avec un cas d'utilisation artificiel qui ne nécessite pas de connaissances particulières pour comprendre et sur lequel nous pourrions demander à l'utilisateur d'effectuer le même type d'interactions pour atteindre des objectifs similaires.

En pratique, pour le cas d'utilisation 2D, nous avons substitué la pile d'image tomographique par une pile d'images génériques sur laquelle effectuer les interactions (zoom, traductions, etc.). Dans ce cas, pour imiter la transition à travers la pile d'images, notre exemple est en fait constitué de quelques images clés (ou images stables) et les images intermédiaires dans cette pile sont en fait des images compensant une transition de type morphing entre deux images clés, simulant ainsi une navigation continue dans une pile d'images de manière cohérente. Le but pour les chirurgiens, lorsqu'ils consultent des scans tomographiques, est de naviguer pour trouver et isoler des images spécifiques, et des zones sur ces images; nous croyons que notre cas d'utilisation artificiel fournit une équivalence appropriée pour notre protocole pour des sujets “ tout-venants ” non-experts.

De plus, notre protocole se déroule en utilisant notre propre logiciel, qu'on désignera par “ démonstrateur ”, développé par nos soins. Cette application, décrite plus en détail dans le chapitre dédié suivant 4, nous permet un meilleur contrôle de l'expérience interactive par rapport à la possibilité de faire passer nos protocoles sur des applications DiCoM. Concrètement,

le démonstrateur nous permet plus facilement de disposer d’une séquence de tâches préparées dont le déroulement est scripté. D’autre part, cela nous permet de proposer des métaphores interactives, elles aussi détaillées dans le chapitre dédié, qui sont une aide notable au sujet pour comprendre les effets des interactions et l’état courants de l’application.

En somme, notre cas d’usage artificiel, appuyé par les fonctionnalités de notre démonstrateur a pour but de faciliter la compréhension et alléger la charge cognitive induite sur l’utilisateur pour que ce dernier puisse plus facilement se concentrer sur la tâche d’élicitation de gestes, dans un premier temps, puis sur sa tâche d’évaluation dans un second temps.

3.3.3 Vocabulaire gestuel pré-défini

Pour rappel, la motivation derrière un vocabulaire gestuel prédéfini, malgré le fait d’avoir déjà collecté des intuitions de gestes pour les interactions, est que leur efficacité n’est pas garantie. Même s’il est facile pour les sujets de produire des gestes intuitifs “ à la volée ”, cela ne garantit par l’efficacité ou l’utilisabilité à long terme des qualités pratiques que sont, par exemple, l’efficacité à long terme, l’évitement du stress musculaire dans un usage durable ou répété, une charge cognitive réduite, ou même la compatibilité technique avec le capteur.

Notre vocabulaire gestuel prédéfini repose sur le principe des “ gestes déclencheurs ”, voir au chapitre précédent la section 2.3.2. Cela signifie que notre vocabulaire utilise certaines postures de la main pour faire basculer notre application dans des modes interactifs associés qui sont, par exemple, concernant notre cas d’usage 2D : le mode translation, le mode zoom, le mode navigation dans une pile d’image, etc. Le mode choisi reste actif tant que le sujet maintient ses mains dans la posture correspondante. L’interaction est induite par un vecteur virtuel $\overrightarrow{AB}$ calculé à partir de A , la position de l’espace à laquelle la posture a été adoptée, et B , la position courante de la main (voir aussi Figure 4.2). Comme évoqué dans la section précédente, ce principe est expliqué, démontré et testé par les utilisateurs avant leur participation à la partie “ évaluation ” de l’expérience.

Toutes les interactions reposent ainsi sur ce même principe : déclencher un mode d’interaction en adoptant la posture de main correspondante et en modulant ensuite l’interaction induite en déplaçant la main depuis point d’adoption de la posture. Ce principe présente trois avantages : une réalisation motrice minimisée, un stress musculaire réduit, une charge cognitive réduite.

Puisque notre vocabulaire gestuel prédéfini est à une main, nous l’avons évalué pour les mains dominante et non dominante. Nous avons toujours évalué d’abord notre vocabulaire avec la main dominante puis avec la main non dominante pour voir si l’utilisation de la main non dominante induit un inconvénient critique en termes d’efficacité. Pour comparer directement l’impact de l’efficacité, les deux évaluations ont été réalisées sur la même séquence d’interactions.

3.3.4 Résultats

3.3.4.1 Observations & analyses

Nous présentons brièvement les observations extraites de notre méthode du premier cas d'utilisation étudié : la consultation des scans tomographiques 2D. Notre protocole expérimental de type " Magicien d'Oz " a été réalisé sur un groupe de 12 sujets.

Translation : La plupart des gestes ont été effectués d'une seule main dominante; les gestes ont toujours été effectués dans le plan parallèle au bureau; la posture la plus populaire était une main ouverte (42%) suivie d'une métaphore saisissante (25%).

Zoom : La plupart des gestes ont été effectués d'une seule main dominante (17% des gestes bi-manuels); le geste le plus populaire était un geste répété de type pincement (58%) appuyé par un mouvement de type push-pull typé avec différentes postures (25%).

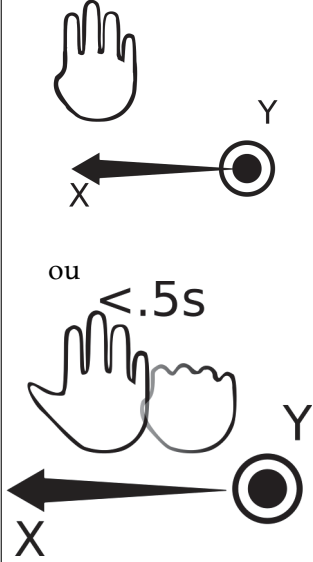
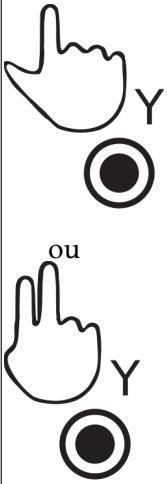
Navigation dans la pile d'image : Les gestes ont toujours été exécutés avec une seule main dominante; la modalité la plus courante fut d'adopter une posture (variable) accompagnée d'un mouvement vers le haut ou vers le bas (25%); les autres gestes furent variables en modalité et en nature (gestes de la main pivotants, mimiques de type push-pull, mimique pouce vers le haut/le bas).

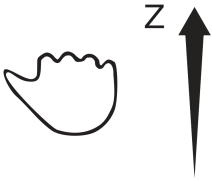
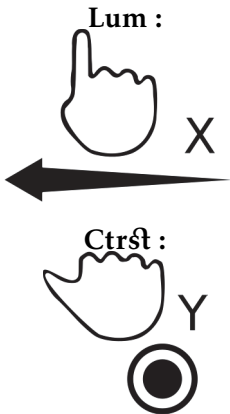
Gestion du contraste/luminosité : La tendance récurrente (84%) fut de regrouper ces deux interactions sous une certaine modalité identique (posture (25%) et direction du mouvement (42%)), et de distinguer laquelle est induite avec une autre modalité (quelle main est utilisée, mouvement de la main et direction du mouvement). Cela est probablement dû à la nature étroite de ces deux interactions.

3.3.4.2 Vocabulaire intermédiaire

Les observations listées ci-dessus ont permis d'établir le vocabulaire gestuel résumé par la table ci-dessous.

Vocabulaire “ intermédiaire ”, issu de la première phase « tout-venant » du protocole.

	Translation : — Paume plate et ouverte, déplacement sur le plan XY (plan parallèle à la table), main dominante — Geste d’agrippement : ouverture suivi d’une fermeture du poing. La transition, pour être détecté doit être effectuée sous un temps inférieur à 0.5 secondes. Déplacement sur le plan XY. <i>Nous avons ici inclus le geste d’agrippement, en tant qu’alternative, de manière compatible à notre mécanique d’interaction : ici l’action permettant d’entrer dans le mode “ translation ” n’est plus une posture mais une transition de posture, reproduisant le geste d’agrippement.</i>
	Zoom : — Extension du pouce et de l’index, mouvement sur la composante Y, main dominante. — Extension du majeur et de l’index, mouvement sur la composante Y, main dominante. <i>Le principe ici est de conserver une des composantes gestuelle de la mimique du pincement, en l’occurrence la composante du geste consistant à tendre deux doigts.</i>

	Navigation dans la pile d'image : — extension du pouce, mouvement sur la composante Z (de haut en bas)
	Modification de la luminosité : — extension du pouce, mouvement sur la composante Y (approcher/éloigner de l'écran sur le plan parallèle à la table) Modification du contraste : — extension de l'index, mouvement sur la composante X (gauche/droite sur le plan parallèle à la table) <i>Le choix de ces gestes est expliqué au paragraphe 3.3.4.2</i>

Le cas de la luminosité et du contraste : La manipulation du contraste et de la luminosité étant de nature très similaire, les utilisateurs « tout-venant » n'ont pas su proposer des tendances gestuelles claires permettant de les discriminer. Par ailleurs, lors de l'évaluation de vocabulaire ces deux interactions étaient souvent confondues.

Pour extraire le vocabulaire gestuel (Table 3.3.4.2), nous avons dû éviter des gestes similaires ou trop proches pour différentes interactions afin de minimiser le risque d'interactions indésirables dues à des gestes qui se chevauchent ou des transitions manquées entre eux. De plus, nous avons jugé nécessaire d'éliminer tous les gestes non continus basés sur le mouvement (pincements répétés pour un zoom continu) et de les adapter pour qu'ils soient utilisés en mouvement continu, de la même façon que le vocabulaire prédéfini que nous proposons, cette initiative étant appuyée par les commentaires des utilisateurs récoltés du questionnaire de comparaison.

Pour ce vocabulaire intermédiaire nous avons donc décidé, suite à ces résultats, de conserver le même mécanisme interactif, que celui employé pour notre vocabulaire conceptuel : reposant sur un basculement d'état via une posture d'enclenchement et un contrôle via vecteur d'influence (voir 2.3). Ce principe interactif, qui a été conçu en premier lieu pour répondre aux critères établis dans le chapitre précédent, n'a pas rencontré de retour disqualifiant ou d'incompatibilité fondamentale dans les gestes recueillis. C'est pourquoi nous avons fait le choix

de raffiner ce vocabulaire à l'aide de l'élicitation des utilisateurs « tout-venants » en conservant son principe fondamental : cela répondait en effet à notre besoin d'intégrer au mieux la composante « d'intuitivité » tout en conservant notre réponse aux critères établis précédemment.

3.4 Protocole « experts »

Pour la deuxième étape de notre démarche, qui a consisté à affiner le vocabulaire gestuel précédemment déduit à l'aide de l'apport des sujets experts, la procédure a été essentiellement la même.

Pour cette deuxième étape, les objectifs sont d'une part d'observer si un consensus de gestes intuitifs se dégage parmi ces sujets experts et, si tel est le cas, d'observer et d'évaluer son degré de différence par rapport aux gestes intuitifs extraits de sujets " tout-venants " ; et d'autre part d'évaluer ce vocabulaire précédent et de cerner les aspects qui doivent être affinés pour parfaire son utilisation par les utilisateurs experts cibles.

3.4.1 Cas d'usage médical

Vis-a-vis des sujets " experts ", qui sont des chirurgiens et donc des utilisateurs familier avec le cas d'usage et contexte cibles, nous ne recourons pas à un cas d'usage artificiel via un jeu de données pré-construit comme effectué pour le protocole sur les utilisateurs " tout-venant " (voir 3.3.3 et 2.3.2). Dans cette itération de notre protocole, comme indiqué par le tableau 3.1, nous soumettons les sujets à un cas d'usage plus proche du cas réel via un jeu de données correspondant à des coupes tomographiques réelles. Cela a été fait considérant que les utilisateurs expérimentés n'auront pas besoin d'un environnement interactif plus adapté étant donné leur expérience dans la manipulation des jeux de données, que sont les piles d'images tomographiques. Bien que ce jeu de données est un jeu de données réel, nous utilisons néanmoins encore notre " démonstrateur " pour cette partie de l'expérience, et non un logiciel de visualisation DICOM spécialisé. Cela a pour but naturellement de conserver un degré de similarité fort entre les deux expériences, nécessaire pour que les retours des utilisateurs soient compatibles et recevable si l'on veut intégrer ceux-ci dans notre vocabulaire intermédiaire.

3.4.2 Résultats

Comme indiqué plus haut, nous avons également réalisé un protocole " Magicien d'Oz " pour les utilisateurs finaux, qui sont des collègues chirurgiens ayant une expérience antérieure du logiciel DiCOM. Nos observations sur les résultats de cette expérience suivent celles que nous avons recueillies pour les sujets " tous-venants ". En comparant les deux, aucune différence significative n'est apparue, en ce sens qu'aucun nouveau geste, aucune nouvelle posture ou métaphore d'interaction n'ont été proposés.

Selon la méthodologie que nous avons présentée (Section 3.4), nous avons évalué les performances des utilisateurs finaux en utilisant le vocabulaire gestuel que nous avons pu déduire

du protocole utilisateur “ tout-venant ”. Ces observations ont été effectuées sur un groupe de cinq collègues chirurgiens.

Le cas d'utilisation de chaque sujet était composé de chaque type d'interactions dans un ordre différent. Pour chaque type d'interaction, trois tâches ont été effectuées. De plus, chaque cas d'utilisation a été fait deux fois pour chaque sujet (main dominante puis main non dominante). Une tendance générale que nous avons observée est que la performance a augmenté pour chaque tâche individuelle à travers les deux cas d'utilisation répétée. Le temps d'achèvement a été réduit ainsi que le nombre de “ lacunes ”, même si la main non dominante a été utilisée dans la deuxième séquence, comme le montre la figure 3.3, indiquant que la performance augmente en raison de l'effet d'apprentissage.

Ici, ce que nous appelons le nombre de “ gaps ”, qui est montré dans la Figure 3.3a, est le nombre de fois que le sujet a dû corriger la position actuelle de sa main pour affiner le vecteur inducteur d'interaction, nécessaire pour atteindre le but de la tâche donnée. De plus, nous constatons que le nombre d'écarts a diminué pour toutes les interactions du même type, confirmant ainsi notre observation. Pour évaluer les performances que nous avons mesurées, Figure 3.3b montre le temps d'achèvement pour chaque type de tâche (y compris les interactions de la main dominante et non dominante) pour chaque utilisateur. Même si les tâches (du même type) peuvent différer en termes d'amplitude (c'est-à-dire la distance entre l'état de départ et l'objectif), nous avons estimé que cela était couvert par le fait que notre vocabulaire nous permet de traiter rapidement les petites et grandes amplitudes d'interaction en modulant le vecteur virtuel utilisé pour produire l'interaction (voir Section 2.2.3.2).

Les questionnaires présentés aux sujets nous ont permis d'obtenir les commentaires suivants :

- **La translation** fut jugée confortable et utilisable par les sujets.
- **Le zoom** fut jugé généralement confortable tout en donnant un contrôle satisfaisant sur la tâche.
- **La navigation dans la pile d'image** fut considérée comme offrant un bon contrôle par les sujets.
- **Le contrôle de la luminosité et du contraste** furent jugés bien contrôlables et utilisables.

De ce retour d'expérience, il ressort que le vocabulaire raffiné que nous avons proposé a été bien accueilli par rapport aux gestes spontanément proposés lors de la phase “ Magicien d'Oz ”.



(a) Nombre d'écarts par utilisateur, pour chaque type de tâche

(b) Temps d'exécution (en secondes) par utilisateur, pour chaque type de tâche

FIGURE 3.3 – Évolution de la performance de “ l'utilisateur expert ” en termes de : (*gauche*); et le temps de complétion (*droite*).

3.5 Limites & difficultés techniques

Au cours de nos deux expériences, les sujets ont parfois éprouvé des difficultés de détection manuelle qui ont été rencontrées de manière incohérente tout au long de la procédure. Certains de ces problèmes pourraient être attribués à des imperfections dans le système de détection du LeapMotion™, notamment lors du changement de posture des doigts, bien que l'on puisse diminuer significativement ces erreurs en apprenant à adopter des postures plus prononcées des mains. Cependant, il semblait que certains de ces problèmes de détection étaient causés par des conditions d'éclairage naturel (heure du jour, temps, etc.) et si le sujet portait des accessoires réfléchissants comme des bagues, un bracelet réfléchissant ou une montre. Même si les paramètres peuvent être contrôlés dans la salle d'opération, ils restent une contrainte à laquelle il faut se méfier. Dans le même ordre d'idées, il faut garder à l'esprit la possibilité qu'un autre membre de l'équipe d'exploitation intervienne involontairement et souligner la nécessité d'un espace (restreint) d'interaction dédié. Bien sûr, le LeapMotion™ offre une interaction mono-utilisateur par conception, ce qui convenait à notre travail, mais aussi en limitant la portée de ce travail à un seul contexte utilisateur. Par conséquent, même si notre approche et nos résultats ne sont pas liés à un dispositif, un autre système de détection pourrait être utilisé pour étendre ce travail afin de développer une approche qui permette d'utiliser le vocabulaire de conception dans un contexte multi-utilisateur simultané.

3.6 Comparaison et discussion concernant une étude similaire

Jurewicz et al. [16] présentent un travail similaire, construit autour d'une approche similaire. Il est toutefois important de noter que, pour chaque étude, même si elles partagent le même domaine d'application, c'est-à-dire la salle d'opération, le cas d'utilisation est radicalement différent. Alors que notre travail se concentre sur la manipulation des images médicales, Jurewicz et al. se concentrent sur les tâches anesthésiques assistées par ordinateur. Néanmoins, les deux études visent à proposer une méthode pertinente pour concevoir des vocabulaires gestuels pour une interaction sans contact en milieu chirurgical.

Dans l'étude de Jurewicz et al. et dans la nôtre, le principe repose sur la réalisation d'une expérience " magicien d'Oz " sur deux populations de sujets : les experts et les non-experts. D'après notre compréhension, Jurewicz et al. ont considéré les deux populations comme des utilisateurs cibles, utilisant des vocabulaires différents pour chacune d'elles si les vocabulaires gestuels recueillis ne présentent pas suffisamment de similitudes pour converger. Notre étude, cependant, ne considérerait pas nos utilisateurs " tout-venants " (ou non-experts) comme des utilisateurs finaux. L'un des principes clés de notre approche était que, comme nous l'avons expliqué, nous avons utilisé le vocabulaire gestuel extrait " tout-venants " pour définir la base d'un vocabulaire intuitif qui serait affiné à l'aide des commentaires et comparaisons ultérieurs de nos utilisateurs cibles. Nous estimons que cette divergence méthodologique prend forme à partir de deux considérations différentes. Tout d'abord, nous pensons qu'avec la plupart des in-

terfaces logicielles modernes, soit l'ensemble des interactions disponibles de l'interface novice est un sous-ensemble des interactions disponibles pour les utilisateurs experts, soit les deux ensembles, si différents, restent disponibles en même temps. Dans tous les cas, plus les fonctionnalités sont riches et nombreuses, plus la probabilité de trébucher sur le problème du chevauchement des gestes pour différentes interactions est grande. Deuxièmement, nous croyons que, dans le contexte hautement spécialisé de la salle d'opération, les utilisateurs novices sont soit des utilisateurs pour une courte période de temps, soit des utilisateurs en formation et visent à devenir des utilisateurs experts. Une autre différence notable dans notre étude est que, comme mentionné dans la section 3.3.3, nous ne considérons pas l'intuitivité comme le seul critère important pour la conception d'un vocabulaire gestuel pertinent. D'où l'apparition de notre vocabulaire gestuel prédéfini, qui est une composante de notre méthodologie qui ne répond, nous le croyons, à aucune équivalence directe dans l'étude de Jurewicz et al.

De notre point de vue, notre étude a bénéficié d'un cas d'utilisation (manipulation d'images médicales) moins complexe, en termes de nombre d'interactions pour mapper notre vocabulaire. La manipulation d'images est aussi composée d'interactions qui sont certes plus répandues que celles qui doivent être réalisées pour accomplir des tâches anesthésiques assistées par ordinateur. Cela a probablement facilité notre déduction d'un vocabulaire gestuel basé sur les commentaires des utilisateurs novices et qui correspond de près au vocabulaire qui semble convenir aux utilisateurs finaux. Jurewicz et al, d'autre part, ont rencontré un nombre accru de propositions d'interaction qui étaient significativement différentes de celles proposées par les utilisateurs novices ou experts. Cela nous a amenés à envisager la possibilité que, en utilisant notre approche pour leur cas d'utilisation, nous ayons eu besoin d'une étape supplémentaire de protocole basé sur le "Magicien d'Oz" pour affiner correctement notre vocabulaire gestuel utilisateur final. Il convient également de noter que Jurewicz et al. ont introduit des métriques, telles que, notamment, la métrique « temps de réaction », comme mesure d'intuitivité qui aurait également pu être utile à notre travail.

Néanmoins, le fait que nos deux approches aient été couronnées de succès, tout en étant toutes deux basées sur le "Magicien d'Oz", nous conforte dans notre conviction que ce type de méthodologie est pertinent pour la conception de vocabulaires gestuels efficaces et intuitifs pour une interaction sans contact en salle opératoire.

CONCEPTION & DÉVELOPPEMENT

4.1 Introduction

Dans ce chapitre nous présentons le développement informatique réalisé au cours de cette thèse. En effet, pour remplir le cahier des charges protocolaires établi dans le chapitre précédent (chapitre 3), nous avons développé une application de visualisation spécifique à notre usage expérimental. Il est courant pour des cas d'études similaires (voir 1.3) d'utiliser des systèmes de visualisation existants pour y greffer le système d'interaction désiré, dans notre cas une interaction gestuelle sans-contact au LeapMotion™. Néanmoins, comme nous l'avons évoqué précédemment, la contextualisation et la bonne compréhension du déroulement de l'expérience interactive via des métaphores visuelles est d'une importance cruciale. Ainsi donc, nous avons fait le choix de développer une application à part entière, et plus particulièrement une interface visuelle sur laquelle nous pourrions y intégrer sans barrières ou contraintes techniques trop importantes tout élément visuel que nous jugerions approprié. Bien que les logiciels de visualisation disponibles bénéficient d'interfaces standardisées et bien connues par les utilisateurs experts, nos protocoles expérimentaux incluent des utilisateurs « tout-venant » qui ne sont pas familiers avec de tels systèmes. De fait, développer cette application a pour but de nous offrir un contrôle avancé sur chaque aspect de l'expérience interactive. De plus, cela nous permet alors de nous assurer de disposer d'autant de fonctionnalités que nécessaire à nos protocoles tels que présentés dans les chapitres précédents. Ces dites fonctionnalités, ainsi que le développement et la construction de notre application, sont détaillés ci-après.

4.2 Démonstrateur

Pour répondre aux besoins protocolaires détaillés dans le chapitre 3, du temps développement a été investi dans une application logicielle. Cette application¹, que l'on désignera ici en tant que “ démonstrateur ”, a pour objectif d'implémenter les fonctionnalités nécessaires aux tests utilisateurs employant le LeapMotion™.

4.2.1 Dépendances et spécifications

Mise à part son interface graphique, ce démonstrateur a été codé en C++. L'interface graphique, pour ce qui est de la partie présentée aux sujets, a été codée en QML (“ Qt Modeling

1. https://icube-forge.unistra.fr/hurstel/touchlessurg_protocol.git

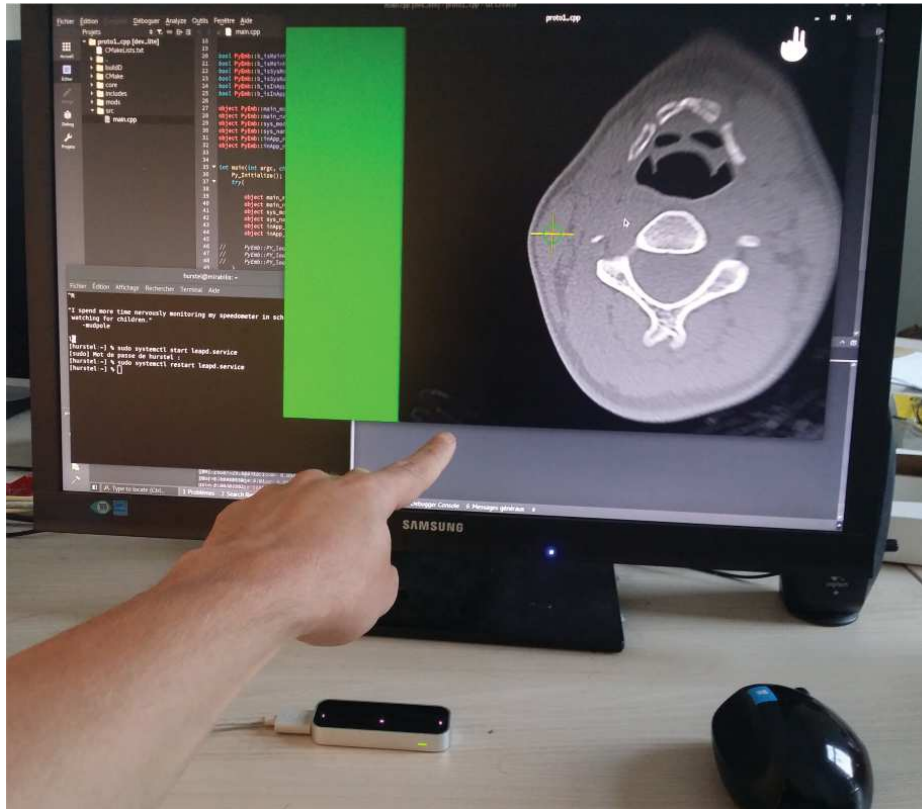


FIGURE 4.1 – Photographie lors du développement du démonstrateur.

Language ”). Le code source fait appel aux libraires suivantes :

- **La librairie standard C++** — En dehors des structures de données classiques (files, piles, etc.), l’implémentation des *thread* (processus légers), de la librairie standard sont utilisés. L’application étant découpée en différents modules, ces threads sont utilisés pour implémenter une file de communication de messages entre les différents modules.
- **La librairie Boost** — Utilisée pour certaines implémentations de structures de données ; en l’occurrence “ des pointeurs dynamiques ”, permettant d’assurer facilement la pérennité des données échangées entre les différents modules.
- **Le framework Qt** — Utilisé pour implémenter les éléments d’interface graphique destinés à l’expérimentateur (chargement de fichiers, etc.). Il intègre l’interpréteur de langage QML, et permet de faire le lien entre l’application et les éléments d’interface codés via ce langage. Qt offre aussi bon nombre d’outils d’utilité globale, en l’occurrence un parseur et générateur de documents XML, format que nous utilisons pour la sauvegarde de divers données persistantes.
- **LeapMotion Desktop SDK (v2)** — Framework permettant d’implémenter l’interaction gestuelle via le capteur LeapMotion™.

Le programme a été développé et déployé sous une distribution GNU/Linux (ArchLinux x86-64, maintenue à jour), exploitant une machine aux spécifications suivantes :

Mémoire 15,6 Gio

Processeur Intel® Core™ i7-3770 CPU @ 3.40GHz × 8

Carte Graphique GeForce GTX 560 Ti/PCIe/SSE2

4.2.2 Compilation et construction

La compilation de l'application est gérée en utilisant l'outil de construction logiciel CMake. La flexibilité offerte par l'outil en matière de gestion de la compilation via son langage, permet de facilement fragmenter l'application. De fait, l'application est organisée de façon modulaire : le " cœur (core) " de l'application est compilé en tant qu'exécutable, tandis que les divers fonctionnalités sont réparties en différents modules, compilés en tant que bibliothèques dynamiques, puis liées au cœur lors de la phase d'édition de liens. Fonctionnellement, le cœur réalise les différents appels de fonctions nécessaires à l'initialisation (et la destruction, en fin de vie de l'application) des différents modules, et gère la communication entre les différents modules. Cette organisation du développement permet une souplesse d'altération de fonctionnalités implémentées, car la suppression, l'ajout ou la modification de modules isolés n'affecte pas le fonctionnement des autres modules et donc le reste de l'application. Cela facilite donc le développement et l'intégration de fonctionnalités expérimentales et temporaires.

À titre d'exemple, la captation des données produites par le LeapMotion™, l'interpréteur de gestes, et l'interface graphique constituent 3 modules différents. Lors d'une exécution standard de l'application, se déroule le processus suivant :

1. Une main est reconnue par le capteur qui relève des données de mouvements.
2. Les données brutes reconnues, via le premier module sont envoyées par celui-ci au reste de l'application.
3. Le cœur de l'application qui gère les messages et leur distribution informe le module d'interprétation de gestes de l'arrivée de ces données.
4. L'interpréteur reconnaît ces données comme étant, par exemple, une translation.
5. L'interpréteur envoie alors à son tour un message informant de son interprétation des données.
6. Le cœur de l'application informe le module d'interface graphique de la translation reconnue.
7. La translation correspondante est induite à l'écran.

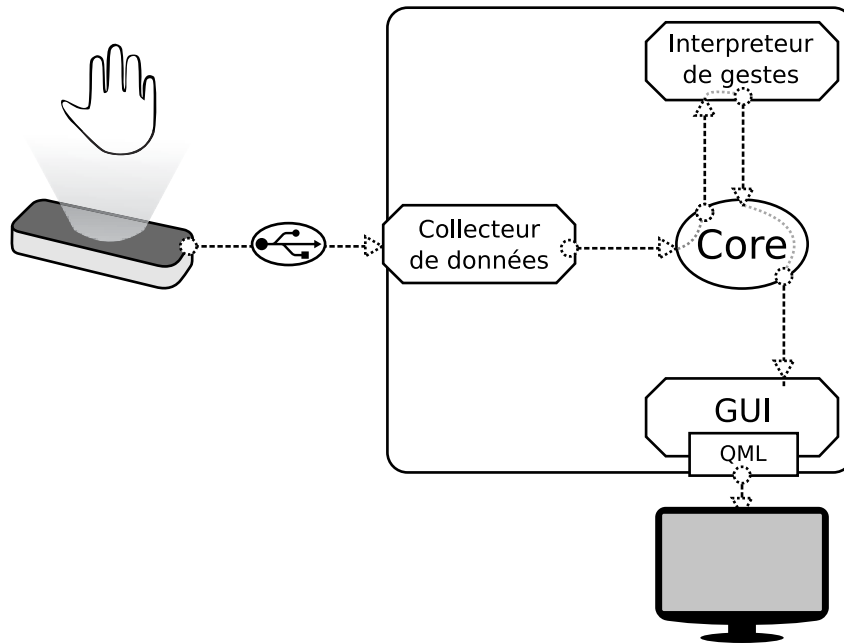


FIGURE 4.2 – Schéma simplifié de l'application “ démonstrateur ” — Dans cet exemple l'application est construite à partir de trois modules : le collecteur de données depuis le capteur, l'interpréteur de gestes, et l'interface graphique.

L'organisation modulaire permet dans un tel exemple, de substituer aisément l'interaction gestuelle via le capteur de geste, par un autre mode d'interaction, en ayant un impact minimal sur le reste de l'application. De même, substituer le module d'interprétation de gestes par un autre équivaut à changer la façon dont sont interprétés les gestes sans affecter l'affichage, altérer le collecteur de données, ou réorganiser les dépendances fonctionnelles au sein de l'application. Concrètement cela nous a permis, au fil de l'avancement de la thèse, de substituer un interpréteur de gestes statique (vocabulaire gestuel fixé, codé “ en dur ”), par un interpréteur flexible modulé par des fichiers externes décrivant le vocabulaire à suivre via un format XML. De même, l'intégration ultérieure d'un lanceur automatique de séquences d'interaction (à des fins démonstratives pour les sujets), a nécessité uniquement l'ajout d'un module parsant des fichiers XML, et envoyant des messages correspondants aux interactions (translation, zoom, etc.) lues séquentiellement dans ce fichier. Cette fonctionnalité d'automatisation des interactions offerte par ce module, faisait office d'interaction induite par l'opérateur lors de la phase de “ proposition d'interactions (cf 4) ” par les utilisateurs lors de nos protocoles.

Lors de son utilisation pour réaliser les tests utilisateurs, l'application était composée des modules suivants :

- Module de persistance des réglages de l'application. Lecture au lancement et sauvegarde à la fermeture sur le disque dur les paramètres génériques de l'application : répertoires de chargement des séquences, activation des logs, etc. Ces paramètres lus sont communiqués aux autres modules si nécessaire.
- Module liant l'application avec l'interface QML destinée aux tests utilisateurs.
- Module permettant à l'expérimentateur de charger une pile d'image nécessaire aux tests

utilisateur via une interface graphique standardisée.

- Module de lecture des données brutes reçues par le LeapMotion™ (via les fonctions du LeapMotion SDK).
- Module interprétant les données reçues par le capteur. Ce dernier est chargé de reconnaître certaines séries de postures et/ou mouvement pour déterminer l'interaction qui doit être induite. Ce module gère de fait également la contextualisation, dans le cas notamment du vocabulaire gestuel défini dans la section 3.3.3, pour déterminer dans quel contexte est réalisé un mouvement de la main (i.e. : contexte de translation, zoom, etc.). Il traduit cette reconnaissance et association en instruction concrète en interaction (e.g. : par exemple, une translation d'une certaine amplitude) envoyé au module d'interface graphique qui aura à sa charge de réaliser le retour visuel approprié et accompagné des métaphores visuelles associées.
- Module chargeant des séquences interactives automatisées, et offrant une interface standardisée pour " jouer " ces séquences.

4.2.3 Interprétation de gestes

Les vocabulaires gestuels élaborés dans cette thèse, repose sur un mécanisme de geste déclencheur. En effet, comme expliqué dans les chapitres précédents (voir 2.3.1), l'adoption de certaines postures permet le basculement de l'application dans un " mode " d'interaction correspondant, dans lequel le mouvement de la main seront interprétés comme des entrées utilisateurs pour ce mode, suivant certaines contraintes de réalisation. Ce principe, de par sa nature est facilement implémentable comme montré dans la figure 4.3. En effet, dans les grandes lignes il s'agit de retranscrire le mécanisme énoncé en programme informatique :

- L'API LeapMotion™ est utilisée pour collecter les informations gestuelles nécessaires produites par l'utilisateur (position de la main, vitesse de mouvement, posture, etc.) par le module « collecteur des données »
- Si l'application n'est pas déjà dans un " mode d'interaction ", alors si une posture définie dans le vocabulaire gestuel couramment employé est adoptée, « l'interpréteur de gestes » fait basculer l'application dans le mode correspondant.
- Lorsque l'application est dans un mode d'interaction, l'interpréteur de gestes traduit les mouvements de la main en action à produire selon les contraintes liées à cette interaction et définie dans le vocabulaire. Par exemple, dans le cas où l'application est dans le mode " translation " dans le cas de notre vocabulaire pré-défini, établi dans 3.3.3, les mouvement de la main ne considérés que sur le plan parallèle à la table. Par conséquent, ce seront les coordonnées X et Y du déplacement qui moduleront les composantes du vecteur induisant la translation ; les variations sur la coordonnée Z (vers le haut/bas) sera ignorées.
- Les interactions interprétées sont alors traduites par le module gérant l'interface graphique et la production effective des actions utilisateurs.
- Lorsque le module interpréteur de gestes détecte qu'une posture de la main est abandonnée pour une posture inconnue ou neutre (poing fermé, ou main ouverte pour nos

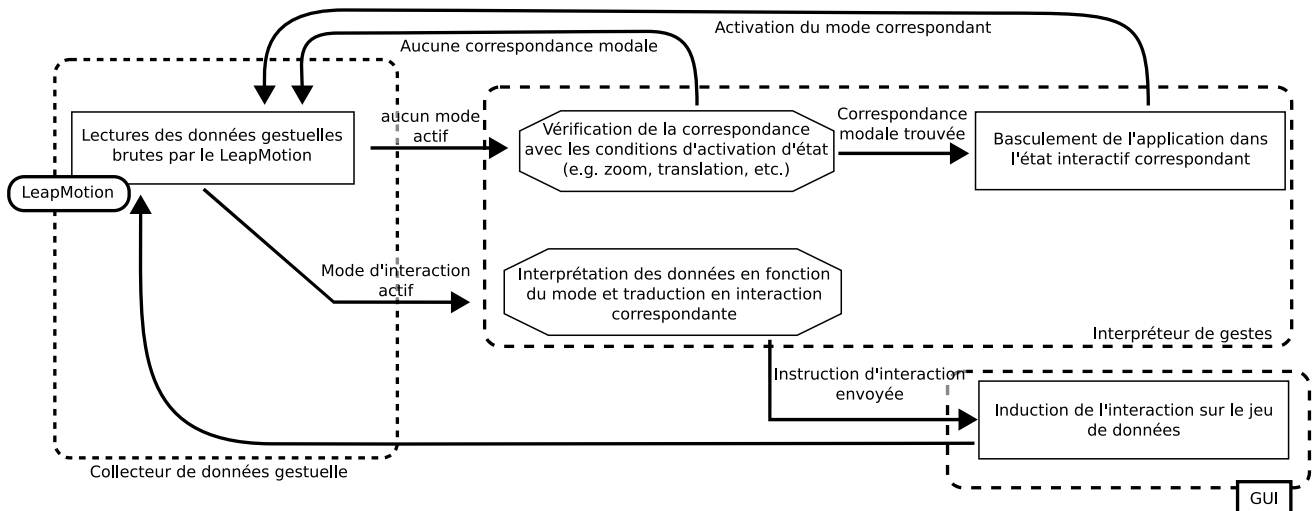


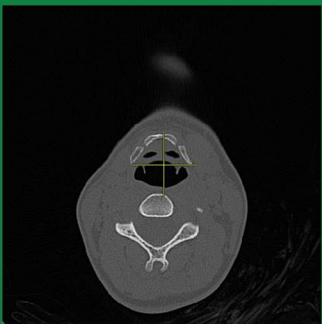
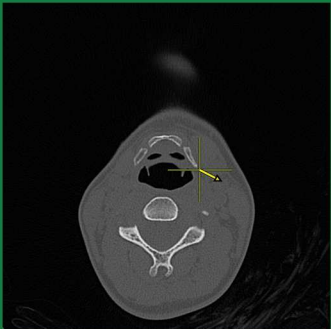
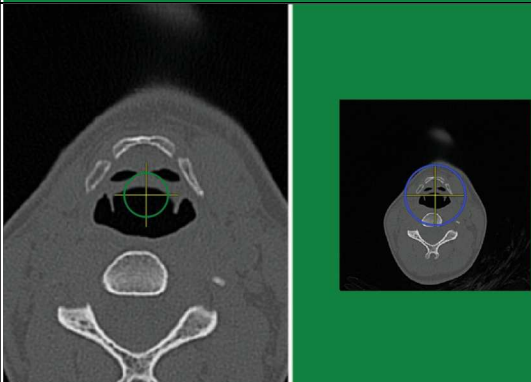
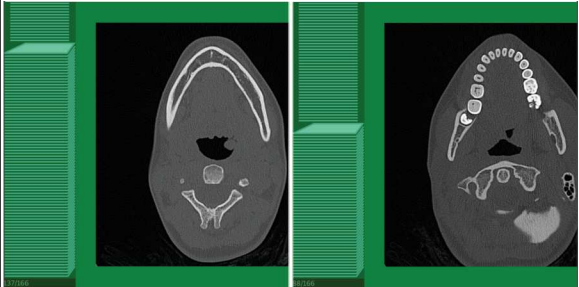
FIGURE 4.3 – *Processus de reconnaissance et interprétation des gestes utilisateurs en interaction.*

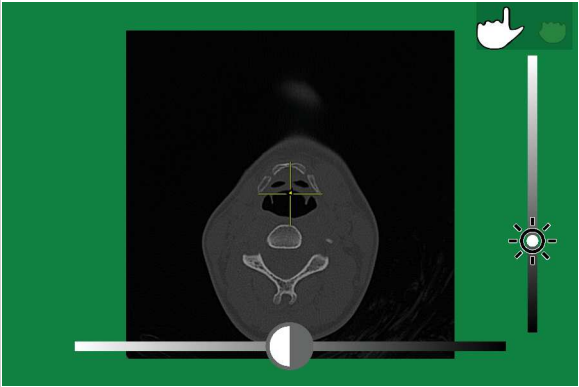
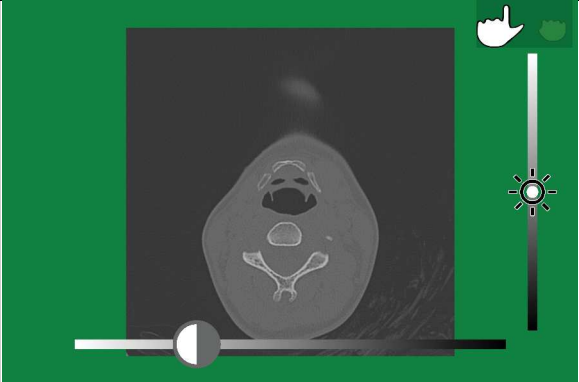
vocabulaires gestuels conçus), l'application abandonne le mode interactif courant. L'interpréteur n'interprétera alors plus les mouvements de la main comme une interaction à produire. À noter également que l'utilisateur peut transiter d'un mode d'interaction à un autre, sans passer par une posture et donc un mode neutre, en adoptant directement une nouvelle posture reconnue par le vocabulaire.

4.3 Interface

Comme détaillé précédemment, nous avons créé notre propre interface de visualisation pour nos tests utilisateurs. A contrario de travaux, évoqués dans des parties précédentes, qui utilisaient des systèmes de visualisation existants et normalisés, nous avons voulu avoir un plus grand contrôle sur l'interface, et plus particulièrement vis-à-vis des métaphores visuelles rendant compte des actions utilisateurs. En effet, nos expériences utilisateurs faisant appel à des utilisateurs « tout-venant » dans un premier temps, l'objectif fut de proposer des métaphores visuelles pour les actions réalisées qui soient aisément compréhensibles par de tels utilisateurs qui ne sont a priori pas expérimentés avec les interfaces de visualisations d'images médicales.

Le tableau ci-dessous liste les métaphores visuelles adoptées par l'interaction implémentée, dans la version du démonstrateur utilisé pour les tests utilisateurs :

	Repère : Un point repère sous forme de croix orthogonale. Celui-ci apparaît lors d'un zoom ou d'une translation. Son but est d'aider à mieux cibler la zone visée par la translation, et à mieux centrer précisément le zoom, ces deux interactions s'utilisant en pratique conjointement.
	Translation : Un vecteur se dessine lors d'une translation. Ce vecteur modélise la translation en guise de métaphore visuelle : la durée, la norme et la direction du vecteur varient respectivement en fonction de la durée, l'intensité, et la direction de la translation induite.
	Zoom : Un cercle centré se dessine lors d'un zoom. Ce cercle modélise le zoom en guise de métaphore visuelle : le rayon du cercle, et sa couleur dépendent respectivement de l'intensité (vitesse du zoom) et de la direction (bleu= avant vert= arrière) du zoom induit.
	Navigation dans la pile d'images : Une métaphore visuelle représentant une pile se dessine sur le côté gauche de l'écran. Cette représentation évolue en fonction de la position dans la pile courante. Le nombre d'image n'étant pas fixe, cette représentation est relative et non pas absolue. Un indicateur numérique est néanmoins présent (formaté <code>numero_image_courante</code> <code>nombre_total_images</code>).

	Gestion de la luminosité : Une métaphore visuelle sous forme de slider vertical. Le curseur positionné sur le slider représente le niveau de luminosité courant.
	Gestion du contraste : Une métaphore visuelle sous forme de slider horizontal. Le curseur positionné sur le slider représente le niveau de luminosité courant.
	Indicateur : Lorsque le vocabulaire présenté dans 3.3.3, a été intégré au capteur, nous avons rajouté un “ <i>indicateur de posture</i> ” à l’interface graphique. Celui-ci permet d’assurer que la posture utilisateur de la main est bien reconnue. La majorité de gestes y étant discriminés à la posture construite en fonction de l’extension des doigts, cet indicateur sert à rendre compte de quels doigts sont reconnus comme tendu par le capteur, permettant ainsi à l’utilisateur de s’assurer de la bonne détection de sa posture.

Remarque : Dans le tableau ci-dessus, les sliders agissant sur le contraste et la luminosité sont respectivement orientés verticalement et horizontalement.

L’interface graphique, composée des métaphores visuelles présentées ci-dessus, a pour but de remplir deux fonctions :

- Premièrement, rendre compte, auprès de l’utilisateur de l’état courant de la visualisation et de l’avancement de l’interaction que celui-ci est en train de réaliser.
- Deuxièmement, contextualiser les interactions. L’apparition d’éléments visuels spéci-

fiques à chaque interface, lorsque l'utilisateur entre dans " *mode d'interaction* " spécifique (associé à une interaction spécifique), permet de rendre compte du contexte d'interaction courant, voir d'un éventuel changement contexte. Cela est d'autant plus utile que cela permet accessoirement de signifier à l'utilisateur de la bonne compréhension par le système de son interaction (via sa posture).

4.4 Conclusion et récapitulatif des fonctionnalités

Pour pouvoir répondre aux besoins de nos expériences et protocoles (voir chapitre 3) une application de visualisation a été développée. Celle-ci nous a permis d'implémenter nos vocabulaires gestuels, tout en proposant une interface contextualisée offrant des métaphores visuelles adaptées pour les sujets « tout-venant ». En effet, le but a été d'offrir des retours visuels permettant à des utilisateurs n'étant pas nécessairement familiers avec le contexte de manipulation d'outils de visualisation d'images médicales. Nous avons donc utilisé des idiomes communs et des métaphores explicites pour rendre compte des interactions induites : par exemple la translation est représentée par un vecteur à l'écran (représentation courante du déplacement), la navigation dans la pile d'images par une représentation symbolique de celle-ci, la gestion du contraste ou de la luminosité (valeurs comprises dans un intervalle) par des sliders qui sont un idiome classique pour ces fonctionnalités dans d'autres logiciels. Seul le zoom ne bénéficie pas d'idiomes standardisés ou référencés, mais nous estimons qu'une codification par couleur (cercle vert pour le zoom arrière, bleu pour le zoom avant) couplés aux effets directs de l'interaction (images s'agrandissant / rétrécissant à l'écran) permet une bonne compréhension de cette interaction uni-dimensionnelle.

La dite application, en plus de l'interface graphique destinée aux sujets expérimentaux, est construite de façon modulaire de sorte à proposer plusieurs fonctionnalités :

- chargement d'une pile d'image au sein de laquelle il est possible de naviguer image par image.
- lecture automatisée d'une séquence d'interactions chargée et stockée via un fichier XML.
- chargement d'un vocabulaire gestuel décrit dans un fichier XML.
- interprétation des gestes réalisés par le sujet via le capteur LeapMotionTM et traduction en interactions conformément au vocabulaire gestuel sus-cité.

Le développement de ce " démonstrateur " nous a donc permis d'implémenter tous les aspects du protocole établi en 3.2, nous permettant ainsi de présenter aux sujets, aussi bien « tout-venant » que experts, un environnement interactif compréhensible, clair et contextualisé.

CONCLUSION

5.1 Bilan

Cette thèse est inscrite dans le cadre du projet 3D-Surg dont l'un des objectifs est la conception de " l'optimisation et le prototypage de nouveaux modèles d'interaction sans contact en vue d'une interaction en temps réel " pour une intégration dans le bloc opératoire. Cet objectif répond à une problématique particulière apportée par la consultation des données médicales en per-opératoire : la perte de stérilité du chirurgien lors de l'emploi de périphériques d'interaction classiques (clavier souris). La réponse à cette problématique dans notre cas est l'interaction sans contact, et plus précisément " l'interaction gestuelle en vol à mains nues ". Dès lors, nos travaux ont porté sur l'élaboration et l'implémentation d'une méthodologie de conception de vocabulaire gestuel intuitif adapté à une intégration dans le bloc opératoire.

Après avoir revu le cheminement et développement dans la recherche des systèmes d'interactions gestuelles depuis leur apparition jusqu'à leurs occurrences actuelles, nous nous sommes penchés sur les cas d'utilisation liés au bloc opératoire. Constatant que le concept a fait ses preuves lors de plusieurs études de cas, nous avons décidé d'aborder un problème alors encore inhérent à ces divers travaux : celui de l'utilisabilité. La recherche sur l'interaction gestuelle appliquée au domaine médical abonde en approches techno-centrées, alors que sur d'autres domaines elle tend vers des approches centrées sur l'utilisateur. En conséquence, dans le cadre de cette thèse, nous avons décidé de suivre les grandes lignes de la conception centrée utilisateur, tout en s'adaptant aux contraintes et contexte imposés par le cas d'usage qu'est la consultation de données médicales.

Suite à cela, nous avons effectué une analyse d'un cas pratique qui nous a été présenté : une opération de laparoscopie. Cette analyse nous a permis de bien comprendre le contexte d'utilisation ainsi que les exigences utilisateurs et organisationnelles. Outre une observation de l'environnement et des conditions d'usages, cela nous a également permis d'être confrontés aux logiciels employés. Dès lors nous en avons extrait des taxonomies d'interactions grâce auxquelles nous avons pu isoler les tâches élémentaires à couvrir par notre solution. Le but fut de substituer les interactions classiques (clavier souris), qui couvraient jusque-là la réalisation de ces tâches, par des interactions gestuelles adaptées et efficaces. C'est pourquoi nous avons déterminé un ensemble de critères à satisfaire : les critères contextuels liés à l'environnement d'utilisation et ses contraintes ; les critères conceptuels, dont fait partie l'utilisabilité, permettant une interaction efficace ; et enfin les critères techniques liés aux contraintes techniques issues du capteur utilisé. Pour chaque tâche élémentaire, nous avons alors itéré sur l'analyse et l'implémentation d'un grand nombre de solutions possibles afin de converger vers un panel de

gestes dont chaque geste, respectivement à sa tâche, satisfait au mieux les critères mentionnés ci-dessus.

Au-delà des critères de stricte performance, nous avons choisi de considérer un critère d'intuitivité et de cohérence d'utilisation partagée, qui, en cohérence avec les méthodes de conception UX et centrées utilisateurs, va de pair avec l'utilisabilité. C'est pourquoi nous avons implémenté une méthodologie de type " Magicien d'Oz ". Les méthodes de ce type consistent à observer les gestes que produisent naturellement les utilisateurs confrontés à la nécessité de proposer une interaction pour accomplir une tâche spécifique. Cette démarche permet d'extraire des vocabulaires gestuels " d'intuition " propre à chaque utilisateur. Il s'agit alors de filtrer les gestes recueillis en fonction de nos critères pour le synthétiser en un vocabulaire générique d'intuition. Cette démarche, nous l'avons alors employée en deux temps : une extraction de tels vocabulaires auprès d'utilisateurs " tout-venants " ; suivie d'une extraction auprès d'utilisateurs cibles. Le premier vocabulaire d'intuition extrait, celui des utilisateurs " tout-venants ", est alors confronté à celui des utilisateurs experts en vue de synthétiser une proposition finale de vocabulaire gestuel d'intuition. Nous avons alors appliqué ce qui constitue donc notre méthodologie auprès de tels groupes d'utilisateurs pour extraire un vocabulaire gestuel intuitif raffiné adapté à notre cas d'usage, en l'occurrence la consultation d'images de types CT-scan pour le bloc opératoire. Pour valider ce vocabulaire, nous l'avons présenté à un panel d'utilisateurs experts (c'est-à-dire des chirurgiens). Il a été perçu comme étant confortable et utilisable, tout en offrant un degré satisfaisant de contrôle sur les tâches présentées. De plus, à mesure que l'utilisateur se familiarise avec le vocabulaire, nous avons observé une diminution générale du taux d'erreurs ainsi que du temps d'exécution, ce qui démontre un effet positif d'apprentissage.

5.2 Limites & perspectives

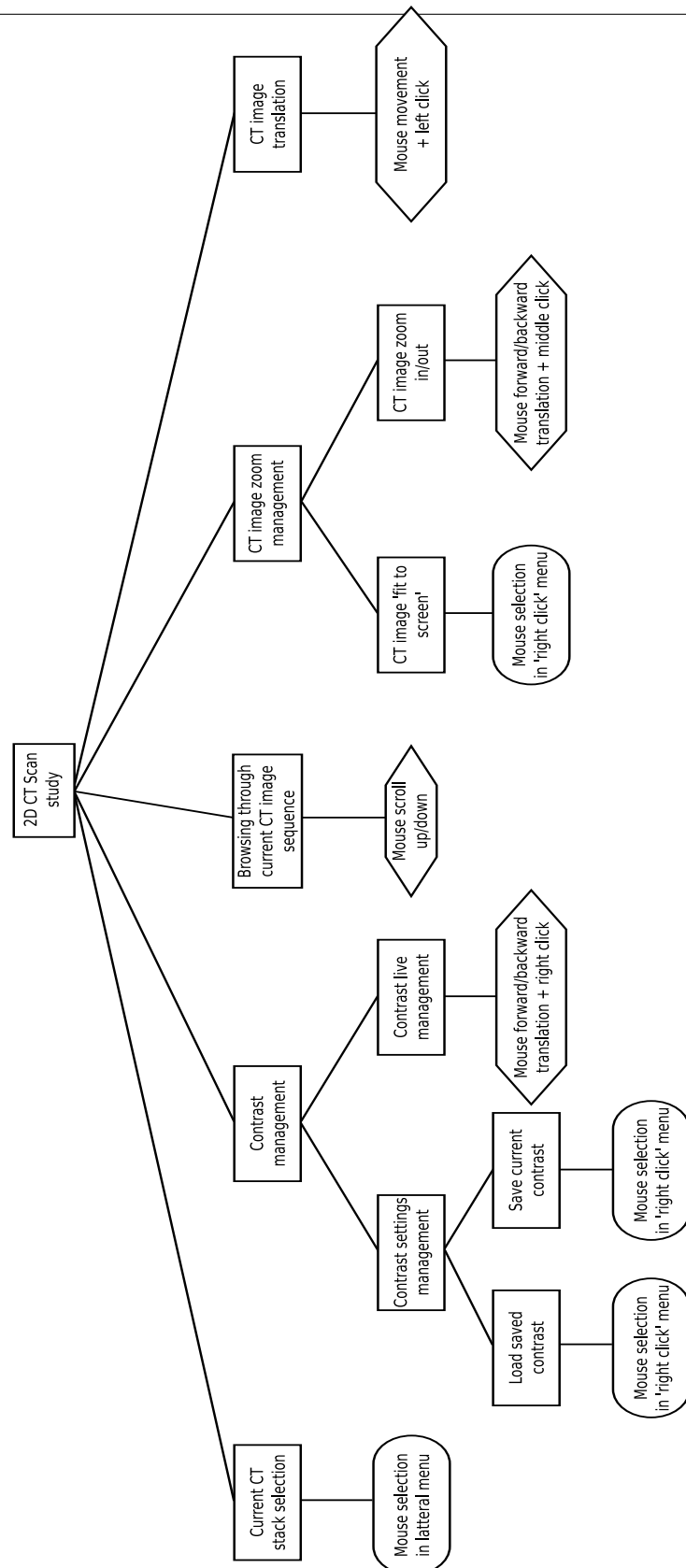
Pour compléter notre objectif, qui était de proposer une interaction sans contact intuitive et efficace à utiliser en per-opératoire par les chirurgiens, une perspective à court terme serait d'appliquer notre méthodologie à un autre cas d'utilisation, par exemple la consultation d'images 3D. Comme nous l'avons fait pour le cas d'utilisation 2D, nous devons valider les données extraites du vocabulaire gestuel intuitif découlant de la démarche proposée, en effectuant des tests d'utilisabilité sur des prototypes d'applications intégrant ces vocabulaires en tant que moyens interactifs impliquant les utilisateurs cibles qui sont des chirurgiens et le personnel des équipes chirurgicales.

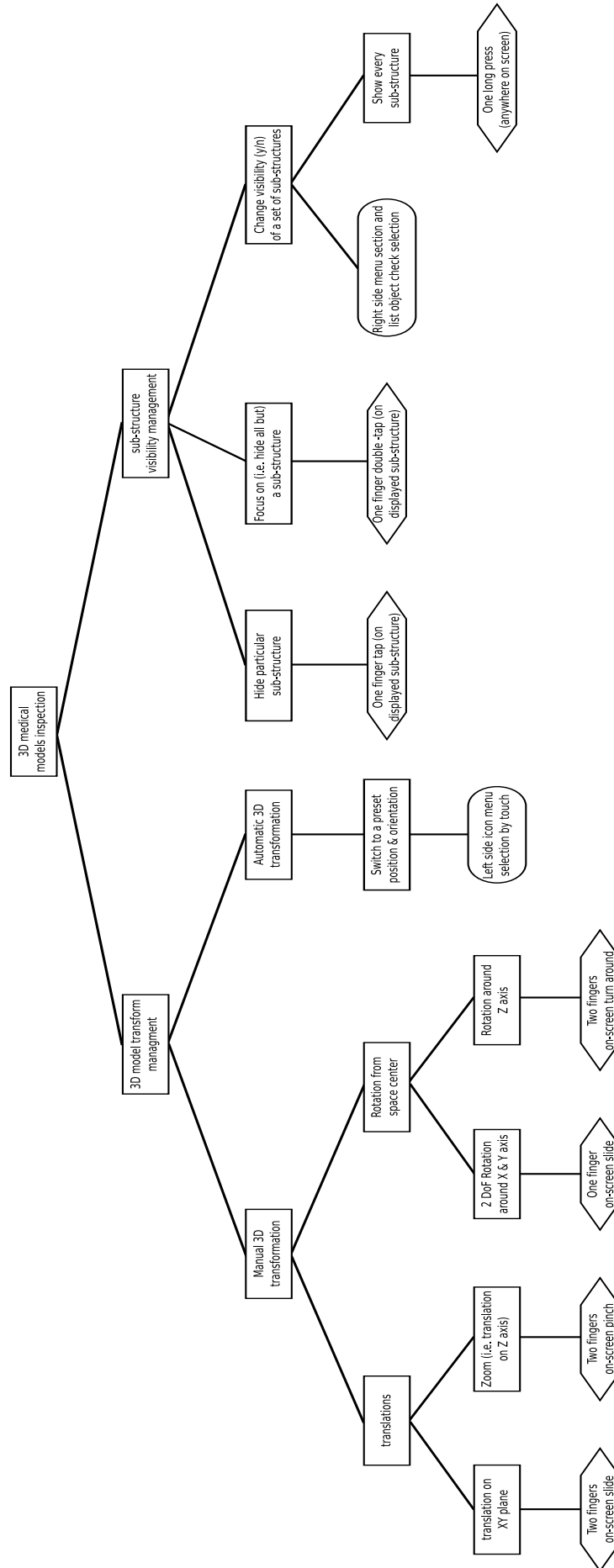
Bien que les travaux présentés ont été réalisés avec la participation de chirurgiens et que la proposition de vocabulaire gestuel a été approuvée par ces derniers, les limites de son intégration restent encore à explorer. En effet bien que concluant d'un point de vue conceptuel, il faudrait évaluer l'impact de l'intégration effective de cette interaction à moyen terme. Concrètement, notre méthodologie étant inspirée de méthodologies (UX, CCUs) itérative, il semblerait pertinent de réitérer une expérience de sollicitation d'utilisateurs cibles. Dès lors l'intégration effective au sein du bloc opératoire afin d'enrichir encore notre méthode, d'étapes post-

intégration pourraient ainsi générer une solution d'interaction bénéficiant d'un degré d'intuitivité et de pertinence encore plus élevé.

L'orientation du travail de cette thèse a été fortement contraint par le contexte et l'environnement d'utilisation. Comme expliqué dans le chapitre 2 cela a fortement influé sur nos critères contextuels et techniques. L'environnement et contexte nous ont poussé d'une part à restreindre notre vocabulaire à des gestes d'amplitude restreinte et d'autre part à considérer en priorité l'usage du capteur LeapMotionTM. En conséquence cela a fortement impacté l'élaboration de notre proposition de vocabulaire de départ (section 2.3), qui a servi de point de base à notre méthode de conception. Dès lors, il semblerait intéressant d'observer l'applicabilité de notre démarche dans des scénarios où les contraintes contextuelles sont différentes. Plus spécifiquement, on peut se demander si cette étape initiale de proposition d'un vocabulaire basé sur la technique et le contexte serait encore pertinente dans d'autres cas d'usage moins contraints, que se soit en changeant la nature des données consultées (type d'images, format), l'environnement d'application ou encore le domaine d'application.

TAXONOMIES





PROTOCOLE : ÉLÉMENTS ADDITIONNELS

B.1 Questionnaires

Ci-après, les questionnaires soumis aux sujets lors des itérations de nos protocoles expérimentaux. Il s'agit d'une transcription, les questionnaires présentés ici ont été remplis et soumis aux sujets sous la forme de questionnaires « *Google forms* »¹.

B.1.1 Profil utilisateur

Ce questionnaire est le questionnaire relatif au “ profil utilisateur ”. Il a été envoyé aux sujets inscrits à nos protocoles au préalable via mail, qui ont pu y répondre avant leur participation effective à l'expérience.

1. Age

2. Vous êtes...

- ☐ Gaucher
- ☐ Droitier
- ☐ Parfaitement ambidextre

3. Possédez-vous des appareils de type smartphone ou tablette?

- ☐ Les deux
- ☐ Oui, une tablette
- ☐ Oui, un smartphone
- ☐ Non

4. Si oui, comment évalueriez-vous votre fréquence d'utilisation de ces appareils?

très occasionnelle ☐—☐—☐—☐—☐ Intensive

5. Jouez-vous souvent à des jeux vidéo?

Jamais ☐—☐—☐—☐—☐ Très fréquemment

6. Si c'est le cas, lequel de ces 2 supports utilisez-vous le plus

- ☐ Mobile
- ☐ Consoles ou PC

1. Une version traduite en anglais des questionnaires a été réalisée mais n'a été utilisée que pour un seul cas de sujet cible

7. Avez-vous déjà utilisé des systèmes interactifs employant de la détection de mouvement (totale ou partielle du corps) sans contact (Kinect, ou LeapMotion)?
☐ Oui
☐ Non
8. Utilisez-vous des logiciels de traitement ou création d'images 2D matricielles ou vectorielles complets (Photoshop, TheGimp, Adobe Illustrator, etc.) et si oui, à quelle fréquence?
 Jamais ☐ — ☐ — ☐ — ☐ — ☐ Très souvent
9. Utilisez-vous des logiciels de traitement ou création d'images 3D (Blender, Maya, etc.) et si oui, à quelle fréquence?
 Jamais ☐ — ☐ — ☐ — ☐ — ☐ Très souvent
10. Utilisez-vous des logiciels de DAO/CAO (AutoCad, etc.) et si oui, à quelle fréquence?
 Jamais ☐ — ☐ — ☐ — ☐ — ☐ Très souvent

B.1.2 Expérience d'élicitation

Ce questionnaire a été soumis aux sujets à la suite de la phase " d'élicitation de gestes " de nos expériences utilisateurs.

1. Les tâches à effectuer étaient bien comprises
 D'accord ☐ — ☐ — ☐ — ☐ — ☐ Pas d'accord
2. J'avais une bonne compréhension des actions qui se produisaient à l'écran durant l'expérience
 D'accord ☐ — ☐ — ☐ — ☐ — ☐ Pas d'accord
3. Décrivez/expliquez le geste utilisé pour implémenter l'interaction de translation d'image. Aviez-vous une intuition ou référence en tête? Si oui laquelle?

4. Ce geste peut-être utilisé efficacement et confortablement de façon répétée/intensive sur de longues durée.
 D'accord ☐ — ☐ — ☐ — ☐ — ☐ Pas d'accord
5. Décrivez/expliquez le geste utilisé pour implémenter l'interaction de zoom sur une image. Aviez-vous une intuition ou référence en tête? Si oui laquelle?

6. Ce geste peut-être utilisé efficacement et confortablement de façon répétée/intensive sur de longues durée.
 D'accord ☐ — ☐ — ☐ — ☐ — ☐ Pas d'accord
7. Décrivez/expliquez le geste utilisé pour implémenter l'interaction de gestion du contraste de la luminosité. Aviez-vous une intuition ou référence en tête? Si oui laquelle?

-
8. Ce geste peut-être utilisé efficacement et confortablement de façon répétée/intensive sur de longues durée.

D'accord ☐—☐—☐—☐—☐ Pas d'accord

B.1.3 Test & évaluation de vocabulaire

Ce questionnaire a été soumis aux sujets à la suite de la seconde phase de nos protocoles qui consistait en une évaluation et comparaison de notre vocabulaire pré-défini, pour les sujets « tout-venants », ou l'évaluation de notre vocabulaire intermédiaire par les sujets cibles (voir 3.1).

1. Les tâches à effectuer étaient bien comprises (par moi)

D'accord ☐—☐—☐—☐—☐ Pas d'accord

2. Mes postures étaient bien reconnues/interprétés par le système

D'accord ☐—☐—☐—☐—☐ Pas d'accord

3. Il était facile de basculer entre les différentes interactions

D'accord ☐—☐—☐—☐—☐ Pas d'accord

4. J'ai fait l'expérience d'interférences entre les différents modes d'interactions

D'accord ☐—☐—☐—☐—☐ Pas d'accord

5. L'impact sur la qualité d'interaction entre ma main dominante et ma main non dominante était

Extrêmement faible ☐—☐—☐—☐—☐ Très fort

6. J'avais un contrôle satisfaisant sur la tâche de translation

D'accord ☐—☐—☐—☐—☐ Pas d'accord

7. Le geste/la posture rendait cette interaction (translation) facile à utiliser

D'accord ☐—☐—☐—☐—☐ Pas d'accord

8. Le geste/la posture rendait cette interaction (translation) confortable

D'accord ☐—☐—☐—☐—☐ Pas d'accord

9. Entre ce geste/posture et ce que vous avez proposé lors de l'expérience précédente, lequel vous semble le plus adapté? (translation)

Celui du vocabulaire ☐—☐—☐—☐—☐ Ma proposition

10. Expliquez et commentez cette réponse. Avez-vous des suggestions?

-
11. J'avais un contrôle satisfaisant sur la tâche de zoom

D'accord ☐—☐—☐—☐—☐ Pas d'accord

12. Le geste/la posture rendait cette interaction (zoom) facile à utiliser
D'accord ☐—☐—☐—☐—☐ Pas d'accord
13. Le geste/la posture rendait cette interaction (zoom) confortable
D'accord ☐—☐—☐—☐—☐ Pas d'accord
14. Entre ce geste/posture et ce que vous avez proposé lors de l'expérience précédente, lequel vous semble le plus adapté? (zoom)
Celui du vocabulaire ☐—☐—☐—☐—☐ Ma proposition
15. Expliquez et commentez cette réponse. Avez-vous des suggestions?
-
16. J'avais un contrôle satisfaisant sur la tâche de navigation dans la pile d'image
D'accord ☐—☐—☐—☐—☐ Pas d'accord
17. Le geste/la posture rendait cette interaction (navigation dans la pile d'image) facile à utiliser
D'accord ☐—☐—☐—☐—☐ Pas d'accord
18. Le geste/la posture rendait cette interaction (navigation dans la pile d'image) confortable
D'accord ☐—☐—☐—☐—☐ Pas d'accord
19. Entre ce geste/posture et ce que vous avez proposé lors de l'expérience précédente, lequel vous semble le plus adapté? (navigation dans la pile d'image)
Celui du vocabulaire ☐—☐—☐—☐—☐ Ma proposition
20. Expliquez et commentez cette réponse. Avez-vous des suggestions?
-
21. J'avais un contrôle satisfaisant sur la tâche de gestion du contraste et de la luminosité
D'accord ☐—☐—☐—☐—☐ Pas d'accord
22. Le geste/la posture rendait cette interaction (gestion du contraste et de la luminosité) facile à utiliser
D'accord ☐—☐—☐—☐—☐ Pas d'accord
23. Le geste/la posture rendait cette interaction (gestion du contraste et de la luminosité) confortable
D'accord ☐—☐—☐—☐—☐ Pas d'accord
24. Entre ce geste/posture et ce que vous avez proposé lors de l'expérience précédente, lequel vous semble le plus adapté? (gestion du contraste et de la luminosité)
Celui du vocabulaire ☐—☐—☐—☐—☐ Ma proposition
25. Expliquez et commentez cette réponse. Avez-vous des suggestions?
-

BIBLIOGRAPHIE

- [1] Baromètre du numérique 2018 (rapport credoc). <https://labo.societenumerique.gouv.fr/wp-content/uploads/2018/12/barometredunumerique2018.pdf>, 2018.
- [2] Ergonomie de l'interaction homme-système — partie 11 : Utilisabilité — définitions et concepts. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:fr>, 2018.
- [3] John Annett. Hierarchical task analysis. *Handbook of cognitive task design*, 2 :17–35, 2003.
- [4] Javier A Bargas-Avila and Kasper Hornbæk. Old wine in new bottles or novel challenges : a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2689–2698. ACM, 2011.
- [5] Bruce P Bernard and Vern Putz-Anderson. Musculoskeletal disorders and workplace factors ; a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. 1997.
- [6] Nicola Bizzotto, Alessandro Cozzano, Leonardo Bizzotto, Dario Regis, Andrea Sandri, and Bruno Magnan. Leap motion gesture control with osirix in the operating room to control imaging : first experiences during live surgery, 2014.
- [7] Marc-Eric BOBILLIER-CHAUMON, Sarah Carvallo, Franck TARPIN-BERNARD, and Jacqueline Vacherand-Revel. Adapter ou uniformiser les interactions personnes-systèmes ? *Revue d'Interaction Homme-Machine Vol*, 6(2), 2005.
- [8] Patricia Ivette Cornelio Martinez, Silvana De Pirro, Chi Thanh Vi, and Sriram Subramanian. Agency in mid-air interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2426–2439. ACM, 2017.
- [9] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies—why and how. *Knowledge-based systems*, 6(4) :258–266, 1993.
- [10] Lars C Ebert, G Hatch, MJ Thali, and Steffen Ross. Invisible touch—control of a dicom viewer with finger gestures using the kinect depth camera. *Journal of Forensic Radiology and Imaging*, 1(1) :10–14, 2013.
- [11] Luigi Gallo, Alessio Pierluigi Placitelli, and Mario Ciampi. Controller-free exploration of medical image data : Experiencing the kinect. In *2011 24th international symposium on computer-based medical systems (CBMS)*, pages 1–6. IEEE, 2011.
- [12] Chauncey Graetzel, Terry Fong, Sébastien Grange, and Charles Baur. A non-contact mouse for surgeon-computer interaction. *Technology and Health Care*, 12(3) :245–257, 2004.
- [13] Guillaume Gronier and Carine Lallemand. *Méthodes de design UX : 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles, 2015.

-
- [14] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. Consumed endurance : a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1063–1072. ACM, 2014.
- [15] Jean-François Jégou. *Interaction basée sur des gestes définis par l'utilisateur : Application à la réalité virtuelle*. PhD thesis, 2013.
- [16] Katherina A Jurewicz, David M Neyens, Ken Catchpole, and Scott T Reeves. Developing a 3d gestural interface for anaesthesia-related human-computer interaction tasks using both experts and novices. *Human factors*, page 0018720818780544, 2018.
- [17] Myron W Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. Videoplace—an artificial reality. In *ACM SIGCHI Bulletin*, volume 16, pages 35–40. ACM, 1985.
- [18] Piyush Kumar, Jyoti Verma, and Shitala Prasad. Hand data glove : a wearable real-time device for human-computer interaction. *International Journal of Advanced Science and Technology*, 43, 2012.
- [19] Effie Lai-Chong Law, Paul van Schaik, and Virpi Roto. Attitudes towards user experience (ux) measurement. *International Journal of Human-Computer Studies*, 72(6) :526–541, 2014.
- [20] Agata Manolova. System for touchless interaction with medical images in surgery using leap motion. *Organized by*, page 113, 2014.
- [21] Andre Mewes, Bennet Hensen, Frank Wacker, and Christian Hansen. Touchless interaction with software in interventional radiology and surgery : a systematic literature review. *International journal of computer assisted radiology and surgery*, 12(2) :291–305, 2017.
- [22] Christophe Mignot, Claude Valot, and Noelle Carbonell. An experimental study of future “natural” multimodal human-computer interaction. In *INTERACT’93 and CHI’93 Conference Companion on Human Factors in Computing Systems*, pages 67–68. ACM, 1993.
- [23] George A Miller. The magical number seven, plus or minus two : Some limits on our capacity for processing information. *Psychological review*, 63(2) :81, 1956.
- [24] Meredith Ringel Morris. Web on the wall : insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 95–104. ACM, 2012.
- [25] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, Jacob O Wobbrock, et al. Reducing legacy bias in gesture elicitation studies. *interactions*, 21(3) :40–45, 2014.
- [26] Michael Nielsen, Moritz Störring, Thomas B Moeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *International gesture workshop*, pages 409–420. Springer, 2003.
- [27] Donald A Norman. Natural user interfaces are not natural. *interactions*, 17(3) :6–10, 2010.
- [28] Kenton O’Hara, Gerardo Gonzalez, Abigail Sellen, Graeme P Penney, Andreas Varnavas, Helena M Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, et al. Touchless interaction in surgery. *Commun. ACM*, 57(1) :70–77, 2014.

-
- [29] Vladimir I Pavlovic, Rajeev Sharma, and Thomas S Huang. Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7) :677–695, 1997.
- [30] Francis KH Quek. Toward a vision-based hand gesture interface. In *Virtual reality software and technology*, pages 17–31. World Scientific, 1994.
- [31] Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 197–206. ACM, 2011.
- [32] G Ruppert, P Amorim, T Moares, and J Silva. Touchless gesture user interface for 3d visualization using kinect platform and open-source frameworks. In *Proceedings of the Fifth International Conference on Advanced Research in Virtual and Rapid Prototyping*, pages 215–219, 2011.
- [33] Douglas Schuler and Aki Namioka. *Participatory design : Principles and practices*. CRC Press, 1993.
- [34] Ben Shneiderman. 1.1 direct manipulation : a step beyond programming languages. *Sparks of innovation in human-computer interaction*, 17 :1993, 1993.
- [35] Édimo Sousa Silva and Maria Andréia Formico Rodrigues. Design and evaluation of a gesture-controlled system for interactive manipulation of medical images and 3d models. *SBC Journal on Interactive Systems*, 5(3) :53–65, 2014.
- [36] Helman I Stern, Juan P Wachs, and Yael Edan. Optimal hand gesture vocabulary design using psycho-physiological and technical factors. In *7th International Conference on Automatic Face and Gesture Recognition (FGRo6)*, pages 257–262. IEEE, 2006.
- [37] Matt Strickland, Jamie Tremaine, Greg Brigley, and Calvin Law. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Canadian Journal of Surgery*, 56(3) :E1, 2013.
- [38] Justin H Tan, Chong Chao, Mazen Zawaideh, Anne C Roberts, and Thomas B Kinney. Informatics in radiology : Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics*, 33(2) :E61–E70, 2013.
- [39] Ricardo Manuel Millán Vaquero, Jan Rzepecki, Karl-Ingo Friese, and Franz-Erich Wolter. Visualization and user interaction methods for multiscale biomedical data. In *3D Multiscale Physiological Human*, pages 107–133. Springer, 2014.
- [40] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. Gesture elicitation studies for mid-air interaction : A review. *Multimodal Technologies and Interaction*, 2(4) :65, 2018.
- [41] Juan Wachs, Helman Stern, Yael Edan, Michael Gillam, Craig Feied, Mark Smithd, and Jon Handler. Real-time hand gesture interface for browsing medical images. *International Journal of Intelligent Computing in Medical Sciences & Image Processing*, 2(1) :15–25, 2008.
- [42] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1083–1092. ACM, 2009.

-
- [43] Thomas G Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *ACM SIGCHI Bulletin*, volume 18, pages 189–192. ACM, 1987.

Approche pour une interaction intuitive et sans contact en bloc opératoire

Résumé

En chirurgie contemporaine, la visualisation des données médicales joue un rôle crucial, sous forme d'images 2D ou 3D, aussi bien en pré-opératoires, qu'en per-opératoires où la stérilité doit y être préservée à tout instant. Un système d'interaction sans-contact, faciliterait grandement la préservation de la stérilité. Ici nous détaillons notre approche méthodologique pour concevoir un tel système. Le cas d'usage pratique pris en référence par notre travail est une opération de laparoscopie et le logiciel couramment associé. Une analyse de la problématique, du cas d'usage et de sa contextualisation nous a permis d'élaborer une méthodologie, concrétisée en une phase exploratoire formalisée, et une phase expérimentale. Cette dernière, elle-même découpée en 2 temps: une étape visant des sujets " tout-venants " permettant d'extraire un vocabulaire d'intuitions génériques, ensuite affiné en utilisant les retours des sujets " experts ".

Interaction sans-contact ; Conception centrée utilisateur ; méthode d'élicitation de gestes

Résumé en anglais

In modern surgery, the visualization of medical data plays a crucial role, in the form of 2D or 3D images, both in per-op and pre-op, where sterility must be preserved at all times. A touchless interaction system would greatly facilitate the preservation of sterility. Here we detail our methodological approach to design such a system. The practical use case referred to in our work is a laparoscopic operation and the case of use associated software. After analyzing the problem, the use case and its contextualization we developed a methodology, which was concretized in a formal exploratory phase and an experimental phase. The latter, itself divided into two stages: a stage aimed at "all-coming" subjects allowing us to extract a generic intuition vocabulary, which is then refined using feedback from "expert" subjects.

Touchless interaction; User Centered Design ; gestures elicitation