HAL
open science

Interaction basée sur des gestes définis par l'utilisateur : Application à la réalité virtuelle

Jean-François Jego

To cite this version:

Jean-François Jego. Interaction basée sur des gestes définis par l'utilisateur : Application à la réalité virtuelle. Autre [cs.OH]. Ecole Nationale Supérieure des Mines de Paris, 2013. Français. NNT : 2013ENMP0033 . pastel-00944726

HAL Id: pastel-00944726

https://pastel.hal.science/pastel-00944726v1

Submitted on 11 Feb 2014

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.

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

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

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité « Informatique, temps réel, robotique, automatique »

présentée et soutenue publiquement par

Jean-François JEGO

Interaction basée sur des gestes définis par l'utilisateur : Application à la réalité virtuelle

Directeur de thèse : **Philippe FUCHS**
Co-directeur de thèse : **Alexis PALJIC**

Jury :

Annelies BRAFFORT, Directeur de Recherche, LIMSI, CNRS UPR 3251, Université Paris XI	Président
Pierre De LOOR, Professeur d'Université, Lab-STICC, CNRS UMR 6285, École Nationale d'Ingénieurs de Brest	Rapporteur
Indira THOUVENIN, Enseignant-Chercheur, Heudiasys, CNRS UMR 7253, Université de Technologie Compiègne	Rapporteur
Frédéric BEVILACQUA, Chargé de Recherche, STMS, CNRS UMR 9912, IRCAM-Centre Pompidou	Examinateur
Philippe FUCHS, Professeur, Centre de Robotique, Mines ParisTech	Examinateur
Alexis PALJIC, Enseignant-Chercheur, Centre de Robotique, Mines ParisTech	Examinateur
Jean-Paul DEPARTE, Ingénieur, Laboratoire Électronique, CRMMF de Kerpape	Invité

T H È S E

MINES ParisTech
Mathématiques et systèmes
60 Bd Saint-Michel 75272 Paris cedex 06

RÉSUMÉ

Ces travaux de recherche proposent une nouvelle méthode d'interaction gestuelle. Nous nous intéressons en particulier à deux domaines d'application : la thérapie à domicile par la réalité virtuelle et les arts scéniques numériques. Nous partons du constat que les interfaces standardisées ne sont pas adaptées à ces cas d'usage, car elles font appel à des gestes prédéfinis et imposés. Notre approche consiste à laisser la possibilité à l'utilisateur de faire apprendre ses gestes d'interaction au système. Cela permet de prendre en compte ses besoins et aptitudes sensorimotrices. L'utilisateur réutilise ensuite son langage gestuel appris par le système pour interagir dans l'environnement virtuel. Cette approche pose des questions de recherche centrées sur la mémorisation, la rétroaction visuelle et la fatigue. Pour aborder ces trois aspects, nous étudions d'abord le rôle de l'affordance visuelle des objets et celui de la colocalisation dans la remémoration d'un vocabulaire de gestes. Ensuite, nous évaluons l'influence de différents types de rétroactions visuelles sur l'évolution des gestes répétés par l'utilisateur dans une série de tâches de manipulation. Nous comparons également les performances entre des gestes d'amplitude réaliste et des gestes d'amplitude faible pour effectuer la même action. Aussi, nous attachons une importance à rendre l'interface accessible en utilisant des dispositifs bas coûts et peu intrusifs. Nous explorons les moyens de pallier les contraintes techniques liées aux systèmes peu performants. Pour cela, nous avons conduit des expériences où plus de six mille gestes proposés par une quarantaine d'utilisateurs ont été étudiés.

Mots clés : réalité virtuelle, facteur humain, interaction gestuelle, 3DUI, apprentissage artificiel

ABSTRACT

In this thesis, we propose and evaluate new gestural interfaces for 3DUI. This work is motivated by two application cases: the first one is dedicated to people with limited sensory-motor abilities for whom generic interaction methods may not be adapted. The second one is artistic digital performances, for which gesture freedom is part of the creative process. For those cases, a standardized approach is not possible and thus user-specific or dedicated interfaces are needed. We propose a user-defined gestural interaction that allows the user to make the system learn the gestures that he has created, in a specific phase, prior to using the system. Then, the user reuses his created gestures to interact in the virtual environment. This approach raises research questions about memorization of gestures, effects of fatigue and effects of visual feedbacks. To answer those questions, we study the memorization of user created gestures regarding the role of affordances and colocalization on gesture recall. Then, we study the role of different visual feedbacks on gesture repetitions for a set of manipulation tasks. We also compare full-collocated gestures to loose gestures with lower amplitude. Also, the approach has been designed to be affordable using low-cost devices. We explore solutions to deal with the lower data quality of such devices. The results of the user-studies are based on the analysis of six thousand gestures performed by forty subjects.

Keywords: virtual reality, human factors, gestural interaction, 3DUI, machine learning

Remerciements

Mes premiers remerciements sont destinés aux membres du jury. Je commence par Pierre De Loor et Indira Thouvenin, rapporteurs de ce manuscrit, en les remerciant pour leur travail de relecture méticuleux et leurs commentaires très constructifs. Je remercie également Frédéric Bevilacqua, Annelies Braffort et Jean-Paul Departe d'avoir accepté d'examiner mes travaux avec enthousiasme et attention. Mes remerciements les plus sincères à Philippe Fuchs qui a accepté de diriger ces travaux : je confirme, cela a été un honneur, merci pour ton œil plus qu'avisé avec toujours un temps d'avance que j'espère pouvoir acquérir un jour. Merci à Alexis Paljic, codirecteur de ces recherches : ta disponibilité, ton écoute et tes conseils, j'ai beaucoup appris avec toi et toujours dans la bonne humeur, je n'oublierai jamais cette formidable collaboration. J'adresse aussi mes remerciements à Claude Laurgeau et Arnaud de la Fortelle, fondateur et directeur du Centre de Robotique de Mines ParisTech pour l'accueil et leur confiance pendant ces belles années au laboratoire. Un grand merci à Jean-Marc Goachet et Thibaud Vaerman pour le temps accordé aux reportages vidéo très pros (et qui font aussi de très bons souvenirs). Mes pensées vont à Erik Geslin qui m'a permis mes premiers pas dans la "VR", merci à toi ; à Stéphane Bouchard pour la collaboration très humaine et toujours avec la joie de vivre. Merci à Marie-Hélène Tramus pour la confiance et pour m'avoir permis d'effectuer des vacations d'enseignement au sein de la formation Arts et Technologies de l'Image si stimulante. Pour les mises en relation et les échanges toujours constructifs, merci à Evelyne Klinger. Merci également à Sophie Lavaud pour l'expérience enrichissante et avoir fait vivre l'installation aux quatre coins du monde. Pour les petits conseils et toujours avec humour, merci à Patrick Callet. J'adresse un grand merci aux participants des prétests et des expérimentations pour leurs disponibilités et leur patience. Merci également aux patients volontaires du CMRRF de Kerpape pour leur enthousiasme et leur disponibilité. Ha mes courageux relecteurs : Anne-Sophie Puthon, Clément Larrivé, Elsa Wolff, Guillaume Bertinet, Judith Guez, Leslie Guilbot, Raoul de Charette, Soizig Le Stradic, Tom Giraud, Tristan Benoist, un gros merci pour la minutieuse chasse à la coquille ! Je pense aussi à ceux sans qui je n'aurais pas persévéré : mes acolytes et co-créateurs, Florian, Adrien et Dimitri pour les projets fous qui ont vu le jour (et sont à venir !), les nombreux soutiens malgré la distance : Jane, Sarah, Anthony (RedWings en force) et amis du Québec, une pensée pour toi Céline. Mes soupapes de décompression en support 24/7 : dédicaces particulières aux potes^{MX} et de très longue date : que serais-je sans vous mes amis ? Anne-Laure et ses anlorismes qui me font tant rire, Clément toujours dispo quand il faut et avec un humour culbutant^{TM}, notre ange Éric et son âme plus que charitable, Maxime et ses méta-melons, Soizig la Brigitte Bardot du pistil à l'énergie débordante et communicative, Tristan toujours motivé et motivant, mais jamais à l'heure et jamais à l'heure à la fois ! Je pense aussi aux moments de bien mettance avec Clément, Marie, Nico, Pierre, Sonia, Yoann. Clin d'œil à mes chers encourageurs Judith (enactive-ducky), Tom (santé/karma !), Raoul (bécots thésocentrés) et Vincent (aka VW qui bosse en fait pour PSA, et toujours dispo quand il faut). « Allez viens manger », « cafet' ? », « ministère ?», heureusement que les collègues sont là pour lever la tête du guidon. Quatre ans de fun déjà, merci à mes voisins de bural et psychologues malgré eux : Alina, Ben, David, Éva, Florent, JE, Jean, Jorge, Lucile, Martin, Martyna, Max, Olivier, Olivier (ils se reconnaitront), Raphaëlle, Sylvain (ça roule ? Ok cette blague n'est pas drone), Tao-Jin, Philippe, Sylvère, Victor. Pour l'accueil et la bonne ambiance, merci à Bogdan, Clément, Fawzi, Fabien, François, Joël, Sébastien et les anciens Ayet, Fatin, Laure, Lghani, Keerthi, Nan, Oussama, Omar, Taha, Vincent. Pour votre aide importante et nos discussions passionnantes, un grand merci à Christine, Christophe, David, Jacky et Sotiris. Dédicaces aux sportifs du Luco avec qui j'ai eu plaisir à user mes galoches : Amaury, Bruno, Fernando et Yu-Fei. Merci enfin à Jérôme et Stéphanie pour le soutien (les week-ends nantais ont été plus que ressourçant pour le p'tit). Je termine par mes parents, Evelyne et Michel, qui m'ont toujours épaulé dans mes aspirations : ne pouvant vous rendre autant que vous m'avez donné, il est naturel que je transmette à mon tour vos si belles valeurs de respect, de courage et de partage.

TABLE DES MATIÈRES

CHAPITRE 1 INTRODUCTION 1

1.1. LA RÉALITÉ VIRTUELLE, PLUS ACCESSIBLE ? 2
1.1.1. Vers une interaction plus « naturelle » 2
1.1.2. L'usage récent du geste comme modalité d'interaction 5
1.1.3. Besoins d'adaptation aux usages et aux usagers 6
1.2. CONTEXTE DE L'INTERACTION GESTUELLE ADAPTÉE EN RÉALITÉ VIRTUELLE 8
1.2.1. Contexte local 8
1.2.2. Réalité virtuelle et rééducation fonctionnelle 8
1.2.3. La réalité virtuelle support des arts scéniques numériques 9
1.3. POSITIONNEMENT DE LA THÈSE 12
1.3.1. Objectifs centrés sur l'adaptation 12
1.3.2. Problématique de recherche 12
1.3.3. Organisation du manuscrit 14

CHAPITRE 2 ÉTAT DE L'ART SUR L'INTERACTION GESTUELLE ADAPTÉE EN ENVIRONNEMENT VIRTUEL 15

2.1. UNE INTERACTION ADAPTÉE EN ENVIRONNEMENT VIRTUEL 16
2.1.1. L'adaptation de l'interaction pour l'expérience utilisateur 16
2.1.1.a Considérations sur l'adaptation 16
2.1.1.b Acquisition et réutilisation des connaissances 17
2.1.2. Principales approches sur l'adaptation de l'interaction 19
2.1.2.a Utilisateur ← Système : l'interaction standardisée 19
2.1.2.b Utilisateur ⇆ Système : l'interaction adaptable ou adaptative 21
2.1.2.c Utilisateur → Système : l'utilisateur définit l'interaction 23
2.1.3. Concevoir une expérience centrée utilisateur 26
2.1.3.a Principes généraux de conception centrée-utilisateur 26
2.1.3.b Principes de conception centrée-utilisateur en réalité virtuelle 29
2.1.4. Synthèse 31
2.2. CARACTÉRISER ET RECONNAITRE LE GESTE POUR INTERAGIR EN ENVIRONNEMENT VIRTUEL 32
2.2.1. Caractériser la gestualité de l'homme 32
2.2.1.a Caractérisations générales du geste 33
2.2.1.b Systèmes de représentation et de transcription du geste 35
2.2.1.c Segmentation en unités gestuelles 37
2.2.2. L'interaction gestuelle par apprentissage 37
2.2.2.a Acquisition du geste 39
2.2.2.b Représentation des données : traitement et descripteurs gestuels 42
2.2.2.c Classification et reconnaissance du geste 43
2.2.2.d Interprétation du geste et rétroaction 48
2.2.3. Considérations sur l'usage de l'interaction gestuelle 49
2.2.3.a Interaction et rétroaction 49
2.2.3.b Ergonomie du geste en interaction 51
2.2.3.c Variabilité et comparaison de gestes 51
2.2.4. Synthèse 53

CHAPITRE 3 CONCEPTS SUR L'APPROCHE OÙ L'UTILISATEUR DÉFINIT L'INTERACTION GESTUELLE 55

3.1. L'UTILISATEUR DÉFINIT L'INTERACTION GESTUELLE 56
3.2. PRINCIPALES PROBLÉMATIQUES SOULEVÉES PAR L'APPROCHE UDIG 58
3.2.1. Répétabilité du geste et variabilité 58
3.2.2. Mémorisation et charge cognitive 59
3.2.2.a Charge cognitive et coordination motrice 59
3.2.2.b Mémoire à court terme et empan mnésique 59

3.2.3. Rétroactions et couplage 60
3.2.3.a Le choix de la rétroaction visuelle 60
3.2.3.b L'interaction en continu 60
3.2.3.c Colocalisation 61
3.2.4. Considérations techniques pour la mise en œuvre de l'UDIg 61
3.2.4.a Acquisition avec des systèmes peu performants 61
3.2.4.b Interprétation avec des systèmes peu performants 62

CHAPITRE 4 EXPÉRIMENTATION SUR LA REMÉMORATION DE GESTES PROPOSÉS PAR L'UTILISATEUR 65

4.1. QUESTION DE RECHERCHE ET HYPOTHÈSES 66
4.1.1. Hypothèse A : affordances des objets 66
4.1.2. Hypothèse B : contraintes et colocalisation 68
4.1.3. Hypothèse C : mémorisation et empan mnésique 69
4.2. L'EXPÉRIMENTATION 70
4.2.1. Tâche 70
4.2.2. Protocole 70
4.2.2.a Conditions expérimentales 71
4.2.2.b Dispositif expérimental 72
4.2.3. Métriques de comparaison 75
4.2.4. Population 75
4.2.5. Prétests 75
4.3. RÉSULTATS 77
4.3.1. Résultats hypothèse A : influence de l'affordance 78
4.3.2. Résultats hypothèse B : influence de la colocalisation 78
4.3.3. Résultats hypothèse C : influence du nombre de gestes 79
4.3.4. Limites et biais possibles 79
4.4. DISCUSSION ET CONCLUSION 81

CHAPITRE 5 EXPÉRIMENTATION SUR LA RÉPÉTABILITÉ DU GESTE ET LES RÉTROACTIONS 83

5.1. QUESTIONS DE RECHERCHE ET HYPOTHÈSES 84
5.1.1. Rétroactions visuelles 85
5.1.1.a Hypothèse A : influence du couplage de la rétroaction sur le geste à l'interaction 86
5.1.1.b Hypothèse B : influence sur le geste du moment d'apparition de la rétroaction couplée 86
5.1.1.c Hypothèse C : influence sur le geste des degrés d'informations perçues en continu 87
5.1.2. Répétabilité 87
5.1.2.a Hypothèse D : existence d'un geste de faible amplitude basé sur un schème 87
5.1.2.b Hypothèse E : influence de la répétition sur le geste à l'interaction 88
5.2. L'EXPÉRIMENTATION 89
5.2.1. Tâche 89
5.2.2. Protocole 90
5.2.2.a Choix des rétroactions 90
5.2.2.b Choix des objets à manipuler 92
5.2.2.c Dispositif expérimental 93
5.2.2.d Conditions d'ergonomie 94
5.2.3. Métriques de comparaison 95
5.2.4. Population 95
5.3. RÉSULTATS 96
5.3.1. Rétroactions visuelles 97
5.3.1.a Résultats hypothèse A : influence du couplage dans l'interaction 97
5.3.1.b Résultats hypothèse B : influence sur le geste du moment d'apparition de la rétroaction couplée 98
5.3.1.c Résultats hypothèse C : influence sur le geste des degrés d'informations perçues en continu 99
5.3.2. Répétabilité et fatigue 102
5.3.2.a Résultats hypothèse D : existence d'un geste de faible amplitude basé sur un schème 102
5.3.2.b Résultats hypothèse E : influence de la répétition sur le geste à l'interaction 104
5.3.3. Limites et biais possibles 105
5.4. DISCUSSION ET CONCLUSION 107

CHAPITRE 6 CONCLUSION ET PERSPECTIVES 109

6.1. CONCLUSION SUR LA MISE EN PLACE DE L'APPROCHE UDIG 110
6.1.1. Comment faciliter la proposition et la remémoration des gestes ? 110
6.1.2. Quel retour sur l'action et quand ? 110
6.1.3. Comment permettre la répétabilité du geste ? 112
6.1.4. Considérations sur les technologies peu performantes 113

6.2. PERSPECTIVES D'USAGE HORS LABORATOIRE ET LIMITES 115
6.2.1. Prototype « Haut la main ! » à destination du grand public 116
6.2.1.a L'appréhension de l'approche par les usagers 116
6.2.1.b Nature des gestes proposés 117
6.2.1.c Rétroaction en avance 118
6.2.2. Prototype proposé aux patients du centre de rééducation fonctionnelle de Kerpape 118
6.2.2.a L'appréhension de l'approche par les usagers 119
6.2.2.b Utilisation par les patients 120
6.2.2.c Considérations technologiques avec du matériel peu performant 122
6.2.2.d Perspectives d'usage en rééducation 123

6.3. RÉCAPITULATIF DES PRÉCONISATIONS D'USAGE DE L'APPROCHE OÙ L'UTILISATEUR DÉFINIT L'INTERACTION GESTUELLE 124

6.4. PERSPECTIVES 127

BIBLIOGRAPHIE 129

TABLES DES ILLUSTRATIONS 141

TABLE DES FIGURES 142
LISTE DES TABLES 144

APPENDICES 145

ANNEXE A TABLEAU COLLABORATIF DESCRIPTIF DES CAPTEURS GRAND PUBLIC SELON L'USAGE ET LES SPÉCIFICATIONS TECHNIQUES 146
ANNEXE B MATÉRIEL ET RÉSULTATS DÉTAILLÉS EXPÉRIMENTATION 1 148
ANNEXE C MATÉRIEL ET RÉSULTATS DÉTAILLÉS EXPÉRIMENTATION 2 150
ANNEXE D TESTS UTILISATEURS GRAND PUBLIC FUTUR EN SEINE 2013 155

CHAPITRE 1
INTRODUCTION

« Indeed, the important thing about gestures is that they are not fixed. They are free and reveal the idiosyncratic imagery of thought. »

David McNeill,
Hand and Mind: What Gestures Reveal About Thought, 1992

« The machine was rather difficult to operate. For years, radios had been operated by means of pressing buttons and turning dials; then, as the technology became more sophisticated, the controls were made touch sensitive […] now all you had to do was wave your hand in the general direction of the components and hope. It saved a lot of muscular expenditure of course, but meant you had to stay infuriatingly still if you wanted to keep listening to the same programme. »

Douglas Adams,
The Hitch Hiker's Guide to the Galaxy, 1979

1.1. LA RÉALITÉ VIRTUELLE, PLUS ACCESSIBLE ?

1.1.1. Vers une interaction plus « naturelle »

L'homme, grâce au potentiel de la réalité virtuelle, a la possibilité de s'échapper de sa réalité quotidienne, pour des raisons aussi bien artistiques que culturelles ou professionnelles (Fuchs, Berthoz, et al. 2006). L'immersion et l'interaction, les deux clés de voûte permettant de vivre cette expérience, doivent pouvoir se faire de façon immédiate et sans complexité d'utilisation. L'*Usage* implique ainsi une prise en main simple, efficace et qui n'oblige pas à suivre un mode d'emploi (laissons les manuels d'utilisation dans le monde réel !). L'expérience doit être accessible et transparente, en somme « naturelle ».

L'interface utilisateur « naturelle »

On constate depuis la dernière décennie l'apparition de nombreuses interfaces dites « naturelles » (*Natural User Interfaces* ou *NUI*), que ce soit dans des dispositifs de réalité virtuelle ou de jeu vidéo ou en Interaction Homme Machine (IHM). On peut effectuer des tâches primitives à priori différentes, par exemple la sélection ou l'édition pour une IHM, et la manipulation en réalité virtuelle (Card et al. 1983; Fuchs, Berthoz, et al. 2006). Comme le note (Lee 2010) ces deux domaines se concentrent sur l'activité ; le principe d'une interface étant de permettre à l'utilisateur de concrétiser une idée, tout en minimisant l'effort pour y parvenir. Il existe ainsi des points de convergence entre les modalités d'interaction de ces deux domaines qui tendent à exploiter des interfaces plus naturelles.

Or, en pratique, (Norman 2010) considère que ces interfaces naturelles ne le sont pas nécessairement d'une personne à une autre, surtout si elles sont de culture différente. Pour (Bird 2011), « tout l'enjeu est de caractériser ce qui fait le propre d'une interaction vraiment naturelle : cela implique de croiser différents domaines comme l'art, le design, l'IHM, l'informatique et la psychologie ». Norman évoque, dès la fin des années 80, un potentiel ordinateur du futur, transparent, où l'utilisateur effectue des tâches plutôt que d'utiliser une machine, qui serait devenue invisible. Afin qu'il soit adapté à l'utilisateur, à ses besoins et possibilités, il propose que la conception devrait prendre en considération l'homme et se centrer sur les besoins utilisateurs (Norman 1989a). C'est bien le but des systèmes de réalité virtuelle qui permettent à l'homme de réaliser des activités dans un monde artificiel qui peut être imaginaire, symbolique ou de simuler certains aspects du monde réel (Fuchs et al. 2003).

L'homme étant au cœur du dispositif, les auteurs considèrent naturellement que la conception du monde virtuel soit centrée sur l'homme et sur son activité in virtuo, avec une réelle transparence de l'interface. Une interface est transparente si l'utilisateur oublie son action sensori-motrice pour se concentrer au niveau cognitif sur la tâche à accomplir. Les auteurs parlent ainsi d'interfaces comportementales. Outre le fait que ces interfaces doivent s'affranchir des contraintes matérielles, elles doivent aussi exploiter un comportement humain, naturel et sans acquis préalables. Cela s'accorde aux propos de (Lee 2010) relatifs au consensus sur la définition du terme « naturel », mettant en avant l'intuitivité et l'invisibilité de l'interface. Fuchs et al. préfèrent employer le terme « pseudo-naturelle » car l'interface reste basée sur la technologie.

Historique d'une interaction centrée sur l'activité

En remontant à l'origine des interfaces, des systèmes construits par besoin scientifique aux interfaces plus accessibles et ubiquitaires d'aujourd'hui, on constate que l'interaction est passée d'une complexité réservée à des outils experts à une simplification extrême (Valli 2008). En parallèle, la simulation puis la réalité virtuelle ont toujours fait appel à des situations exploitant des comportements et des gestes orientés vers l'activité, à travers des dispositifs concrets. Citons par exemple le gant de données, le volant ou le manche de pilotage de simulateur de transport. En observant l'intégration du geste dans les systèmes informatiques, on distingue plusieurs grandes ruptures où émerge le mouvement à la place de traditionnelles actions et contrôles par boutons de commande, et des interfaces en ligne de commande (*CLI* pour *Command Line Interface*), (Figure 1-1).

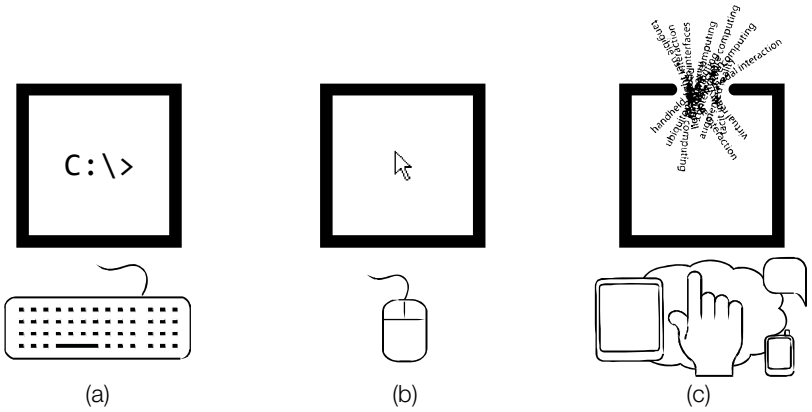


Figure 1-1 : L'interaction en trois générations, : (a) lignes de commande (CLI), (b) manipulation directe (GUI), (c) styles d'interactions émergents et basés sur le monde réel (NUI), adapté de (Jacob et al. 2008)

Dans un souci d'accessibilité de l'outil informatique, (Engelbart et al. 1968) proposent un concept pour faciliter et organiser le travail individuel et collaboratif. Dans la démonstration baptisée plus tard *The Mother Of All Demos*, ils présentent un dispositif qui deviendra la souris, permettant la manipulation directe de l'information sur un écran. Cette démonstration propose aussi d'autres avancées centrées sur l'usage comme la métaphore du bureau, la visioconférence, le courrier électronique et l'hypertexte définissant les bases du paradigme *WIMP* (acronyme anglais pour fenêtre, icône, menu, pointeur). On assiste ainsi à la première évolution de l'interaction qui emploie l'usage des métaphores et incite l'utilisateur à explorer l'interface pour accéder à des fonctionnalités étendues à travers une interface graphique (*GUI*). Il est à noter que ce paradigme demeure toujours à l'heure actuelle majoritairement employé pour la production de contenu numérique (Lee 2011).

Pour en revenir aux environnements virtuels, dès le milieu des années 1970, Krueger et al. créent l'installation expérimentale *Videoplace* (Figure 1-5) avec la volonté de permettre une interaction plus immédiate tout en évitant d'appareiller l'utilisateur. L'aspect visuel choisi s'affiche délibérément en rupture avec les interfaces *WIMP* de l'époque (Krueger et al. 1985). La nouveauté de l'approche réside dans le fait d'utiliser sa propre silhouette pour agir et interagir avec des objets et créatures virtuelles grâce à un dispositif d'extraction vidéo en temps-réel. Bien que l'installation conduise à une interaction en temps-réel plus transparente

en utilisant le corps, elle ne permet pas une interaction avancée comme la saisie de données ou la modélisation d'objet 3D. Dans cette même optique de s'affranchir de la commande ou du *GUI*, (Bolt 1980) propose un prototype qui permet la manipulation directe associée à des commandes vocales simples. La démonstration *Put that there* permet une interaction à distance simplement en pointant un écran du doigt et en dictant des commandes permettant cette fois quelques créations d'objets et manipulations comme la sélection et le déplacement, sans bouger de son fauteuil. Bien que l'on puisse noter une émergence du geste du bras, les interactions restent en « vase clos », car prédéfinies et contraintes (voir Figure 1-1). On note cependant une simplification de l'interface, privilégiant la manipulation directe tout en évitant en partie les métaphores.

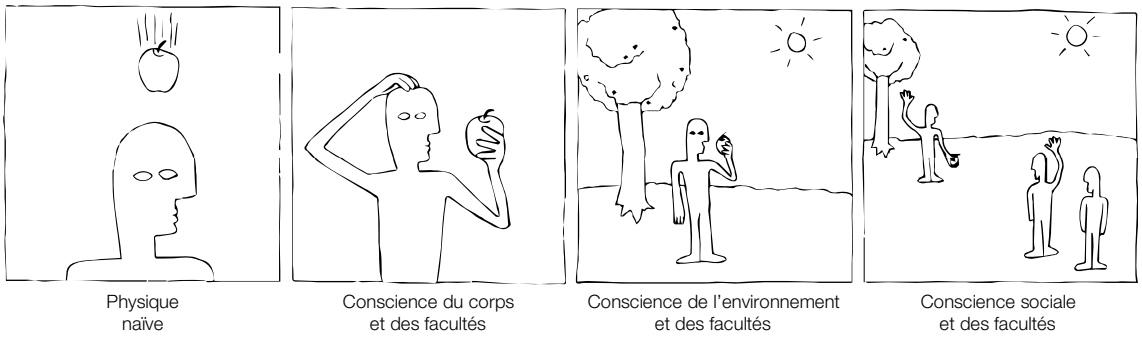


Figure 1-2 : Les quatre thèmes de l'interaction basée monde réel selon (Jacob et al. 2008)

Depuis la fin des années 1990, entrainée par les avancées technologiques et une meilleure compréhension de la psychologie humaine, l'interaction connait une seconde évolution (Jacob et al. 2008). En plus de garder la manipulation directe, elle s'ouvre et se contextualise (Figure 1-1). C'est l'émergence d'interactions dites *post-WIMP* que les auteurs proposent de nommer interactions basées sur le monde réel (*reality-based interaction*). Ils y distinguent quatre thèmes principaux : la physique naïve consistant, par exemple, à ajouter des effets d'inertie dans le défilement de contenu, les contextualisations corporelles, celles de l'environnement et enfin l'aspect social (Figure 1-2). À la fin des années 1990, van Dam évoque des interfaces *post-WIMP* ultimes, impliquant plusieurs sens, et une communication basée sur un langage naturel et multi-utilisateur (van Dam 1997). Il cite deux dispositifs de ce type : la reconnaissance de l'écriture manuscrite sur *PDA*, et l'utilisation de techniques et périphériques non *WIMP* dans les jeux vidéo, par exemple les volants des jeux d'arcade.

Cette seconde évolution (Figure 1-1) se rapproche de plus en plus de ce qui caractérise intrinsèquement la réalité virtuelle : interagir dans l'environnement grâce à ces propriétés d'abstraction basées sur le monde réel. Les deux domaines s'enrichissent mutuellement et la convergence semble s'orienter vers l'activité de l'utilisateur. Les deux critères principaux demeurent la transparence de l'interface et l'accessibilité au niveau cognitif sous la notion d'interface intuitive, rappelant les préconisations de (Fuchs, Burkhardt, et al. 2006) citées précédemment. Mais il reste une différence essentielle dans leurs finalités, l'IHM étant orientée plutôt vers la communication et la réalité virtuelle vers l'activité.

1.1.2. L'usage récent du geste comme modalité d'interaction

Diminution des verrous technologiques

En parallèle des avancées algorithmiques, la dernière décennie aura aussi permis de s'affranchir de verrous technologiques permettant la création de systèmes immersifs et interactifs plus polyvalents (Norman 2010). De plus, leurs coûts se réduisent, facilitant leur propagation. L'interaction gestuelle, canal de communication spontanée, se démocratise et de plus en plus de systèmes intègrent cette modalité. Celle-ci se révèle plus commode, car pratique et facile d'accès. Elle a même l'avantage d'être plus hygiénique pour les dispositifs grand public et les hôpitaux (Schultz et al. 2003). On la retrouve dans les kiosques ou bornes interactives de musées, et dans la navigation des menus des consoles de dernière génération. Elle est implémentée plus récemment dans certaines télévisions en remplacement de la télécommande. Ce concept est d'ailleurs sujet de recherches depuis de nombreuses années (Freeman & Weissman 1995). Le geste, et par extension les mouvements du corps, implique davantage le joueur dans le cas du jeu vidéo. Par conséquent, la plupart des consoles de jeux intègrent la possibilité de jouer en utilisant plutôt des mouvements du corps, en complément ou en remplacement des manettes et joysticks. Les principaux constructeurs proposent cette modalité : citons la *Wiimote* de *Nintendo*, la caméra *Eyetoy* et la manette *PlayStation Move* de *Sony* ainsi que la caméra *Kinect* pour la console *Xbox360* de *Microsoft*. On note également que l'utilisateur est de moins en moins instrumenté, rendant l'utilisation plus immédiate et contribuant à la transparence de l'interfaçage.

Des modalités d'interaction uniformisées

Autre conséquence de leur propagation, voire de leur mondialisation, les modalités d'interaction gestuelle permettant l'usage le plus intuitif tendent paradoxalement à s'uniformiser, en héritant parfois du paradigme *WIMP*. Les derniers jeux vidéo sont un exemple de modalités gestuelles standardisées, leur marché étant mondial. De nombreux de jeux musicaux ou de rythme, de danse ou de chorégraphie sont apparus utilisant différents dispositifs afin de capter la gestuelle de l'utilisateur ((Charbonneau et al. 2009) en détaillent certains). Ces jeux ont le plus souvent un but commun : faire le bon geste au bon moment ou selon différents critères rythmiques ou d'amplitude pour marquer des points. Les possibilités d'interaction sont prédéfinies et bornées par des contraintes visuelles ou spatiales, et la tolérance du système est ajustée selon l'expertise désirée du joueur. En ce qui concerne les interfaces tactiles, l'interaction se fait de prime abord par manipulation directe, c'est-à-dire un schème de sélection et de pointage. Il est ainsi possible de toucher directement le contenu, de saisir du texte grâce à un clavier virtuel, de tapoter pour cliquer, sélectionner, déplacer, et même de faire défiler l'information par effets d'inertie (voir physique naïve (Jacob et al. 2008)). Cependant, un usage plus élaboré hérite du *GUI* et de l'emploi de métaphores, par exemple : zoomer à deux doigts symbolise un étirement de l'information, rester appuyer permet de punaiser une information sur une carte routière, etc. En conséquence, par familiarisation et phénomènes d'apprentissage, les usagers les réutilisent sur des dispositifs similaires.

1.1.3. Besoins d'adaptation aux usages et aux usagers

Adaptation à l'usage et émergence du « fait-maison »

La diversification des usages et la démocratisation de dispositifs d'interaction plus polyvalents et permettent une réduction des coûts. On note l'émergence de détournements de dispositifs issus des jeux vidéo en dispositifs de simulation et de réalité virtuelle personnalisés et plus abordables. Dès 1991, Pausch envisage déjà de réduire les coûts en détournant un gant de données issu du jeu vidéo (*Nintendo PowerGlove*) qu'il associe à un traqueur magnétique professionnel (Pausch 1991). Le gant ne se révèle pas véritablement bas coût (5000 dollars), mais demeure nettement moins cher que certains systèmes commerciaux, avec même une qualité supérieure. Dans cette idée, Lee modifie en 2008 la *Wiimote* pour en faire un dispositif de suivi du point de vue (Lee 2008). À noter qu'il existe de nombreuses solutions *open source* qui participent à cette démocratisation par l'échange et la collaboration. Citons l'existence du protocole standardisé de communication réseau *VRPN* (Russel et al. 2001), ou des plateformes de conception 3D comme *VRJuggler* ("VR Juggler" 2007) ou *Blender* (Blender Foundation 1995). Il existe des forums ("Meant to Be Seen" 2009, "VRoot.org" 2006), des associations d'amateurs ("Project Holodeck" 2012, "The VR Geeks Association" 2010). De nombreuses sociétés et startups apparaissent sous l'impulsion des financements participatifs (*crowdfunding*). En créant du contenu, détournant, modifiant et combinant leurs efforts, les passionnés créent à leur tour leurs propres expériences de réalité virtuelle (Kuntz et al. 2012).

Figure 1-3 : Deux utilisateurs en immersion et interaction avec des dispositifs à bas coûts : à gauche casque Vuzix VR920, Gametrak et Wiimote / à droite casque Sony HMD, SpacePoint Fusion et Razer Hydra. Photos de l'association VRGeeks extraites de (Kuntz et al. 2012)

Adaptation aux usagers

La technologie a acquis une certaine maturité étant plus robuste, plus précise et plus répandue, et de moins en moins intrusive. Mais nous observons que la démocratisation tend à standardiser les interfaces. Ces dernières offrent une accessibilité et une intuitivité apparentes, qui se font souvent au prix de contraintes parfois fortes et d'une certaine rigidité. Cela tend à exclure certains usagers qui ne peuvent pas ou ne souhaitent pas utiliser ces modalités. L'implémentation d'interfaces dites standards est inappropriée dans certains cas

d'usage. En effet, elles ne sont pas adaptées aux utilisateurs aux possibilités motrices limitées, ou souhaitant en faire un usage avancé avec leur propre vocabulaire gestuel. Il nous semble important de considérer ces usagers aux besoins plus spécifiques en prenant en compte leurs intentions, leurs possibilités et leurs habiletés, qu'elles soient sensorimotrices ou cognitives, afin d'effectuer une série d'activités en environnement virtuel. L'objet de cette thèse est d'explorer les moyens d'interaction vis-à-vis des différentes habiletés et dextérités des utilisateurs. Nous ciblons deux cas d'usage qui nécessitent une interaction gestuelle adaptée en environnement virtuel : la réhabilitation cognitive de patients, et les performances artistiques en environnement virtuel. Dans les sections suivantes, nous ferons un état de l'art des applications existantes et nous ciblerons les besoins actuels.

1.2. CONTEXTE DE L'INTERACTION GESTUELLE ADAPTÉE EN RÉALITÉ VIRTUELLE

1.2.1. Contexte local

Le centre de robotique de Mines ParisTech poursuit des recherches sur la modélisation et l'interaction gestuelles à travers différents projets industriels (Chaire de recherche *Robotique et Réalité Virtuelle*, *PSA Peugeot-Citroën*), de préservation du patrimoine immatériel comme la retransmission de savoirs artisanaux et de performances artistiques (projet FP7 *i-Treasures*), mais aussi de santé (projet ANR *AGATHE*). Ce dernier projet a permis le développement d'un outil générant des applications thérapeutiques de réalité virtuelle pour des patients ayant perdu leur autonomie (Klinger et al. 2013). Suite à l'altération de fonctions cognitives due à des accidents tels qu'un traumatisme crânien ou un accident vasculaire cérébral, certains patients ne savent plus s'organiser ni dans le temps ni dans l'espace. Cet outil permet la création d'espaces d'interaction qui simulent des situations de la vie quotidienne pour dans un premier temps évaluer l'état du patient et ensuite intervenir en rééducation. L'idée prospective de cette thèse est d'explorer des modalités d'interaction transparentes pour l'utilisateur, indépendamment de sa condition motrice avec éventuellement des perspectives de déploiement à domicile.

Figure 1-4 : Un patient accompagné d'une thérapeute s'entraine avec une application de réalité virtuelle (Klinger et al. 2013)

1.2.2. Réalité virtuelle et rééducation fonctionnelle

La rééducation fonctionnelle concerne aussi bien la rééducation motrice, ou la physiothérapie que la réhabilitation cognitive afin d'aider le patient à retrouver une certaine autonomie. Depuis la fin des années 1990, plusieurs études relatent les avantages de la réalité virtuelle pour ce type de réhabilitation (Voir par exemple (Cicerone et al. 2000; Klinger 2006)). Concernant les avantages, (Rizzo et al. 1997) montrent qu'il est possible de concevoir des applications pour diagnostiquer, évaluer les besoins et assister les patients dans les Activités de la Vie Quotidienne (AVQ). Les environnements virtuels permettent d'aider à focaliser

l'attention sur l'activité, et de placer le patient dans un environnement contrôlé avec plus de sûreté et de sécurité. Ils notent aussi la possibilité d'enregistrer les expériences et de se diversifier en implémentant par exemple des activités plus ludiques donnant plus de motivation. Aussi, les applications virtuelles ont l'avantage de permettre de répéter les exercices plus facilement et indéfiniment (Gourlay et al. 2000).

Perspectives du déploiement à domicile

Dans ces perspectives d'accessibilité et de facilitation du déploiement de ces systèmes à domicile pour une rééducation, (Gourlay et al. 2000) montrent que cela réduirait le temps et les coûts de déplacements pour le patient et ses accompagnateurs. Le patient restant chez lui, cela évite la désorientation due à un milieu non familier. En parallèle, plusieurs études proposent d'utiliser et de modifier des jeux vidéo à des fins de réhabilitation. Par exemple, les travaux de (Herbelin et al. 2011) offrent de très bonnes perspectives dans l'utilisation de la caméra *Eyetoy* de *Sony* pour la réhabilitation suite à un AVC, ou la console *Nintendo Wii* pour les exercices physiques et d'entrainement cérébral des personnes âgées à domicile. Cependant, ces jeux détournés requièrent souvent une certaine habileté motrice, comme pouvoir saisir une manette ou lever les bras devant une caméra. Il est courant que des personnes atteintes de troubles cognitifs aient également des troubles de la motricité et soient dans l'incapacité d'interagir facilement. Comme l'expliquent (Keates et al. 1998), cela se révèle d'autant moins évident dans un environnement virtuel, notamment par la charge cognitive qu'impliquent l'immersion et l'interaction en environnement virtuel, incluant l'étape de familiarisation avec l'interface.

→ L'enjeu de nos travaux de recherche est de mettre en place une interface qui assiste le patient dans la réhabilitation d'activités de la vie de tous les jours, et qui soit adaptée au handicap de la personne. Il semble important de mettre en place une interaction qui prenne clairement en compte ses capacités sensorimotrices. Ici, la transparence de l'interfaçage se révèle cruciale pour permettre une activité cognitive optimale. Plus concrètement, pour la thérapie cognitive, nous envisageons de proposer des solutions pour l'aide à la réalisation des séquences d'AVQ où le patient pourra utiliser une gestuelle qui lui est propre.

1.2.3. La réalité virtuelle support des arts scéniques numériques

La réalité virtuelle peut être définie comme un médium d'expression artistique (Slater et al. 2000). La technologie est devenue un support au processus de création de performances artistiques et peut aussi servir de modèle à l'écriture (Bardiot 2013). Nous situons notre apport dans le processus de création des arts scéniques numériques impliquant le corps, aussi appelés *digital performances.* (Dixon et al. 2007) les définissent comme les productions des arts de la scène, où les technologies informatiques jouent un rôle clé que ce soit dans le contenu, la technique, l'esthétique ou le résultat final. Nous choisissons de nous situer côté performeur[1], plutôt que du côté spectateur-acteur, bien qu'il se révèle de plus en plus possible d'impliquer ce dernier dans ces performances immersives et interactives.

[1] « Performeur » est pris dans le sens de celui qui exécute la pièce comme un acteur, un mime, un danseur, etc.

Typologie des espaces scéniques virtuels

(Bardiot 2013) propose de regrouper les espaces scéniques virtuels sous quatre catégories de dispositifs scénographiques :

- L'image-scène qui évoque le théâtre traditionnel, mais qui reste souvent réservée à un seul spectateur amené à agir dans/sur la représentation. Elle se présente principalement sous la forme d'installations interactives (historiquement du CD-Rom aux dispositifs immersifs) ;
- La téléscène qui correspond à des scènes à distance ou en réseau ;
- La scène-image envahie d'écrans, où les images deviennent des espaces malléables, ce qui rejoint la tradition théâtrale de scène-tableau faisant de la scène une image ;
- La scène augmentée avec adjonction ou superposition d'un espace virtuel à un espace réel à la manière de la réalité augmentée.

Bien entendu, les frontières sont minces et il y existe des possibilités de combinaison. Par conséquent, selon les dispositifs d'immersion visuelle et d'intégration d'objets réels, on peut parler de réalité augmentée, mixte ou virtuelle, ce qui renvoie au continuum réalité-virtualité de (Milgram et al. 1994).

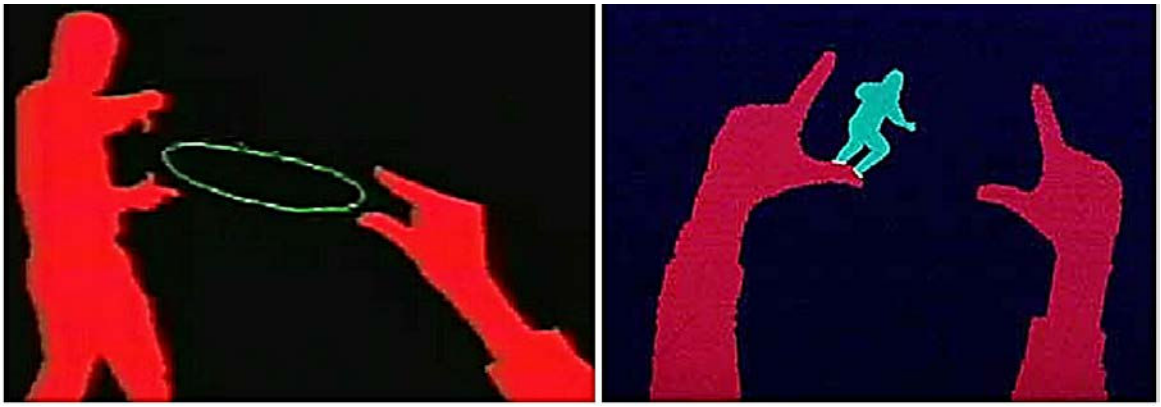


Figure 1-5 : Captures d'écran de l'installation interactive *Videoplace* (Krueger et al. 1985)

Les rapports au corps du performeur et au jeu scénique virtuel

Parmi toutes ces possibilités scénographiques, recentrons-nous sur deux aspects principaux décrits par (Bardiot 2013) : le rapport au corps et le rapport au jeu scénique. Premièrement, dans le rapport au corps, l'espace scénique se définit à partir du corps du ou des performeurs. Le costume peut être l'interface elle-même et les capteurs embarqués ou à poste fixe comme les capteurs par vision. Différentes zones d'interaction sont définies, et ce sont elles qui organisent et créent l'espace et le jeu. L'espace scénique se définit à partir du corps lui-même. On peut citer l'installation artistique *Videoplace* décrite précédemment, la silhouette du spectateur-acteur permet l'interaction dans un espace en deux dimensions. La performance *Messa Di Voce* (Levin et al. 2003) reprend cette idée et exploite des interactions du même type en y ajoutant l'ombre et la voix comme générateurs de formes. Il existe aussi des plateformes pour VJ[2] qui permettent à l'utilisateur de définir des commandes gestuelles.

[2] Le terme VJ (VideoJockey) est l'équivalent de DJ (DiskJockey), où le media mixé n'est pas un son, mais une vidéo ou un effet visuel temps-réel.

Citons par exemple la récente *Kinectar Performance Platform* (Vik 2011) qui permet de définir soi-même des instruments virtuels personnalisés (par sa fonctionnalité *Instrument builder*).

Deuxièmement, dans le rapport au jeu scénique, les dispositifs d'interaction sont disposés sur scène et ce sont les déplacements du performeur qui créent le jeu. L'une des premières performances jouant sur cette approche est *Variations V* (Cage et al. 1965) où les mouvements des danseurs interviennent directement sur la composition musicale. Techniquement, au bord de la scène, des capteurs de proximité, comme des cellules photoélectriques, déclenchent ou interrompent les sons et des antennes capacitives (thérémines) s'activent par proximité au passage des danseurs. Petite parenthèse historique, la thérémine inventée par Termen en 1919 est considérée comme le premier instrument pouvant être joué sans être touché, le son est modulé en temps-réel par la gestuelle. On note ici une vraie transparence de l'interface, les danseurs n'étant pas instrumentés. Le geste d'interaction est défini dans le rapport du corps à l'objet ou à l'espace. Dans le domaine de la musique contemporaine, la pièce de percussions virtuelles *Voir-Toucher* (Pagliei 2013) associe le son et le geste au moyen d'une lutherie composée d'un alphabet de gestes. La performance utilise des « méta-instruments » associant des gestes libres, et des contacts avec des pièces de bois. Ici, les notions de souplesse émergent dans l'outil devant satisfaire les possibilités créatives et de composition.

Selon (Bardiot 2013), il existe d'autres performances intégrant davantage la gestuelle et hybridant ces espaces corps et scène. L'auteur décrit le spectacle augmenté *Nou I_D* (Sanchez et al. 2006) comme un dialogue entre interprètes et environnement scénique, avec une écriture par conséquent non linéaire. Pour ce faire, les auteurs croisent la reconnaissance du geste en temps-réel avec des systèmes de captation sans fil et la vision par ordinateur. L'hybridation et le dialogue apportent ici des réflexions sur l'intelligence ambiante ou le *living art*. Dans cette même idée de dialogue, mais cette fois avec un avatar, citons l'installation *La Funambule* (Bret et al. 2002). Outre l'implémentation d'interactions littérales comme le mouvement du balancier de l'usager qui produit une rotation de la funambule sur elle-même, l'entité est dotée d'un comportement autonome. La funambule peut produire des réponses non écrites au préalable, mais émergeant de la faculté du système à s'adapter aux gestes de la personne qui interagit. Cela fait référence au domaine plus large de l'autonomie.

→ On constate que les dans les arts scéniques numériques, la gestuelle est intrinsèque à l'expressivité. Ici, c'est la diversité des mouvements et leur interaction avec l'environnement virtuel qui importent car les mouvements participent au processus de création. Un des besoins fondamentaux dans une activité de création est la liberté d'expression, la technologie venant en support aux intentions du performeur. L'acteur, le mime ou le danseur, peut associer un geste propre à une activité temps-réel sur scène comme un effet graphique, une animation d'avatar, etc. Les notions émergentes d'autonomie montrent que l'interaction doit s'adapter, dialoguer, voire s'autoadapter.

1.3. POSITIONNEMENT DE LA THÈSE

1.3.1. Objectifs centrés sur l'adaptation

Nous avons vu que la démocratisation de l'interface gestuelle appuyée par le besoin inhérent d'accessibilité et d'immédiateté entraine simplification et uniformisation. Des populations aux besoins sensorimoteurs et cognitifs spécifiques, les utilisateurs novices ou les experts désirant un usage avancé risquent d'être mis à l'écart. Au regard des besoins d'adaptation, ces travaux de recherche ont pour objectif de proposer et d'évaluer de nouveaux moyens d'interaction gestuels intuitifs et transparents. Les domaines d'application visés sont la rééducation fonctionnelle, la thérapie à domicile, ainsi que les arts numériques scéniques nécessitant la gestuelle en temps-réel. Ces cas d'usage ne permettent pas l'utilisation d'interfaces standardisées qui font appel le plus souvent à des gestes prédéfinis et imposés. Nous nous tournons donc vers une approche centrée utilisateur considérant ses propres gestes pour interagir selon ses aptitudes sensori-motrices et cognitives. Dans un souci d'accessibilité matérielle de l'interface, nous intégrons dans la conception les dispositifs à bas coûts, peu intrusifs et de manipulation facile. Nous avons pour ambition par cette thèse de donner un certain nombre de réponses et de recommandations pour la mise en place d'une interaction centrée utilisateur à travers deux cas d'usage :

- → Pour la rééducation fonctionnelle, nos recherches se focalisent sur la possibilité laissée à l'utilisateur d'agir avec ses propres gestes, afin d'accomplir une activité en environnement virtuel. Nous nous basons sur une séquence d'actions issues de situations réelles écologiques effectuées par les patients en rééducation.
- → Dans la pratique artistique, l'idée est de permettre aux artistes une interaction par le geste sur la base d'une expérience sensorimotrice, en leur proposant de définir leurs propres mouvements et chorégraphies pour l'interaction. Concrètement, nous proposons que les mouvements du corps puissent être couplés à des entités virtuelles ou effets visuels permettant un contrôle ou un dialogue avec ces éléments.

1.3.2. Problématique de recherche

Notre problématique concerne l'interaction gestuelle par mouvements non imposés. Elle pose un certain nombre de questions. Nous situons ces travaux dans l'adaptation de l'interfaçage et de l'interaction aux gestes de l'utilisateur. La question de fond est donc :

- → **Comment permettre une activité en environnement virtuel à partir d'une gestuelle non imposée ?**

Il en découle trois questions centrées sur l'homme et une autre technologique auxquelles nous répondrons dans ce mémoire (Figure 1-6) :

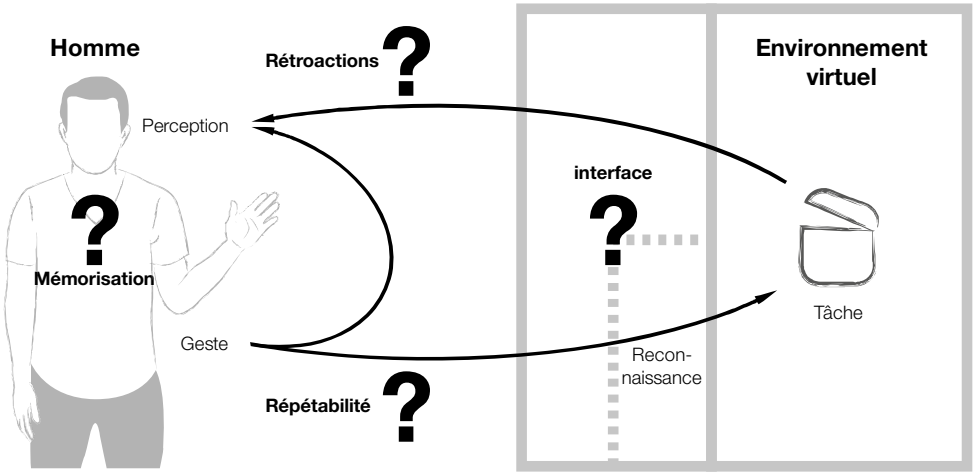


Figure 1-6 : Positionnement des problématiques soulevées par une approche où l'Utilisateur Définit l'Interaction gestuelle

Répétabilité

Nous faisons l'hypothèse d'une approche « sur mesure » clairement adaptée à l'utilisateur, en considérant que ce dernier puisse proposer et réutiliser ses gestes. Or, nous savons qu'il n'est pas possible de répéter à l'identique un geste et qu'il varie par conséquent dans sa répétition. L'interaction gestuelle doit se révéler exploitable malgré cette variabilité. Pour étudier la répétabilité, nous répondrons à trois sous questions :

→ Que signifie refaire un même geste ?
→ L'utilisateur doit-il apprendre à refaire le même geste ?

Nous savons que répéter un geste peut entrainer une certaine fatigue et à la longue une plus grande variabilité dans sa répétition.

→ Quelle est l'influence de la fatigue sur la répétabilité ?

Mémorisation

Outre la répétabilité du geste, la mémorisation d'un ensemble de gestes au regard de différentes tâches est à étudier :

→ Comment favoriser la mémorisation d'un vocabulaire de gestes définis par l'utilisateur ?

Rétroaction

L'interaction gestuelle se fait en environnement virtuel avec un retour visuel qui donne une rétroaction à l'utilisateur. Ce retour visuel peut s'afficher à différentes étapes du geste.

→ Quand et comment exploiter le retour visuel afin d'aider l'utilisateur dans son interaction ?
→ Ce retour visuel influence-t-il sa compréhension de l'avancement dans l'activité ?

Considérations technologiques

Enfin, nous souhaitons exploiter du matériel peu onéreux pour rendre l'interaction gestuelle accessible au grand public. Mais les dispositifs à bas coût peuvent avoir des problèmes inhérents de performance ou de précision.

→ Comment pallier les problèmes des contraintes techniques avec des systèmes peu performants ?

Ces travaux prospectifs ont pour objectif de donner des réponses sur les possibilités d'interaction par des gestes quelconques, à travers des évaluations sur l'usage de l'interaction gestuelle au sens large. Ces évaluations se font par capture et suivi du geste, associés à une rétroaction visuelle. Dans cette étude, nous exploiterons des algorithmes d'apprentissage et de reconnaissance du geste déjà existants.

Nous avons l'ambition de proposer une solution générale à l'interaction gestuelle à partir de mouvements non imposés. Nous menons des évaluations dans un cas général. Nous avons testé, sans réaliser une validation approfondie, **notre interaction basée sur des gestes définis par l'utilisateur** dans les deux domaines applicatifs envisagés : la rééducation fonctionnelle et la pratique artistique.

1.3.3. Organisation du manuscrit

À l'issue de cette introduction présentant le contexte actuel et les cas d'application ciblés, nous répondons aux questions de recherche soulevées par l'état de l'art dans le chapitre 2. Une première partie traite des possibilités d'adaptation de l'interaction à l'usager et une deuxième partie décrit l'usage et la caractérisation du geste pour l'interaction. Les deux contributions de cette thèse sont présentées au sein des chapitres 3 à 5. La première contribution conceptuelle du chapitre 3 décrit le modèle UDIg où l'Utilisateur Définit l'Interaction gestuelle. De ce modèle, nous formulons des hypothèses sur la remémoration des gestes, leur répétabilité et le choix des rétroactions à l'interaction. Les contributions expérimentales décrites dans les chapitres 4 et 5 permettent de valider ces hypothèses. Le chapitre 6 qui conclut ce travail de recherche propose une discussion et une synthèse des résultats obtenus. Des perspectives issues de tests réalisés avec des patients et le grand public ont permis d'établir des préconisations d'usage spécifiques.

CHAPITRE 2
ÉTAT DE L'ART SUR L'INTERACTION GESTUELLE ADAPTÉE EN ENVIRONNEMENT VIRTUEL

Dans ce chapitre, nous proposons un état de l'art en deux volets sur l'interaction gestuelle adaptée. Avant d'adapter un système de réalité virtuelle à l'homme, nous souhaitons dans un premier temps situer les facultés de l'homme en vue de son adaptation au système. Au regard de notre problématique, plusieurs théories inspirent l'élaboration d'une interaction adaptée avec un système de réalité virtuelle. Nous nous centrons sur l'interaction ; il existe plusieurs possibilités d'implémentation d'une interaction adaptée aux besoins de l'utilisateur. Nous décrivons les principales possibilités et définirons nos orientations. La question de leur mise en place au sein d'un environnement virtuel se pose. Nous explorerons les possibilités de conception d'un environnement virtuel au regard du caractère adapté de nos cas d'application.

Le second volet concerne l'interaction gestuelle. En premier lieu nous explorerons le fonctionnement du geste afin d'en isoler ses caractéristiques. Pour une adaptation à l'interaction, l'utilisateur proposera des gestes, mais cela implique un apprentissage. Nous décrirons des possibilités d'apprentissage par le système au regard d'une interaction gestuelle adaptée.

2.1. UNE INTERACTION ADAPTÉE EN ENVIRONNEMENT VIRTUEL

2.1.1. L'adaptation de l'interaction pour l'expérience utilisateur

Nous avons au cœur de notre recherche l'objectif d'une interaction gestuelle adaptée en environnement virtuel. Pour cela nous abordons la technologie informatique comme un outil au service de l'activité. L'adaptation de cet outil requiert de considérer l'activité humaine, le fonctionnement cognitif de l'utilisateur, ses objectifs, ses besoins et ses exigences (Bobillier-Chaumon et al. 2005). L'adaptation se révèle donc être un processus complexe. Nous proposons de caractériser et de décrire quelques aspects de l'homme en rapport à son environnement, à ses connaissances et son aptitude à l'adaptation.

2.1.1.a *Considérations sur l'adaptation*

Dans la mise en place d'une technologie adaptée, toute la difficulté réside dans la considération de l'homme « dans son épaisseur et sa complexité » (Bobillier-Chaumon et al. 2005). Dès le milieu des années 1980, de nombreuses théories se distinguent d'approches trop « technocentrées ». Ces théories redonnent une dimension à l'activité humaine. (Dreyfus 1984) exprime notamment que « le caractère intelligent du comportement humain est lié non pas au logico-symbolique, au calcul, mais à son caractère situé et incarné ». De nombreux travaux en psychologie du développement, en sociologie, en anthropologie et en sciences cognitives envisagent de dépasser les modèles fondés sur le traitement de l'information. La cognition ne se réduit donc pas à un traitement d'informations ou à un enregistrement de données issues d'un monde prédéfini : il s'agit alors de conceptualiser l'homme comme acteur situé, incarné et relationnel.

Ceci implique à l'homme de se référer à sa culture, à son histoire et à la complexité des situations sociales. Pour ce faire, l'approche écologique de (Gibson 1966, 1979) propose notamment de considérer l'homme et sa perception en rapport à l'environnement. Les actions de l'homme ne seraient pas guidées par ses processus cognitifs internes, mais par les contraintes écologiques, c'est-à-dire les structures du monde externe. L'environnement fournit les comportements possibles à travers les affordances. Celles-ci font référence aux attributs des objets qui permettent à l'utilisateur de savoir comment les utiliser. Gibson propose l'idée de perception active, les informations sont perçues grâce aux mouvements permettant de détecter les invariants. Toujours en psychologie cognitive, (Neisser 1976) montre le caractère indissociable de la perception-action, selon lui l'exploration du monde visuel serait dirigée par des schèmes anticipatoires définis comme des plans pour l'action perceptive. Dans les neurosciences, (Berthoz 1997b) considère également la perception comme une activité interprétative dirigée vers l'action.

(Burkhardt et al. 2004) notent que Gibson fonde essentiellement sa théorie sur la perception sans y inclure les processus cognitifs. Afin de mieux considérer la cognition, (Varela et al. 1991) mobilisent la phénoménologie et les aspects neurologiques et biologiques. Ils proposent le paradigme de l'énaction conférant à la cognition un caractère incarné. Ce paradigme permet de considérer la cognition par le couplage structurel entre l'organisme et son environnement. Ceux-ci se révèlent indissociables et se déterminent l'un vis-à-vis de l'autre : l'organisme donne forme à son environnement en même temps qu'il est façonné par lui. Concernant le rapport à l'outil technologique, (Stewart et al. 2004) suggèrent que « tout

artefact, du silex à la voiture, ou à l'ordinateur sont des interfaces énactives qui couplent structurellement l'être humain et le monde dans lequel il évolue ». Les auteurs développent la thèse que l'interface a un rôle non négligeable, car elle peut impacter l'énaction, tout comme « les skis modifient la perception de la montagne » pour reprendre l'analogie de (Thouvenin 2009). Dans le cas d'une activité avec l'informatique, (Bobillier-Chaumon et al. 2005) soulignent « qu'appréhender l'adaptation des technologies, c'est aborder aussi le rapport de l'individu aux dispositifs qui d'une certaine manière le déterminent, mais qu'il structure en retour également ».

Une interface — même adaptée — impacte ainsi son utilisateur. Cela provient naturellement des facultés d'adaptation que l'homme possède. (Berthoz 1997a) souligne que l'adaptation est l'une des propriétés les plus importantes du système nerveux. Il parle de plasticité neuronale, ou plasticité cérébrale. L'auteur montre par exemple que l'adaptation est mise en œuvre suite à des dysfonctionnements des sens, ou des lésions ou des conflits sensorimoteurs. Elle fait intervenir des processus qui modifient des mécanismes locaux par guidage global de l'organisme.

Notre travail de recherche s'attache justement à solliciter au minimum ces facultés d'adaptation de l'utilisateur. En revanche, l'expérience associée aux connaissances a priori de l'utilisateur entrent en jeu lors d'une interaction avec un système, plus spécifiquement avec un environnement virtuel dans notre étude. Certaines connaissances pouvant être sollicitées, nous proposons d'explorer les facultés d'acquisition et de réutilisation de l'expérience.

2.1.1.b *Acquisition et réutilisation des connaissances*

(Hurtienne et al. 2008) proposent de situer les différents types de connaissances que l'homme possède et est en mesure d'acquérir tout au long de sa vie. Les auteurs proposent de les positionner dans un continuum (Figure 2-1) entre proportion d'individus concernés en abscisse, et niveau d'encodage et de réutilisation des connaissances en ordonnée. Cette classification demeure discutable, notamment vis-à-vis de l'ancrage écologique discuté précédemment. Toutefois, cela permet d'avoir une première base afin de situer l'adaptation cognitive.

- Le premier niveau concerne **les connaissances innées** de l'utilisateur comme les réflexes et/ou les comportements instinctifs. Il est considéré par certains experts comme le seul réellement intuitif, car le plus universel. Les auteurs précisent que ce niveau est encore sujet à discussion.
- Le deuxième niveau — appelé niveau sensorimoteur — intègre les **connaissances générales acquises** durant l'enfance et utilisées constamment dans les interactions avec le monde. Il comprend les concepts d'affordances, de psychologie de la forme ou gestaltisme et de schème mental. Nous proposons de détailler ces concepts :
    - Comme décrit précédemment, les affordances mises en évidence par (Gibson 1979) font référence aux possibilités d'action induites par l'environnement ou l'objet, qui peuvent être perçues ou non par l'utilisateur. (Rizzolatti et al. 2011) estiment que l'acte moteur approprié serait choisi au regard d'une zone précise de l'objet présentant des affordances.
    - La théorie des gestalts (Koffka 1935) met l'accent sur les ressemblances structurales entre des ensembles ou des structures globales comme des formes, et non comme une simple addition ou juxtaposition d'éléments. Un ensemble d'étoiles dans une certaine

configuration est interprété par exemple comme la *Grande Ourse*. À noter que cela se rapproche des observations de (Rizzolatti et al. 2011) citées ci-dessus.

  - Enfin, selon (Piaget 1971), un schème est l'assimilation d'un objet physique, ou d'une idée, perçus par l'individu à un ensemble d'objets. L'assimilation peut se faire aussi vis-à-vis d'une situation dont l'individu est déjà familier. Comme le note Piaget, les schèmes sont reproductibles, ont un but. Ils sont utilisés et assimilés de façon inconsciente. Ils ne requièrent plus d'apprentissage pour des activités basiques. Si cette assimilation échoue, par exemple si l'objet ne peut être associé à un schème psychologique préexistant, alors commence un processus d'accommodation. Celui-ci consiste à modifier un schème existant afin de pouvoir intégrer un nouvel objet ou une nouvelle situation. Cela renvoie à la notion d'adaptation précédemment décrite.

- Le troisième niveau comprend **les connaissances culturelles** issues du milieu de l'individu. Elles diffèrent par conséquent énormément d'un milieu à un autre. Cela rejoint justement les considérations de Norman à propos des interfaces dites naturelles qui ne le sont pas nécessairement d'une culture à une autre (Norman 2010).
- Le niveau le plus spécifique est **l'expertise**. Les auteurs la définissent comme l'acquisition d'une connaissance spécifique dans le cas par exemple d'une profession ou d'un hobby.
- Les auteurs distinguent en plus **la connaissance de l'outil** qui se révèle importante pour l'élaboration d'une interface avec métaphores. L'exemple le plus courant est la métaphore du bureau en informatique qui fait référence aux notions de dossier, fichier, corbeille. Nous verrons dans la partie 2.1.3.b que la métaphore reste une option en environnement virtuel lorsqu'il n'est pas possible d'utiliser un schème pour interagir.

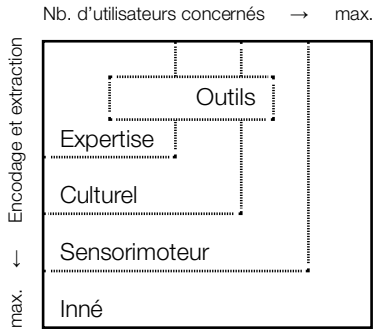


Figure 2-1 : Continuum des sources de connaissances selon (Hurtienne et al. 2009)

En replaçant le geste dans l'apprentissage, on constate qu'il contribue à la mémorisation, aux connaissances. Plusieurs études (Goldin-Meadow 1999; Wagner et al. 2004) montrent que le geste aide à la fois à la remémoration et faciliterait la résolution de problèmes. Les auteurs émettent l'hypothèse d'une réduction de la charge cognitive, car l'apprenant passerait d'une mémoire verbale à une mémoire spatiale. Un très grand nombre de simulateurs d'apprentissage de gestes techniques exploitent directement les gestes issus d'une situation équivalente en réel. Par l'entrainement et la répétition en environnement virtuel, ces gestes sont assimilés et incarnés. (Thouvenin 2009) souligne l'idée que le geste est considéré comme forme de connaissance empirique. L'auteur en distingue trois formes : le geste technique comme copie exacte d'un geste expert, le geste à vocation artistique qui demande un réglage permanent et un ressenti, et enfin le geste sportif, plus axé sur la performance.

→ Nous avons vu que l'homme ancré dans son environnement est capable d'acquérir des connaissances. Étant donné que nous souhaitons minimiser l'adaptation de l'utilisateur, l'interface doit être accessible en lui conférant de la flexibilité et de la tolérance. Le principe est de proposer une interaction qui peut faire appel aux connaissances de l'utilisateur en évitant de trop solliciter un apprentissage supplémentaire ou superflu. Différentes implémentations de systèmes interactifs proposent d'exploiter ces connaissances. Nous proposons d'explorer les grandes approches pour une adaptation de l'interaction.

2.1.2. Principales approches sur l'adaptation de l'interaction

Situer l'adaptation d'un système peut prendre plusieurs formes. (Bobillier-Chaumon et al. 2005) en distinguent trois principales (Figure 2-2) :

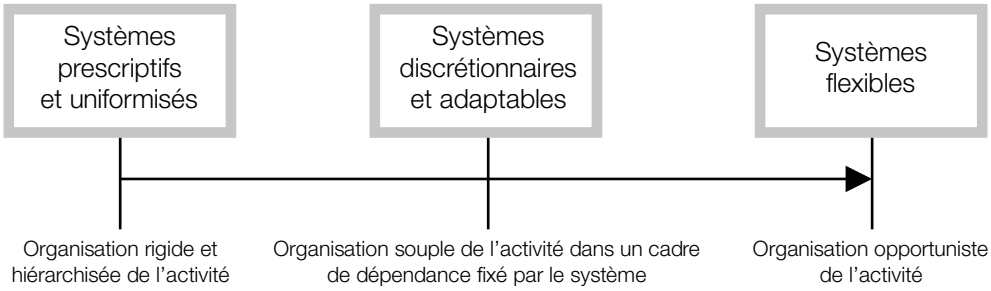


Figure 2-2 : Différents niveaux d'adaptation et de prescription des systèmes et du travail (Bobillier-Chaumon et al. 2005).

- On peut considérer que l'interaction est standardisée et adaptée pour le plus grand nombre, on se trouve dans le cas d'un système uniformisé ;
- Quand la standardisation ne satisfait pas certains utilisateurs ou des cas d'usage, des adaptations de cette interaction sont possibles. Cela présente une option envisageable dans nos cas d'application. Il est possible d'adapter l'interaction, mais le cadre de dépendance reste fixé par le système ;
- Dans le cas d'une adaptation directe, il existe la possibilité d'une autre approche dédiée avec un système flexible.

Nous détaillons ces trois approches dans les parties suivantes.

2.1.2.a *Utilisateur ← Système : l'interaction standardisée*

Comme discuté en introduction, l'approche d'une interaction standardisée cible une utilisation sans prérequis en invoquant des références générales. La standardisation renvoie aux connaissances communes, culturelles innées ou acquises sous la notion d'intuitivité. Plus en détail, (Hurtienne et al. 2008) définissent l'intuitivité comme l'application d'une connaissance inconsciente de l'utilisateur qui permet une interaction optimale. La problématique fréquente d'une conception d'une expérience « utilisateur » reste de choisir la modalité d'interaction moyenne issue d'un consensus, renvoyant aux connaissances générales permettant d'être utilisée par la plupart des utilisateurs.

Une interaction standardisée fait appel au stéréotype qui selon (Rich, 1989) est le modèle d'un groupe d'individus aux « comportements » similaires. Dans ce cas, les utilisateurs sont supposés appartenir à une population homogène. (Bobillier-Chaumon et al. 2005) constatent que les interactions standardisées ont l'avantage de bénéficier d'une grande accessibilité grâce aux références communautaires ou de culture de masse. On observe également un transfert d'apprentissage entre les interfaces. Cela permet une réutilisation entre différentes applications. Cette interopérabilité se retrouve dans l'usage de dispositifs tels que les tablettes, les télévisions interactives, les kiosques d'information, les bornes interactives, les consoles de jeux vidéo, les simulations de réalité virtuelle, etc. Les utilisateurs face à des systèmes électroniques similaires ont tendance à acquérir de l'expérience et à la réutiliser par effet de propagation. Enfin, la simplicité de mise en œuvre d'une interaction standardisée facilite la maintenance, à la sureté et au contrôle du processus.

Les interactions gestuelles implémentées peuvent faire appel à des connaissances générales décrites précédemment (cf. 2.1.1.b) qui peuvent être de l'ordre de l'acquis, comme pointer du doigt, attraper ou saisir. Elles peuvent faire l'objet d'utilisation de métaphores comme balayer d'un geste, écarter les mains pour zoomer, ou emploient des modèles physiques naïfs (Jacob et al. 2008). À première vue, on pourrait penser que ces systèmes permettent une grande liberté d'interaction, permettant à l'utilisateur d'interagir comme bon lui semble. En fait, les mouvements sont clairement induits par le contexte du contenu virtuel, avec rétroactions explicites accompagnées d'aides visuelles, de modes d'emploi ou d'indices comme des pictogrammes, animations, tutoriaux, etc. (LaViola et al. 2011). Le système est borné par un certain nombre de contraintes, le but ici étant d'être accessible par le plus grand nombre, et pour la même raison, mis en place pour un nombre réduit d'activités.

Cette approche d'uniformisation a cependant quelques limites. Norman (Norman 2010) montre que l'accessibilité se révèle au final restreinte : l'interaction prédéfinie est issue de consensus et est destinée à un usage générique. Elle ne prend pas en compte les différences interculturelles et interindividuelles. Au niveau interculturel, prenons l'exemple d'agiter sa main face au dispositif pour signaler que l'on souhaite prendre le contrôle (exemple du manuel (Microsoft 2013)). Ce signe voulant dire « bonjour » dans les cultures occidentales signifie au contraire « non merci » ou « le rejet » dans les cultures indiennes d'Asie ou du Moyen-Orient. Au niveau interindividuel, les limitations sont à prendre en considération comme les spécificités sensorimotrices (l'utilisateur est dans l'impossibilité de lever les bras), ou cognitives (comme savoir lire, la familiarité avec une interface, la compréhension du tutoriel, etc.). (Bobillier-Chaumon et al. 2005) ajoutent qu'il faut prendre en compte les considérations affectives et sociales. Le choix d'une interface standardisée risque de rendre l'interaction non accessible de prime abord, car décontextualisée. Enfin, on observe une certaine rigidité de l'interface où tout est prédéfini, et par conséquent non-adaptable. Dans nos cas d'usage, il semble préférable de nous tourner vers une interface plus adaptée ou paramétrable selon les aptitudes, les besoins ou les habitudes de l'utilisateur.

2.1.2.b *Utilisateur ⇆ Système : l'interaction adaptable ou adaptative*

Afin de pallier les situations que l'interaction standardisée ou uniformisée ne peut satisfaire, il existe des possibilités d'ajustement. Il s'agit alors de modifier ou de paramétrer cette interaction prédéfinie pour qu'elle puisse répondre aux aptitudes sensorimotrices ou cognitives de l'utilisateur. Selon (Rich 1989), l'utilisateur ne correspond plus dans ce cas au stéréotype, mais à un modèle individuel aux besoins spécifiques. (Bobillier-Chaumon et al. 2005) constatent que l'adaptation peut être faite soit du côté utilisateur, soit automatisée par le système. Ils distinguent la personnalisation de la customisation. Dans le cas d'une personnalisation, le système guide le processus d'adaptation. La customisation laisse l'utilisateur choisir entre plusieurs options prédéfinies. À cela s'ajoutent les capacités d'adaptation du système entre adaptabilité et adaptativité. Cela dépend du contrôle laissé à l'utilisateur ou au système, sur quatre différentes étapes :

- L'initiation correspond à la décision de suggérer l'adaptation ;
- La proposition représente les suggestions ou les recommandations ;
- La sélection correspond au processus de choix parmi les propositions de réaction ;
- L'exécution est la mise en œuvre de la réaction choisie.

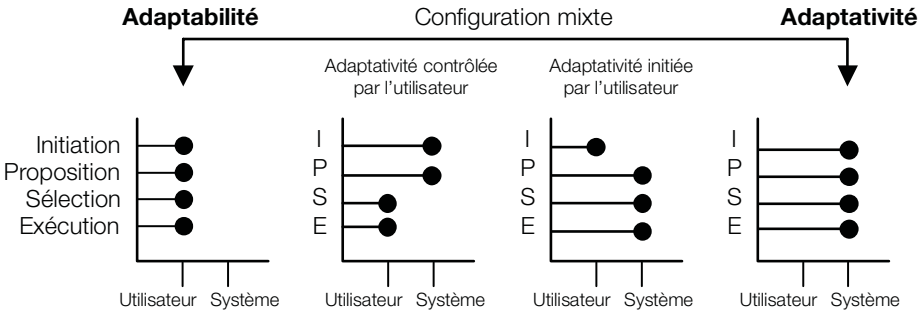


Figure 2-3 : Différents types de processus d'adaptation : de l'adaptabilité à l'adaptativité d'après (Dieterich et al. 1993), extrait de (Bobillier-Chaumon et al. 2005)

Oppermann (Oppermann et al. 1997) propose un continuum (Figure 2-3) entre adaptabilité et adaptativité, qui laisse respectivement le contrôle total à l'utilisateur ou au système. Une interface adaptable ou statique permet ainsi à l'utilisateur de changer lui-même certains paramètres. Un exemple classique d'interaction en 2D adaptable est la possibilité de changer la couleur, la taille et la vitesse d'un pointeur de souris. En réalité virtuelle, (Bowman et al. 2006) suggèrent par exemple de personnaliser l'interface selon les préférences de l'utilisateur comme en laissant la possibilité d'ajout de modification ou de paramétrage des techniques d'interaction appelées « saveurs » (traduit de l'anglais *flavors*). Au niveau ergonomique, cela octroie un confort et une plus grande efficacité dans l'interaction puisque l'adaptation correspond aux besoins de l'utilisateur. Mais là encore, il s'agit de modifier ou de personnaliser une technique définie d'avance. Une autre option donnée par les auteurs est la mise en place d'interactions spécifiques. Celle-ci dépend de la tâche, du domaine, de l'interfaçage, de l'application et de l'utilisateur. Il est possible de considérer son âge, son expertise, ses préférences, etc. Ils estiment cependant que la mise en place de cette interaction spécifique entraine un investissement certain.

Une autre possibilité consiste à laisser le système s'adapter automatiquement à l'utilisateur en présupposant ses besoins. Il s'agit alors de systèmes adaptatifs ou dynamiques. L'approche mixte laisse l'utilisateur choisir l'approche suggérée par le système. Selon (Langley 1999) les interfaces utilisateurs adaptatives sont conçues pour réagir et interagir avec l'utilisateur dans le but de permettre une meilleure expérience utilisateur. Un exemple dorénavant répandu est la conception de sites web adaptatifs (traduction de *responsive design*). Ceci permet d'améliorer l'ergonomie de sites web et l'expérience utilisateur (Gardner 2011). Pour cela, l'affichage du contenu et la navigation s'adaptent automatiquement selon le terminal utilisé (mobile, tablette, grand écran). Cette approche permet par exemple de pallier les contraintes de visualisation ou de restitution des sites web sur mobile, en privilégiant l'allégement du contenu. Certains kits de développements intègrent depuis peu la fonctionnalité de *responsive design*, l'interface s'ajustant en fonction de la proximité de l'utilisateur à l'écran (voir *Distance-Dependent Interactions* et *Adaptive UI* (Microsoft 2013)). Mais, là encore, le système déduit l'adaptation optimale en s'adaptant au dispositif et pas forcément selon les aptitudes de l'utilisateur.

(Bobillier-Chaumon et al. 2005) soulignent que l'interaction ainsi personnalisée et la reconnaissance de l'usager ont pour avantages une plus grande flexibilité et la versatilité grâce à la prise en compte du contexte et de la situation d'usage. Il existe cependant les limitations : un système trop adapté implique un risque d'isolement, une complexité de mise en œuvre et d'évaluation. Ensuite, il peut y avoir des effets négatifs lors d'une mauvaise adaptation. Enfin, penser la conception au regard de processus de réappropriations possibles par l'utilisateur se révèle difficile. La table ci-dessous résume les avantages et limites des deux grandes approches interaction standardisée ou adaptée :

Table 2-1: Intérêts et inconvénients des approches d'uniformisation et d'adaptation, extrait de (Bobillier-Chaumon et al. 2005)

Standardisé	Adapté
✓ Référence communautaire, culture de masse	✓ Ergonomie, confort, efficacité
✓ Transfert d'apprentissage	✓ Personnalisé, reconnaissance de l'usager
✓ Réutilisation et interopérabilité	✓ Flexibilité et souplesse d'usage
✓ Simplicité de mise en œuvre et maintenance	✓ Prise en compte contexte et situation d'usage
✓ Sûreté et fiabilité	
× Différences interindividuelles	× Risque d'isolement
× Décontextualisé	× Complexité de mise en œuvre, évaluation
× Rigidité	× Effet négatif lors d'une mauvaise adaptation

Le choix d'une interaction adaptée plutôt que standardisée reste une question ouverte. En effet, pour une application grand public, avec un nombre réduit de tâches, une approche standardisée semble appropriée. Celle-ci comprend des aides visuelles, une grande tolérance et des tutoriels implicites qui évitent une phase d'apprentissage. Mais l'approche standardisée met de côté les besoins et les aptitudes spécifiques requis par certains usagers. Par ailleurs, pour un usage expert ou des besoins plus spécifiques, une interaction gestuelle prédéfinie incluant une phase d'apprentissage et de familiarisation par l'utilisateur semblerait moins contraignante. De plus, tout comme le continuum d'Oppermann, le cas spécifique où

l'utilisateur définit lui-même le geste pour interagir dans une tâche précise n'est pas considéré.

2.1.2.c *Utilisateur → Système : l'utilisateur définit l'interaction*

Une autre approche possible est de faire apprendre au système informatique la façon dont l'utilisateur souhaite réaliser l'interaction. L'activité est directement prise en compte et ce dernier est force de proposition. Ce cas permet une réelle adaptation aux aptitudes et préférences de l'utilisateur. Le concept général est de présenter différentes tâches d'une activité à effectuer et l'utilisateur doit alors associer une modalité d'interaction à chaque tâche. Les possibilités sont très diverses : la voix, le geste, le suivi du regard, le souffle, ou des informations physiologiques (rythme cardiaque, sudation). L'utilisation d'une interface cerveau-ordinateur (*Brain-Computer Interfac*es ou *BCI*) serait même envisageable comme le montrent des travaux récents de (Lécuyer et al. 2008). Des combinaisons multimodales sont envisageables, comme l'association du geste et de la voix.

Utilisation de cette approche pour l'interaction

Il existe des implémentations de cette approche utilisant un apprentissage par le système que l'on retrouve en informatique et dans les domaines du handicap, de la création artistique ou de la robotique :

(Myers 1998) souligne que Engelbart, dès 1962, suggère de laisser la possibilité à l'utilisateur de définir ses propres macros dans les outils de traitement de texte pour en simplifier l'usage. Par exemple le système EIES (Hiltz et al. 1981) pour la prise informatisée de note en conférence autorise la définition de commandes par l'utilisateur. Les auteurs rapportent que leur utilisation est perçue comme plus utile pour les utilisateurs experts, car cela permet un contrôle proactif. De plus cette personnalisation correspond mieux à leurs besoins. (Rubine 1991) élabore un algorithme de reconnaissance de gestes sur assistant personnel (*PDA*) qui est notamment intégré dans plusieurs systèmes (*NeXT Application Kit*, *Andrew Toolkit* et l'environnent *Garnet*). Cela permet aux concepteurs d'ajouter une interface gestuelle en faisant apprendre au système des exemples de gestes. Cette option fait aussi l'objet de brevets. Par exemple, celui déposé par (Marvit et al. 2007) décrit la possibilité de proposer des gestes pour lancer des applications ou des fonctionnalités sur mobile. À noter que la frontière reste mince avec des interfaces adaptables décrites précédemment, même si ici l'utilisateur est dans ce cas-ci plus à même de proposer des commandes.

À des fins d'aide à la personne, citons l'application *Wii-help* (Wassner 2008). Celle-ci permet à certains usagers inaptes à utiliser une souris de piloter le curseur à distance par les doigts et de prédéfinir des gestes pour accéder à certaines fonctionnalités. Il existe un certain nombre d'études de faisabilité et *benchmarks* de systèmes d'apprentissage artificiel en vue de proposer, entre autres, des systèmes adaptatifs, mais peu d'implémentations existent (Rautaray et al. 2012). Une intégration de (Zhang et al. 2008) a pour vocation de rendre accessible la composition de numéros de téléphone par des gestes définis par les utilisateurs. L'un des objectifs est de faciliter l'usage aux personnes aveugles ou âgées. Les auteurs soulignent que la proposition de réaliser l'apprentissage par les utilisateurs garantit une meilleure reconnaissance qu'avec une base de gestes préenregistrés et « génériques ». Mais l'article met principalement en avant la faisabilité technique.

Dans le domaine de la création sonore, il existe plusieurs applications de création de geste en relation au son. Nous retenons le système *Mogee* (Zamborlin 2012) qui permet de transformer tout objet en instrument de musique avec un microphone de contact. L'utilisateur peut enregistrer ses propres gestes qu'il « trace » sur différents objets et les faire reconnaitre par le système pour produire des sons (Tough 2013).

Enfin, à noter qu'en robotique, cette approche est notamment utilisée pour apprendre aux robots des activités par démonstration et imitation. Il apprennent ainsi à se repérer ou se déplacer dans l'environnement, et exécuter des tâches (Steil et al. 2006).

Utilisation de cette approche en conception inclusive

Cette approche renvoie au principe de conception inclusive généralement mis en œuvre dans les modèles de conception centrée utilisateur (Hurtienne et al. 2008). La conception inclusive a pour but d'aider à concevoir un dispositif utilisable par une majorité d'utilisateurs aux aptitudes différentes. Cela permet de couvrir des situations d'usage très larges, sans avoir recours à des adaptations ou une conception spécifique. De nombreuses études d'élaboration d'interfaces gestuelles pour la réalité mixte ou augmentée font appel à ce principe de conception inclusif. Elles permettent d'élaborer des classifications de gestes et d'aboutir à des consensus. Nous proposons une investigation rapide afin d'en extraire quelques préconisations :

(Good et al. 1984) proposent le concept l'Interface Dérivée de l'Utilisateur (traduit de *User Derived Interface*) dans la conception d'une interface client de gestion de courriel. Ils proposent une étude où l'utilisateur a la possibilité d'utiliser l'application en le laissant dicter ses propres commandes. Une partie de ces commandes est interprétée automatiquement par le système informatique. Celles qui ne sont pas reconnues sont traitées par un opérateur qui se fait passer pour la machine par la technique du « Magicien d'Oz ». De plus, il peut demander à l'utilisateur de les expliciter au besoin.

(Wobbrock et al. 2009) et (Ruiz et al. 2011) utilisent cette approche pour la mise en place d'interactions gestuelles sur dispositifs tactiles. Les auteurs remarquent que la métaphore du bureau influence les modèles mentaux des utilisateurs. Ils notent un certain nombre de consensus, mais les utilisateurs utilisent fréquemment des gestes similaires pour des interactions différentes. En revanche, certaines commandes ne donnent aucun consensus. Ces études ont permis d'aboutir à des classifications de consensus sur la nature, la forme, le référentiel ou la continuité des gestes. Plus récemment, une étude en réalité augmentée de (Piumsomboon et al. 2013) a permis d'établir une classification de gestes pour guider les designers à élaborer des interfaces centrées utilisateurs cohérentes. Les auteurs observent que plus la tâche a une signification abstraite (comme coller, copier, grouper des éléments, etc.), moins il y a de consensus (voir Figure 2-4). (Choi et al. 2012) se sont interrogés sur la pertinence de l'usage de commande gestuelle définie par l'utilisateur pour des systèmes domotiques. Dans un premier temps, ils ont laissé la possibilité aux participants de proposer des gestes associés aux contrôles de différents produits (comme allumer, éteindre, augmenter, diminuer pour la climatisation, la TV, la lumière). Puis, un mois plus tard, les mêmes utilisateurs ont dû choisir des gestes parmi ceux qu'ils ont proposés, ajoutés à ceux des autres participants pour les mêmes actions. Ainsi, 65% des gestes préférés dans la première expérimentation ont été changés dans la deuxième. Et il y a significativement plus de consensus dans la deuxième expérimentation.

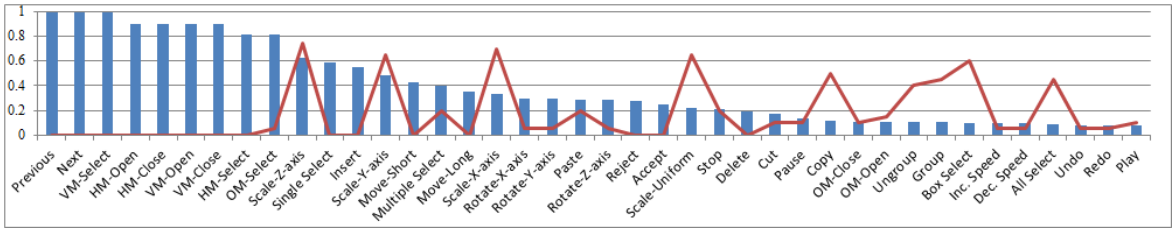


Figure 2-4 : Taux de consensus en ordre décroissant pour quarante tâches (en bleu) et ratio de l'utilisation de gestes bimanuels (ligne rouge) extrait de (Piumsomboon et al. 2013)

Dans le même registre, afin de contrôler une application de réalité augmentée, (Voida et al. 2005) demandent aux utilisateurs de proposer des gestes pour interagir avec de multiples écrans disposés dans la pièce. Si les écrans sont à portée de bras, certains utilisateurs préfèrent les toucher directement. En revanche, pour les écrans à plus de 60 cm de distance des utilisateurs, ces derniers semblent privilégier la manipulation distante plutôt que de se déplacer de leur siège. Les utilisateurs effectuent essentiellement des gestes de pointage du doigt pour désigner les écrans. À noter, les mouvements de faible amplitude de la main ou qui impliquent un déplacement minimal du corps sont privilégiés. Notons toutefois que ces conceptions inclusives tendent au final vers l'élaboration de consensus afin de proposer des interactions standardisées. En résumé, nous obtenons ces tendances et ces limites :

Table 2-2 : Tendances et limites en conception inclusive

✓ Geste mimétique pour action concrète	× Multiplication de l'interaction en usage avancé
✓ Réutilisation des connaissances acquises	× Geste parfois difficilement reproductible
✓ Contact colocalisé privilégié	× Geste similaire pour commandes différentes
✓ Si contact à distance, « télécommande » et mouvements peu fatigant privilégiés	× Geste parfois non approprié à un usage répété

→ Par conséquent, à notre connaissance, il n'existe pas d'évaluation avancée sur l'usage d'une modalité d'interaction définie par l'utilisateur. Cela s'explique par le fait que les cas d'usage dédiés restent trop spécifiques. Comme le précisent (Mäntyjärvi et al. 2004), le système d'apprentissage artificiel reste très complexe à mettre en œuvre, bien que certaines solutions robustes existent. Pourtant cette approche où l'utilisateur définit l'interaction permet une prise en compte de la singularité de l'usager, à la fois axée sur les besoins et sur les possibilités sensorimotrices. Nos perspectives d'usage et cas d'application nous orientent vers cette approche clairement centrée-utilisateur. Comme il ne semble pas exister d'évaluations concrètes, il n'existe pas par conséquent de « guide » de conception d'un système de réalité virtuelle incluant cette approche et son évaluation. Afin de mettre en place cette approche, nous allons nous pencher dans le paragraphe suivant sur des principes de conception avec une approche centrée-utilisateur.

2.1.3. Concevoir une expérience centrée utilisateur

(Burkhardt et al. 2006) distinguent deux principales approches pour la conception en réalité virtuelle : la première est plutôt centrée sur la perception et la copie du réel tandis que la seconde est centrée utilisateur. L'approche centrée-perception se concentre sur les habiletés psychomotrices et omet de considérer avant tout l'environnement virtuel comme un médiateur de l'activité pour l'utilisateur. (Bobillier-Chaumon et al. 2005; Pirhonen 2010) notent que la conception centrée-utilisateur a le double mérite d'offrir une rupture avec une vision « technocentrée » et d'insister d'autre part sur l'aspect qualitatif de la nature de l'interaction. Il s'agit « d'adapter la conception au fonctionnement cognitif de l'utilisateur, à ses objectifs et à ses besoins et exigences ». (Hutchins et al. 1986) proposent d'appréhender l'aspect qualitatif de l'expérience de l'utilisateur et en particulier les aspects critiques de l'usage. Cela permet une meilleure adaptation des interfaces en évitant de n'évaluer que la performance seule. (Laurel 1986) ajoute qu'il est important que l'interface puisse disposer d'un contexte personnalisé et soit agréable à utiliser.

La conception d'une interface n'étant pas une science exacte, on note que la littérature sur les méthodes existantes reste très largement basée sur les heuristiques ou des résultats empiriques. Nous exposons dans cette partie les différentes méthodes centrée-utilisateur afin de concevoir et d'évaluer une modalité d'interaction définie par l'utilisateur.

2.1.3.a *Principes généraux de conception centrée-utilisateur*

Théorie de l'action

Afin de formaliser la conception centrée utilisateur, (Norman 1989a) propose la « Théorie de l'action » en référence à l'approche écologique de (Gibson 1966, 1979). Celle-ci s'articule autour de deux points : les contraintes écologiques et les affordances que nous avons vues précédemment. Norman propose une conception et une évaluation en considérant le modèle de l'usager et le modèle du système. Il définit l'interaction comme un processus dynamique qui peut être décrit en sept phases d'accomplissement d'une activité en partant du but :

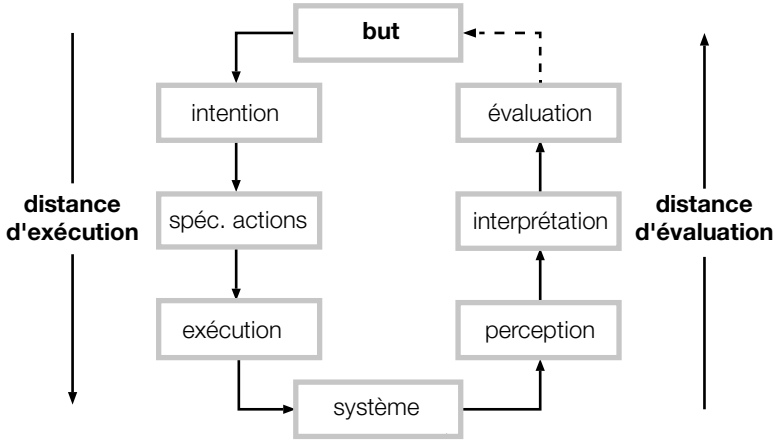


Figure 2-5 : Sept phases de l'interaction homme système selon (Norman 1989a)

- Définir le but ;
- Former une intention ;

- Spécifier une séquence d'actions ;
- Exécution d'une action ;
- Percevoir l'état du système ;
- Interpréter l'état du système ;
- Évaluation du système du point de vue des buts et des intentions.

Toujours selon Norman, la conception en elle-même doit appliquer quatre critères de conception au regard de la tâche :

- Présenter des affordances afin de donner les possibilités d'action en exploitant les connaissances du monde et de l'utilisateur. Cela oblige à simplifier l'interface et évite à l'utilisateur un apprentissage superflu ;
- Visibilité de l'état du système en tout temps. Les fonctionnalités doivent être visibles avant et pendant l'action ;
- Le couplage naturel (traduit de *natural mapping*) signifie qu'il y a une correspondance directe entre l'action et le résultat. Ce couplage peut être effectué dans le cadre de contraintes naturelles ou symboliques. Celles-ci guident le comportement et facilitent l'usage ;
- La rétroaction permet de donner un retour à l'utilisateur qui doit donc pouvoir percevoir l'état du système, et l'interpréter. Si cette rétroaction est continue, elle est liée au couplage.

Norman (Norman 1989b) ajoute qu'il est important de pouvoir gérer les erreurs, en laissant la possibilité à l'utilisateur de faire marche arrière ou de corriger. Enfin, si rien n'est possible, Norman préconise de standardiser. Cela borne fortement l'utilisation et oblige l'utilisateur à un apprentissage, mais la standardisation a l'avantage de fiabiliser ou sécuriser l'utilisation (cf. 2.1.2.a).

De nombreux auteurs proposent des recommandations similaires à celles de Norman, mais détaillent certains des quatre critères de conception décrits précédemment, et en ajoutent d'autres. Par exemple (J. Nielsen et al. 1990) préconisent la cohérence dans l'utilisation, permettant une réutilisation de comportements similaires pour l'usage. La même action aura ainsi toujours le même effet (par exemple, toujours placer un bouton de *retour arrière* au même endroit sur une interface dynamique). Il ajoute qu'il est important de prévoir les erreurs et de les reporter. (Tognazzini 1999) privilégie l'ergonomie en favorisant l'exploration et l'efficience en réduisant la latence. Il propose d'intégrer la loi de Fitts (Fitts 1954) si cela est pertinent. Pour rappel, cette loi modélise les tâches de pointage ou de sélection. Le modèle et l'expérience montrent que redimensionner la cible en fonction de la distance facilite le pointage. Enfin, toujours dans un souci d'utilisation simplifiée, (Shneiderman 1998) préconise de faciliter l'utilisation en montrant explicitement que l'utilisateur a le contrôle. Il recommande aussi d'éviter de solliciter la mémoire à court terme.

(Bobillier-Chaumon et al. 2005) soulignent que le principe de conception de Norman met en évidence les divergences possibles entre les variables psychologiques et les variables physiques du système. Cela permet de mettre au jour des problèmes de corrélations entre les intentions de l'usager transposées en actions admises par le système (distance d'exécution), et en parallèle, l'interprétation de l'état du système en fonction des buts et intentions de l'usager (distance d'évaluation). Ils insistent enfin sur le fait d'accorder une attention aux dimensions

humaines dans la conception d'un logiciel et de son interface. En effet, elles restent pour l'utilisateur l'unique moyen de se représenter le système et d'y agir.

Théorie de l'énaction

Comme nous l'avons expliqué, l'adaptation centrée sur l'homme implique de se référer à son ancrage écologique et au caractère incarné de sa cognition. Il semble pertinent de prendre en compte le principe de l'énaction (Varela et al. 1991) dans la conception d'un environnement virtuel et des interactions. (Thouvenin 2009) rappelle cependant que « ce n'est pas l'interface qui est (ou n'est pas) énactive c'est l'humain, utilisant une interface qui rend le monde énactif ». Ainsi, un certain nombre de travaux en intelligence artificielle proposent des alternatives aux approches informatiques issues du cognitivisme ou du connexionnisme intrinsèquement basées sur des représentations. (Winograd et al. 1989) proposent de repenser la conception d'une intelligence artificielle en se centrant sur la métaphore du vivant, la machine se positionnant comme une ressource à l'activité. (de Loor et al. 2008), proposent aussi un canevas pour la mise en place d'une intelligence artificielle énactive. Les auteurs remarquent deux problèmes : une évolution « non contenue » du système artificiel (ontogénèse) et la pertinence du sens créé par la machine. Les auteurs mettent l'emphase sur le guidage interactif de l'environnement permettant un développement ontogénique approprié. Ils proposent également de considérer la place de l'homme dans la boucle : ceci permettrait de pallier la question du sens.

Le principe d'énaction appliqué à la réalité virtuelle fait appel au paradigme fort du couplage structurel entre l'organisme et son environnement. (de Loor et al. 2008) et (Thouvenin 2009) soulignent que la conception d'une interface de réalité virtuelle est propice à l'implémentation d'une boucle énactive, car il y est possible de modéliser le système, mais également l'interaction entre l'homme et ce système. Ce couplage structurel homme-système ne peut exister que si les boucles sensorimotrices s'auto-adaptent au fil de l'expérience. Ces propos sont illustrés par l'usage d'un dispositif haptique, parce qu'il permet la palpation par l'homme, intrinsèquement active, il donne un ressenti du contour d'un objet, de sa texture, de sa forme, etc. À noter que ce couplage n'existe que si le retour d'effort est recalculé à une fréquence suffisamment élevée (1000Hz). Dans le même registre, le suivi du point de vue dynamique (*headtracking*) peut être considéré comme « énactif », car il permet la compréhension de l'environnement par une perception active. Enfin, un élément constituant du paradigme de l'énaction est l'autonomie des systèmes.

Théorie du *« flow »*

Dans une optique différente paraissant presque à l'encontre d'une adaptation, (Csíkszentmihályi 1990) propose un principe de conception basé sur l'engagement et la réussite par le challenge. Il est à l'origine de la notion de *flow* principalement mise en œuvre dans les jeux vidéo. L'activité doit avoir un but précis et bien compris, et la rétroaction doit être immédiate. La tâche entreprise se présente comme un défi et exige une aptitude particulière. Cela met en jeu l'implication et la concentration de l'utilisateur. L'utilisateur est absorbé et son attention est canalisée. Selon l'auteur, cela donne à l'utilisateur le sentiment d'exercer un contrôle sur son action, tout en entrainant une perte du sentiment de conscience de soi. L'utilisateur ainsi captivé ne voit plus passer le temps et éprouve une grande satisfaction.

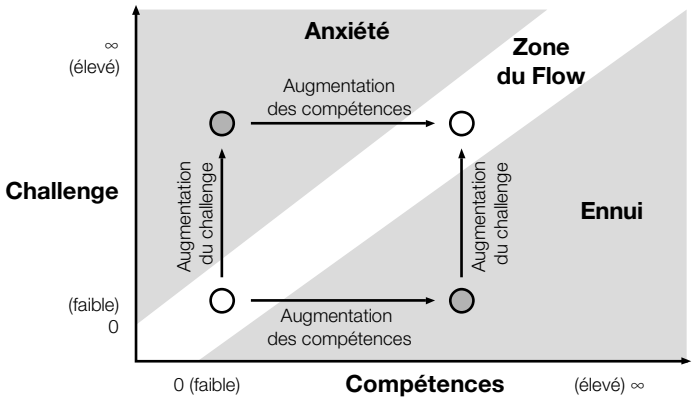


Figure 2-6 : Notion de *flow* (Csíkszentmihályi 1990)

Cette question de challenge et des aptitudes est à prendre en considération notamment dans l'implication en rééducation ou dans la réalisation d'une performance artistique. La motivation y a un rôle important et doit contrer l'ennui, le stress, ou l'anxiété notamment dans l'activité de rééducation. Dans les processus créatifs, l'expressivité du corps en interaction est directement en lien avec les composantes du *flow* citées précédemment. La mise en œuvre de ce principe nécessite de trouver le bon équilibre entre habileté de l'utilisateur et challenge dans l'activité. Son application se destine plutôt à favoriser le dépassement de soi. Elle serait pertinente pour l'ajustement de la tolérance requise pour effectuer une tâche dans un système adapté.

2.1.3.b *Principes de conception centrée-utilisateur en réalité virtuelle*

Principes de conception généraux

Toujours dans l'optique de placer la dimension humaine au cœur des dispositifs de réalité virtuelle, nous allons nous pencher sur quelques méthodes et principes de conception. (Lampton et al. 1994) proposent une méthode de conception pour une application d'entrainement militaire. Ils définissent des tâches spécifiques (se déplacer, communiquer, utiliser un objet) et mettent en place des fonctionnalités du simulateur appropriées (locomotion, vision, manipulation, son, suivi). De leur côté, (Bowman et al. 2004) proposent un ensemble de préconisations pour la conception et le développement d'interfaces utilisateur 3D centrées sur :

- la prise en compte des caractéristiques de l'homme ;
- la définition des rétroactions ;
- les contraintes (comme vu précédemment) ;
- les interactions bimanuelles ;
- la définition de groupes d'utilisateurs ;
- la prise en compte du confort d'utilisation.

Concernant la création de nouvelles interfaces 3D, ils proposent trois stratégies. La première se base sur la réplication du monde réel, la seconde sur l'adaptation de techniques pour l'interaction en 2D. Enfin, ils proposent de « conférer une certaine magie et une esthétique » dans l'implémentation d'interfaces 3D. Comme le notent (Tyndiuk et al. 2003) les deux méthodes décrites précédemment ont pour vocation d'identifier principalement les facteurs

liés à l'utilisateur, à la tâche, au système et à l'environnement qui influencent la performance. Par conséquent, ils ne rendent pas compte des influences réciproques de ces différents facteurs.

D'autres recherches tentent de préciser ces influences comme les travaux de (Verna 2000) qui proposent un « Modèle Représentatif de l'Interaction Cognitive » (MRIC). Il y décrit le processus de virtualisation soit les mécanismes cognitifs qui font qu'à partir d'une situation réelle, l'homme entre progressivement dans le virtuel. Bien que ce modèle prenne en compte les influences entre facteurs cités précédemment, (Tyndiuk et al. 2003) font noter qu'il permet essentiellement de catégoriser les applications de réalité virtuelle.

Principes de conception centrés sur l'activité « 3I² »

Afin d'accorder une importance directe aux dimensions humaines dans la conception, (Fuchs et al. 2003) proposent un modèle en réalité virtuelle qui associe « modèle usager » et « modèle système » sur les besoins fonctionnels. Ce modèle se focalise essentiellement sur les processus cognitifs mis en jeu au regard de l'activité dans le monde virtuel. L'interfaçage permettant l'activité sensorimotrice doit demeurer le plus transparent possible. Pour ce faire les auteurs échelonnent leur modèle sur trois niveaux dont la clé de voute est l'immersion et l'interaction (I²) :

- Le premier niveau concerne les **I² sensorimotrices**. Il permet l'interfaçage, c'est-à-dire l'association de l'activité sensorimotrice de l'utilisateur :
    - avec différentes interfaces motrices ;
    - avec des dispositifs de restitution sensorielle et les interfaces sensorimotrices.

La transparence de l'interface au niveau des I² sensorimotrices reste le critère essentiel, l'utilisateur ne doit pas s'en préoccuper.

- Le deuxième niveau permet des **I² cognitives**, c'est-à-dire que l'homme peut comprendre et interagir avec le système. Les auteurs y distinguent les deux notions décrites précédemment : l'emploi des schèmes (Piaget 1971) et des métaphores pour l'interaction.
    - L'utilisateur peut effectuer une activité en environnement virtuel en adoptant le comportement qu'il aurait habituellement devant une situation similaire réelle, on parle de schème comportemental importé (Piaget 1971) ;
    - Dans les cas où le schème n'est pas envisageable, l'autre possibilité est d'employer une métaphore. Les auteurs distinguent deux types :
        - Les métaphores sans substitution sensorimotrice correspondent à une image symbolique utilisant le même canal sensorimoteur ;
        - Les métaphores avec substitution sensorimotrice font appel à un autre canal. Par exemple, il est possible de symboliser la collision entre deux objets par un changement de couleur.

- Enfin, le troisième niveau, appelé niveau des **I² fonctionnelles**, regroupe les différentes activités possibles dans l'environnement virtuel. Elles sont regroupées en quatre catégories appelées primitives comportementales virtuelles (PCV) :
    - L'observation ;
    - Le déplacement ou la navigation ;
    - L'action sur le monde virtuel ;

- La communication avec autrui ou avec l'application.

Ces trois niveaux se parcourent ainsi : pour concevoir un système de réalité virtuelle, il faut partir de l'activité en définissant les PCV du niveau fonctionnel, puis choisir les modalités d'I² cognitives appropriées et enfin sensorimotrices. Il demeure important de considérer la notion de transparence de l'interface. Pour évaluer un système de réalité virtuelle, il faut parcourir ces trois niveaux dans le sens inverse, en commençant par vérifier la qualité et la transparence de l'interfaçage, la compréhension des modalités cognitives et enfin voir si l'utilisateur arrive à effectuer l'activité désirée dans l'environnement virtuel.

2.1.4. Synthèse

Nous avons vu que l'approche qui permet à l'Utilisateur de Définir l'Interaction (UDI) semble correspondre aux besoins inhérents de nos cas d'application. Nous constatons que bien qu'il existe quelques implémentations, aucune évaluation d'usage appropriée n'a été menée. Afin de concevoir un système qui intègre cette approche centrée-utilisateur, nous nous orientons naturellement vers une conception centrée-utilisateur.

Étant donné qu'il n'existe pas de modèle général, nous nous inspirons de plusieurs méthodes pour mettre en place une activité intrinsèquement centrée-utilisateur. Ainsi, les principes de la théorie de l'action, de l'énaction et le modèle 3I² considèrent explicitement l'homme, le système et leur interdépendance dans l'expérience. Côté utilisateur, Norman souligne l'ancrage écologique et rappelle l'intérêt des contraintes écologiques où l'activité est guidée par l'environnement et celui des affordances qui offrent les possibilités d'action. Concernant l'activité même en environnement virtuel, le modèle 3I² isole clairement les types d'activité grâce au principe de Primitives Comportementales Virtuelles (PCV). L'exécution de ces PCV requiert ainsi une interaction qui a ici un caractère dédié. Nos cas d'usage nous orientent vers une modalité gestuelle qui a le double mérite d'être plus écologique et plus directe. Les gestes du bras ou de la main semblent appropriés afin de tester ce système, mais toute partie motrice du corps est envisageable pour interagir.

Il reste à comprendre ce qu'est un geste et comment il s'exécute, notamment lorsque celui-ci est dirigé vers une activité en environnement virtuel. En vue de l'implémentation d'une interaction gestuelle adaptée et pertinente, comment cette dernière peut-elle être apprise et interprétée par un système ? Un certain nombre d'évaluations d'usage devront être entreprises pour répondre à cette question.

2.2. CARACTÉRISER ET RECONNAITRE LE GESTE POUR INTERAGIR EN ENVIRONNEMENT VIRTUEL

Avant de comprendre la nature ou l'objectif du geste, rappelons qu'il est basé sur un mouvement du corps. (Baudel et al. 1993) montrent que la gestuelle a l'avantage d'être une forme naturelle de communication et d'être rapide à apprendre. (Braffort et al. 2006) remarquent une particularité du geste qui le différencie des autres modalités sensorimotrices : sa fonction motrice et sa fonction sensorielle sont indissociables. De plus, pour ajouter un peu de complexité, il existe une influence réciproque geste et action. Par exemple, (Goldin-Meadow et al. 2013) notent que dans la communication orale, le geste reflète les pensées, y compris le non-verbal, mais il peut réciproquement influencer la pensée de celui qui parle. Dans ce cas, le geste se révèle être à la fois un moyen de communication et un « outil pour la pensée » du locuteur.

Afin de mieux comprendre le geste, nous proposons un état de l'art sur sa caractérisation selon l'approche de (M. Nielsen et al. 2004; Zhao 2001). Nous analysons le geste dans sa naturalité et en vue d'une implémentation en informatique. La première caractérisation est centrée sur la sémantique du geste et sa directivité vis-à-vis d'une activité et de l'environnement. La seconde est une caractérisation descriptive, quantitative ou qualitative, à des fins d'interprétation par un système. Nous décrivons quelques techniques d'apprentissage et de reconnaissance du geste. Nous détaillons ensuite comment mettre en place une interaction gestuelle au regard de notre problématique d'adaptation. Nous verrons enfin quelques considérations d'ergonomie dans le cadre d'une interaction gestuelle.

2.2.1. Caractériser la gestualité de l'homme

> *A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on it's way to hitting a key is neither observed nor significant. All that matters is which key was pressed.*

(Kurtenbach et al. 1990)

Cette citation montre qu'il existe un écart entre le potentiel d'expression d'un geste et l'interprétation possible par un système informatique. (Jensenius 2007) soulève que dans le cas d'une interaction avec un système, il existe plusieurs façons de transcrire le geste aussi bien de façon quantitative que qualitative. Nous souhaitons capturer l'essence du geste nécessaire à la réalisation d'une activité en environnement virtuel. En effet, l'activité d'apprentissage (formation, cours, rééducation, *serious game*) ou la pratique artistique (performance, critères d'esthétique ou d'expressivité) soulèvent la question de l'analyse quantitative ou qualitative du mouvement. Afin de mieux comprendre le geste, nous allons explorer sa nature, et l'analyser en vue de le faire apprendre et reconnaitre par un système informatique.

2.2.1.a *Caractérisations générales du geste*

Classifications sémantiques

Pour comprendre la nature du geste, nous proposons de considérer plusieurs classifications sémantiques du geste selon l'association à la parole. En étudiant l'association du geste à la parole, (Kendon 1988, 2004) propose d'analyser la gestualité selon sa symbolique. Il élabore dès la fin des années 1980 un premier continuum allant de la gesticulation idiosyncrasique à la langue des signes (Figure 2-7), auquel la communauté se réfère encore aujourd'hui :

- La gesticulation correspond à un geste spontané (mouvement des bras ou de la main) qui accompagne la parole ;
- Les gestes quasiment linguistiques aux mêmes caractéristiques que les précédents, mais sont intégrés à la parole ;
- La pantomime, qui correspond aux gestes sans paroles, que l'on retrouve notamment au théâtre ;
- Les emblèmes sont des gestes d'éloge, voire d'insulte ;
- Les gestes des langues des signes incluant les gestes et postures qui forment un système de communication linguistique à part entière.

(McNeill 1992) propose une classification qui considère le geste dans un sens plus large. En effet, il peut refléter les pensées, sentiments, désirs, et actions. Il propose une classification qui n'est plus uniquement rattachée à la parole et qui comprend :

- Les gestes iconiques utilisés pour décrire un objet ou une action concrète ;
- Les gestes métaphoriques qui correspondent à l'image qui se réfère à une idée abstraite ;
- Les gestes déictiques qui servent à désigner un objet et sont utilisés en narration ;
- Les gestes de pulsations aussi appelés battements qui rythment le discours ou en accentuent certaines parties. Ils ne portent pas de sens propre.
- Les emblèmes qui sont en lien avec des signes symboliques, comme dire « au revoir » par un signe de la main.

(Kendon 2004) s'appuyant sur la démarche de McNeill propose d'étayer son continuum en y incluant la parole :

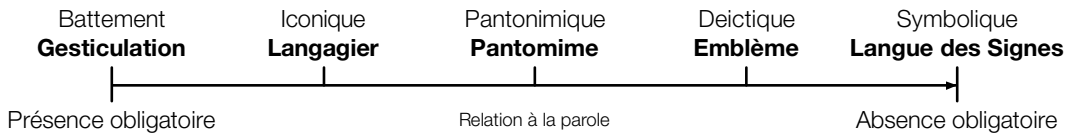


Figure 2-7 : Le continuum de Kendon en relation à la parole

On constate que le geste est porteur de sens et est fortement corrélé au fonctionnement de la parole. Il existe d'autres classifications du geste concernant la finalité du geste, sa fonction ou la direction du geste vis-à-vis de l'environnement.

Formes gestuelles

Dans le domaine des sciences du sport, certains auteurs proposent une distinction entre les habiletés topocinétiques et les habiletés morphocinétiques. On retrouve cette caractérisation

chez plusieurs auteurs (Paillard 1974; Serre 1984) dans les théories motrices et communicatives. (Cole et al. 2002) font également cette distinction en y ajoutant les aspects phénoménologiques dans leur « théorie incarnée du geste » (traduit de *embodied or integrative theory of gesture*). Les auteurs considèrent la topocinèse comme basée sur le contrôle moteur et essentiellement dépendant de la proprioception et des processus liés au schéma corporel. Ils distinguent en parallèle la morphocinèse comme étant plutôt liée au contrôle communicatif ainsi qu'à la sémantique. Serre (Serre 1984) fait une troisième distinction en séparant la sémantique de la morphocinèse et parle ainsi de sémiocinèse. Cela nous donne trois approches :

- La topocinèse (ou téléocinèse) désigne l'activité motrice dont le but est d'atteindre un point de l'espace et est tournée vers l'efficacité de l'action. Le moyen d'y parvenir oblige à effectuer un mouvement pour atteindre un point, c'est le cas dans les sports coopératifs, de combat, de tir et dans une considération plus large une tâche de pointage ;
- La morphocinèse a pour but de produire une forme conforme à un code. L'espace ou le temps sont des moyens de produire cette forme. On retrouve cette notion en gymnastique, en danse, ainsi que dans l'écriture ;
- La sémiocinèse est plutôt considérée comme une forme corporelle pour produire du sens et des émotions chez les spectateurs, que l'on retrouve notamment dans la danse.

Cette distinction de la forme gestuelle serait à prendre en considération au niveau de la description gestuelle plutôt que dans l'activité elle-même, et sous un angle qualitatif. Nous verrons qu'il existe différentes descriptions quantitatives et qualitatives du geste en informatique que nous détaillons dans le paragraphe 2.2.2.b (Représentation des données : traitement et descripteurs gestuels).

Directivité du gestuel par rapport à l'environnement

Dans une perspective fonctionnelle orientée vers un contrôle gestuel, (Cadoz 1994) propose de classer les gestes selon leur rapport à l'environnement, il distingue :

- La fonction ergodique « geste ↔ environnement » qui concerne les actions matérielles de manipulation, régulation, ou modification de l'environnement physique. Cette fonction peut être appelée haptique ou instrumentale (Jensenius 2007) ;
- La fonction épistémique « geste ← environnement » qui consiste à percevoir l'information depuis l'environnement comme le cas d'une exploration haptique ;
- La fonction sémiotique « geste → environnement » qui permet à l'inverse la communication d'informations vers l'environnement. Dans ce cas, le geste transmet et exprime quelque chose (Braffort et al. 2006). On la retrouve dans communication non-verbale et dans l'interaction en environnement virtuel.

Cette distinction permet de comprendre la directivité du geste. La plupart des systèmes d'interaction gestuelle seraient orientés « sémiotique ». Cependant, la fonction ergodique a un rôle par rapport aux rétroactions produites par l'environnement virtuel. (Quek et al. 2002) notent qu'un geste de manipulation d'objets réels relève d'une fonction essentiellement ergodique parce qu'il y a contact physique. La question de savoir si une interaction gestuelle sans retour haptique relèverait de la sémiotique reste ouverte. Si ce retour est une substitution sensorielle (comme la vision), elle relèverait plutôt de l'ergodique.

Le geste pour l'interaction en environnement virtuel

Dans l'application à la réalité virtuelle, (Braffort et al. 2006) remarquent qu'il existe principalement trois types de gestes selon leur capacité à se combiner entre eux ou avec d'autres modalités :

- Les gestes simples correspondent à des gestes de commande (ou contrôle) sans combinaison possible ;
- Les gestes coverbaux peuvent se combiner notamment avec un message vocal afin de former un énoncé plus complet ;
- Les gestes des langues des signes qui se combinent entre eux en gardant des propriétés linguistiques comme celles d'une langue naturelle.

Les auteurs précisent que le geste peut se positionner « en entrée » et « en sortie » du système. Le geste en entrée correspond à l'interaction par le geste (manipulation directe ou commande et contrôle gestuel). Le geste « en sortie » renvoie à la notion d'humanoïde ou d'avatar aux capacités expressives, notamment par le geste. Puisque nous ciblons une interaction par le geste, ces travaux de recherches se consacrent essentiellement au geste « en entrée ».

2.2.1.b *Systèmes de représentation et de transcription du geste*

De nombreuses approches existent sur la transcription du geste. Cela aboutit à des systèmes de notation et de codage gestuels qui peuvent inspirer la caractérisation du geste en descripteurs pour un système informatique.

Transcription pour la langue des signes

Le découpage en structures temporelles a permis la segmentation de geste notamment dans la langue des signes. Il existe également un découpage en structures paramétriques pour les systèmes de notation. (Stokoe 1972) propose que le signe puisse être caractérisé par trois paramètres : configuration, emplacement et mouvement. À cela peuvent s'ajouter des informations sur l'orientation et la mimique du visage. Cette approche encore majoritairement utilisée décrit le geste principalement sous forme iconique (il existe des logiciels de transcription en symboles de gestes issus de vidéo de langue des signes). (Braffort et al. 2006) ajoutent que des approches récentes tâchent de dépasser cette représentation relevant essentiellement de l'iconique. Elles permettent de prendre davantage de paramètres en compte, comme l'intention du signeur.

Systèmes d'annotations du mouvement en danse

Dans l'idée de représenter le mouvement plutôt par la cinématique ou la morphologie du corps en relation avec l'environnement, il existe un certain nombre de systèmes de notation. Le mouvement peut être représenté suivant une dimension temporelle à la manière d'une partition de musique. Ainsi la choréologie permet de formaliser ces mouvements. (Benesh et al. 1956) proposent une notation à trois vocations :

- Une représentation formelle du mouvement (déplacements, positions des parties du corps), des relations entre les individus entre eux et avec l'environnement matériel ;
- L'analyse du mouvement : rythme, phrasé, etc. ;
- La transmission d'une chorégraphie hors support vidéo, image, son, etc.

Il existe aussi le système de notation de Eshkol-Wachman (Eshkol et al. 1958) que l'on retrouve dans l'annotation du mouvement pour la danse, la langue des signes et la physiothérapie. Cette notation se base essentiellement sur la représentation du corps, des membres et/ou des articulations. Il existe d'autres représentations plus spécifiques à la danse et permettant une représentation plus qualitative du mouvement. Parmi les plus connues, le *labanotation* (Laban et al. 1971) permet une représentation symbolique des gestes dansés qui décrit le mouvement selon quatre paramètres :

- Le paramètre corporel (corps) qui identifie les caractéristiques physiques du corps en déplacement ;
- Le paramètre d'espace (espace) qui correspond à l'occupation du corps dans de l'environnement. Laban propose l'idée de « kinésphère » comme espace imaginaire personnel englobant le corps et accessible par les membres ;
- Le paramètre de forme (forme) qui considère le corps comme forme géométrique globale qui peut changer ou évoluer par rapport à l'espace l'environnement ;
- Le paramètre d'effort dynamique (effort) qui décrit les aspects qualitatifs du mouvement.

Quelques modèles du mouvement humain

Il existe dans les théories du contrôle moteur un certain nombre de façons de transcrire le geste. La modélisation biomécanique du mouvement humain peut se faire par techniques cinématique ou dynamique (Bouisset et al. 2006). L'approche cinématique mesure le déplacement des segments corporels. La chaîne cinématique correspond à l'assemblage de segments reliés entre eux par des articulations. Ses segments sont contraints par leur articulation avec d'autres segments. Une chaîne classique est représentée par des segments en série et peut être définie par des degrés de liberté (DDL). La chaîne cinématique directe décrit le mouvement articulation par articulation. La chaîne cinématique inverse décrit le mouvement de façon globale en partant d'un bout de la chaîne, comme la main dans le cas du bras et le mouvement du coude et de l'épaule seront déduits. L'approche dynamique correspond au processus de calcul du mouvement généré par des accélérations. La dynamique inverse est le processus de calcul des forces qui générerait un mouvement donné.

Il existe des modèles de prédiction du mouvement humain basé sur ces modèles. La loi de Fitts (Fitts 1954) décrite précédemment permet de modéliser une tâche de pointage. La loi des puissances 2/3 de (Viviani et al. 1980) lie vitesse de déplacement et rayon de courbure de la trajectoire du geste. Ainsi, la vitesse augmente dans les parties peu courbées de la trajectoire et inversement dans les parties plus courbées. Les mouvements peuvent aussi être modélisés par prosupination (pronation ou supination), flexion et extension. (Boutet 2008) propose par exemple une schématisation des gestes par enchaînements de mouvements pro supinateurs, mettant en avant la dynamique du mouvement.

D'autres approches décrivent le geste au regard de la tâche. Par exemple, en robotique, (Lyons et al. 1989) proposent de modéliser une tâche de préhension de façon qualitative avec des critères centrés sur la forme, la taille de l'objet et la précision ou la fermeté requise pour effectuer la saisie. Enfin, il existe des possibilités de modéliser le mouvement par des données issues de signaux musculaires (électromyographie), mais comme le précise (Bouisset et al. 2006) la mise en œuvre demeure complexe.

2.2.1.c *Segmentation en unités gestuelles*

Afin d'étudier la structure du geste et plus précisément sa structure temporelle, nous allons nous pencher sur quelques travaux de référence, principalement basés sur la gestuelle coverbale (McNeill 1992) et la langue des signes (Liddell et al. 1989). Le geste peut être considéré comme un événement avec un début et une fin et appartenant au domaine de la parole. Le principe est de segmenter le geste en différentes structures qui s'enchaînent temporellement.

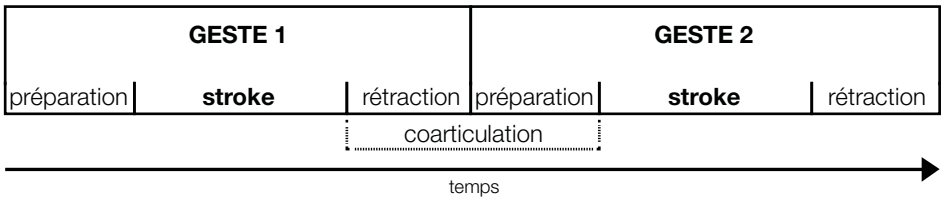


Figure 2-8 : Découpage d'un geste et phénomène de coarticulation entre deux gestes, adapté de (Kita et al. 1998)

Kendon et McNeill proposent de décomposer ces structures sous forme d'unités et de phrases consécutives (Figure 2-8). La partie signifiante du geste peut être de deux types selon sa dynamique. Dans le cas du geste dynamique, la partie signifiante est le *stroke*, soit le mouvement signifiant. S'il s'agit d'une posture, on parle de *hold,* soit un geste tenu. Avant et après la partie signifiante, on distingue respectivement les phases de préparation et rétraction. La phase de préparation du geste permet le déplacement de la main de sa position de départ à la phase de *stroke*. La partie où la main retrouve sa position de repos est la rétraction. Les phases de préparation et de rétraction peuvent être soit omises, soit influencées ou mélangées par les activités précédentes ou suivantes. Ce phénomène est appelé coarticulation.

→ Ces descriptions du geste permettent de mieux comprendre sa nature. Nous avons vu que le geste peut être fortement corrélé à la parole, aussi bien dans le sens que dans sa structure. Le geste peut être dirigé vers une activité. La relation à l'environnement et les réciprocités posent des questions relatives au couplage du geste et de l'action. En vue d'une implantation d'une interaction gestuelle, nous allons nous pencher sur les techniques d'acquisition, de reconnaissance du geste et sur l'interaction.

2.2.2. L'interaction gestuelle par apprentissage

Comme le note (Lee 2010), la reconnaissance du geste partage beaucoup des problématiques de la reconnaissance de la parole. Cela se rejoint d'ailleurs les analyses de McNeill et Kendon. On distingue un certain nombre de problématiques :

- La prise en compte des différents accents et dialectes ;
- L'expression de la pensée par des formulations différentes ;
- La proximité phonétique des mots (homophonie) ;
- La reconnaissance d'un mot dans un contexte est peu fiable, avec du bruit ;
- L'extraction de la partie signifiante du discours (notion d'embrayage et de débrayage).

L'option de laisser l'utilisateur faire apprendre au système ses propres gestes permet de pallier les différences interindividuelles. (Lee 2010) remarque que la différenciation des gestes, tout comme la parole, demeure difficile. Des gestes proches peuvent s'avérer difficiles à différencier. Le choix des techniques d'apprentissage artificiel et celui des descripteurs informatiques du geste sont capitaux pour permettre une reconnaissance optimale.

Embrayage et débrayage

Il est important de déterminer quand l'utilisateur souhaite interagir. Ceci rejoint les problématiques de la reconnaissance de la parole et du « syndrome de l'immersion » ; syndrome que l'on peut illustrer par l'expression « tout ce que vous direz sera retenu contre vous ». On retrouve également dans la littérature le terme de « *Midas Touch Syndrom* » (Lee 2010). Cela provient de la mythologie grecque « tout ce que touche Midas se transforme en or ». Le geste peut être interprété par le système, qu'il soit volontaire ou non. Cela peut empêcher l'utilisateur de communiquer en parallèle avec une autre personne ou un autre système. Dans la reconnaissance de la parole, il existe néanmoins des segmentations manuelles par bouton (comme sur les microphones) ou par un vocabulaire spécifique, comme l'usage de mots signalant le début et la fin d'un enregistrement. Cela amène aux notions d'embrayage et de débrayage.

Gestion d'un système d'apprentissage et de reconnaissance du geste pour l'interaction

La plupart des systèmes de reconnaissance de gestes en informatique suivent quatre principales étapes. Elles se succèdent le plus souvent de façon séquentielle :

- L'acquisition de données, qui inclut la capture des données brutes et leur traitement éventuel ;
- Leur représentation en valeurs numériques structurées (primitives) ;
- La classification qui permet de transformer le flux de données en symbole qui constitue le vocabulaire de gestes de l'application ;
- L'interprétation qui permet de traduire ces éléments symboliques en fonction du contexte.

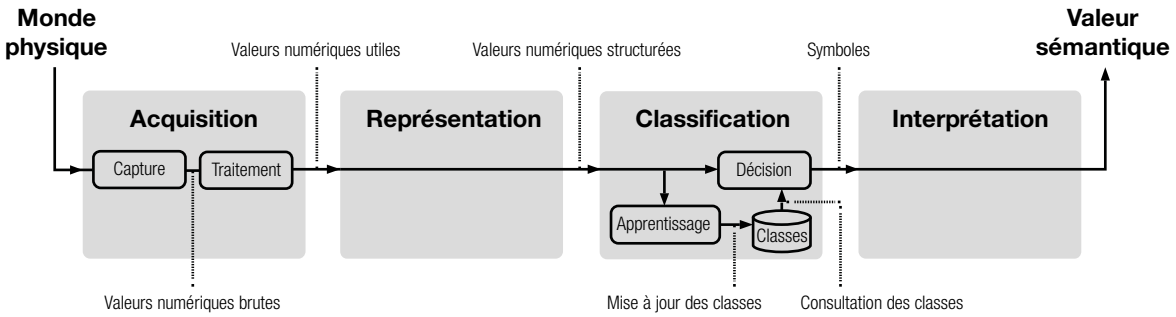


Figure 2-9 : Schéma général d'un système de reconnaissance de gestes (Braffort et al. 2006)

Nous proposons de détailler ce processus dans les paragraphes suivants, en décrivant dans un premier temps les principes d'acquisition du geste et le matériel existant. Nous verrons ensuite comment décrire le geste en vue de sa reconnaissance par un système informatique. Nous recensons quelques techniques d'apprentissage fréquemment utilisées pour la reconnaissance de gestes et enfin, la phase d'interprétation.

2.2.2.a *Acquisition du geste*

L'interaction par le geste est possible grâce aux systèmes de capture de mouvement du corps ou des gants de données. La modélisation des différentes parties du corps peut être très précise, en temps-réel, avec des données transmises à une fréquence élevée (> 60 Hz). Cependant, les systèmes de capture de mouvement restent complexes à mettre en place et à calibrer. Aussi, ils restent coûteux et nécessitent le plus souvent de grands espaces. Ils demeurent donc difficilement utilisables par le grand public. C'est notamment le cas pour des dispositifs optiques ou à centrales inertielles sans fil qui nécessitent de porter des marqueurs ou une combinaison.

Nous ciblons un interfaçage accessible qui soit à la fois peu intrusif et abordable financièrement. Ainsi, une option possible est de se tourner vers des dispositifs grand public principalement issus du jeu vidéo. Nous proposons une liste non exhaustive ainsi qu'un tableau récapitulatif des principaux périphériques (en annexe A).

Figure 2-10 : Capteurs grand public. (a) Caméra de profondeur *Microsoft Kinect*, (b) Bracelet à centrale inertielle *Movea MotionPod*, (c) Manette à capteur magnétique *Razer Hydra* et modification possible

Matériel à bas coût portés par l'utilisateur

Comme expliqué au chapitre 1 (1.1.3), les constructeurs de consoles de jeux vidéo intègrent depuis quelques années la modalité gestuelle en complément ou à la place des manettes et joysticks. De nombreux dispositifs commerciaux ont émergé, souvent vendus afin d'agrémenter l'expérience de jeu. On note l'arrivée de manettes de jeu vidéo capturant les mouvements avec de plus en plus de degrés de liberté et de précision. La taille du marché dans l'industrie du jeu vidéo permet de réduire les coûts de fabrication et de proposer des solutions alternatives.

Citons quelques succès commerciaux comme la manette *Wiimote* de *Nintendo* sortie en 2005. La manette permet d'obtenir quelques données d'accélération grâce aux accéléromètres. Une caméra embarquée dans la manette qui repère deux points lumineux infrarouges d'une barre placée sous l'écran de télévision permet d'obtenir des informations de positions et d'orientations absolues, mais partielles (Otaduy et al. 2009).

Parmi les constructeurs de dispositifs pour ordinateur, nous pouvons citer les manettes de jeux à capteurs magnétiques *Razer Hydra*. Elles donnent les positions et les orientations absolues à 6 DDL des deux manettes dans un volume correspondant à une sphère de rayon 1m autour de la base (Figure 2-10 (c)).

Sony commercialise depuis 2010 le dispositif *PlayStation Move*. Cette manette, au principe proche de celui d'une *WiiMote*, est surmontée d'une boule émettant une lumière colorée. La manette renvoie 6 DDL grâce à la combinaison d'une caméra repérant la boule de la manette (*PlaystationEye*) et d'une centrale inertielle.

Il existe des capteurs issus de la capture de mouvement professionnel comme les *MotionPod* de la société *Movea* (Figure 2-10 (b)). Ces capteurs au coût unitaire élevé se révèlent plus abordables lorsqu'ils sont achetés en grande quantité. À noter qu'on les retrouve par exemple dans des dispositifs grand public comme des télécommandes ou manettes de jeu.

Matériel à bas coût non intrusif basé vision

La reconnaissance de geste peut se faire simplement par webcam ou caméra standard ou caméra de profondeur. Dans ce cas un traitement d'image est nécessaire afin de repérer par exemple, la main ou la tête. (Rautaray et al. 2012) relatent plusieurs techniques combinant détection et suivi. La détection correspond au repérage de la main ou de la tête sur une image par segmentation de la zone de l'image où elle se trouve. Pour cela, il existe de nombreuses méthodes basées sur la couleur de la peau, la forme, le mouvement, des modèles anatomiques, etc. Ensuite, il est possible de procéder au suivi de la partie du corps (*tracking*). Selon la technique de suivi, cette opération demeure difficile pour une modélisation détaillée de la main notamment, car son apparence peut changer rapidement. Le tracking peut correspondre à l'interpolation entre les régions segmentées où se situe la main. Le suivi et sa robustesse sont importants, car ils permettent d'obtenir des trajectoires du mouvement. Ces trajectoires contiennent des informations essentielles et peuvent être utilisées sous forme de données brutes ou après une analyse supplémentaire. Aussi, les méthodes de suivi basées sur des modèles prédictifs permettent de maintenir le suivi même si des données viennent à manquer à certains moments (dues par exemple à des variations de lumière ou des occultations).

Microsoft commercialise la caméra de profondeur *Kinect* qui permet de reconstruire en temps-réel des chaînes cinématiques représentatives du corps d'un ou plusieurs utilisateurs. Une chaîne cinématique comprend le buste, la tête, le bras, les jambes, etc. de l'utilisateur en 3D (Figure 2-10 (a)). Les algorithmes permettent d'associer à cette silhouette un squelette 3D donnant des informations de position et d'orientation de 24 articulations du corps (Shotton et al. 2013; Suma et al. 2011). Concrètement, cela permet d'obtenir les positions ou les orientations de différentes parties du corps et des articulations. Cependant, ce dispositif ne donne pas d'informations suffisamment précises pour certaines articulations telles que les poignets, les chevilles ou la préhension des mains. Enfin, il arrive que la caméra de profondeur ne puisse pas repérer certains segments du corps. C'est le cas lors d'occultation par des éléments externes ou par les autres parties du corps de l'utilisateur.

Plus récemment le dispositif *LeapMotion* permet une capture des deux mains, donnant les positions et orientations respectives des métacarpes ainsi que la position du bout des doigts. Les articulations des doigts peuvent être estimées par chaînes cinématiques. Il est également possible d'avoir la position de l'extrémité d'objet dans le champ du capteur, comme un stylo par exemple.

Détournement des capteurs

Nous avons vu succinctement en introduction que la démocratisation des capteurs à bas coût donne lieu à de nombreux détournements. Certains amateurs fabriquent chez eux des « murs

immersifs » ou des systèmes à base de casque de réalité virtuelle (Kuntz et al. 2012). Ces capteurs font aussi l'objet de modifications. Il est possible de les démonter pour les miniaturiser. Par exemple, une manette *Razer Hydra* peut être transformée en un capteur 6DDL de casque de réalité virtuelle, (Ciger 2012) mais également pour repérer la main (Figure 2-10 (c)).

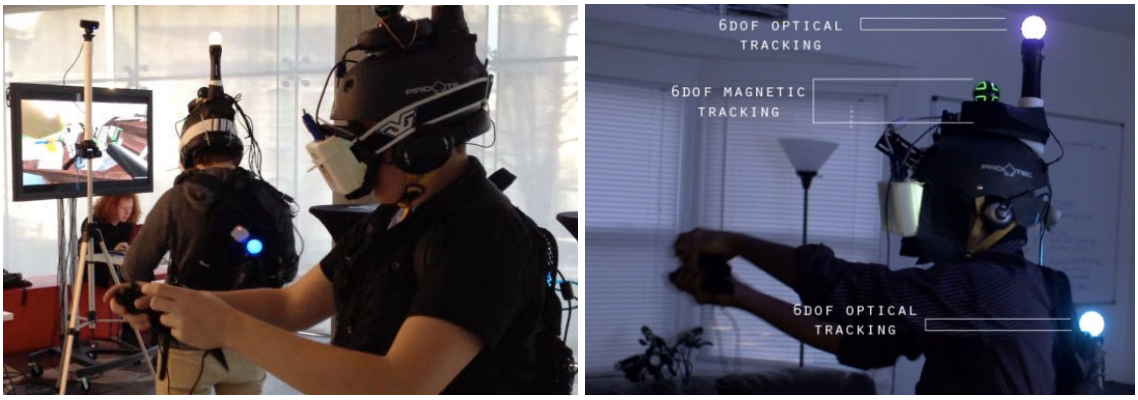


Figure 2-11 : Projet Holodeck, combinaison de capteurs *Playstation Move* et *Razer Hydra*.
Crédit photo : ("Project Holodeck" 2012)

Plusieurs solutions *open sources* permettent de récupérer les données des capteurs. Il est également possible de combiner les capteurs et de fusionner certaines données de façon logicielle. Par exemple, le projet *Holodeck* ("Project Holodeck" 2012) utilise notamment des manettes *Playstation Move* pour repérer la position des utilisateurs dans la pièce et des manettes *Razer Hydra* pour suivre les mains. La base étant installée sur l'utilisateur, les positions des mains peuvent être calculées par fusion de données dans le référentiel de la pièce (Figure 2-11).

→ En résumé, nous proposons un tableau récapitulatif des principaux capteurs à bas coût en Annexe A : Tableau collaboratif descriptif des capteurs grand public selon l'usage. La table ci-dessous résume les avantages et limites des capteurs grand public. Une autre limite différente de la partie matérielle réside dans la considération des descripteurs gestuels. Cela correspond à la façon de décrire le geste en fonction du besoin final. Nous détaillerons cela dans les parties suivantes.

Table 2-3: Avantages et Limites des capteurs à bas coût

✓ Accessible financièrement	× Précision
✓ Robustesse	× Latence
✓ Facilement remplaçable	× Fréquence parfois faible
✓ Données relatives et absolues	× Données parfois seulement relatives
✓ Possibilité de combinaison	× Limitation de l'espace d'utilisation (fil / vision)
✓ Clé en main pour les plus démocratisés	× Modifications matérielles parfois nécessaires
✓ Détournement par la communauté *open source*	× Données partielles (DDL limités) ou occultations (basé vision)

2.2.2.b *Représentation des données : traitement et descripteurs gestuels*

Il existe de nombreuses manières de décrire le geste pour son intégration dans un système informatique. (Papin 2006) distingue les descripteurs issus de « l'anthropométrie dynamique » comme la force, l'amplitude, le sens, la vitesse, la durée ou le nombre de répétitions, mais il en existe bien d'autres, plus qualitatifs. (Kurtenbach et al. 1990) cités précédemment considèrent que faire un salut de la main est un geste alors que presser la touche d'un clavier ne l'est pas. Ils notent en effet que le mouvement du doigt vers la touche n'est pas considéré comme un geste. (Billinghurst et al. 2011) ajoutent que le geste aurait pu être effectué avec « amour ou colère ». Cela montre que le geste est quantifiable et qualifiable.

Par exemple, (Fdili Alaoui et al. 2012) proposent de caractériser la qualité de mouvement en s'inspirant du Laban. Les descripteurs proposés basés sur des modèles physiques représentent qualitativement le mouvement. Dans le domaine des arts, quantifier ou qualifier le mouvement dépend de l'expressivité ou de l'esthétique désirée. Dans le cas de la mise en place d'un système destiné à être utilisé par une personne à mobilité réduite, celle-ci peut avoir un geste limité en amplitude en degrés de liberté. Cela ne signifie pas pour autant qu'elle ne peut pas effectuer un geste précis ou expert. Le choix du descripteur est donc inhérent aux caractéristiques sensorimotrices du geste. La description du geste dépend par conséquent du handicap, de sa singularité, mais également de l'activité désirée. La caractérisation quantitative ou qualitative du mouvement reste un problème ouvert, car dépendant du cas d'usage et de l'activité.

D'après (Braffort et al. 2006), le choix des descripteurs gestuels, aussi appelés primitives de représentation du geste, est important, car il influence directement la qualité et la robustesse du système de reconnaissance de gestes. Ces primitives possèdent plusieurs propriétés, comme la continuité, la réversibilité, la sensibilité, l'indépendance et l'homogénéité temporelles. Les auteurs proposent de regrouper ces primitives en quatre catégories :

- Les primitives directes qui sont principalement les données brutes de capteur normalisées ;
- Les primitives géométriques comme la distance parcourue, une boite englobante, la courbure, etc. ;
- Les primitives visuelles comme les pixels pour un capteur vision ;
- Les primitives cinématiques comme la vitesse, l'accélération.

Les primitives choisies sont ensuite transmises au système d'apprentissage artificiel sous forme de vecteurs de primitives.

→ En résumé, l'acquisition du geste peut se faire par différents capteurs. Il existe aujourd'hui un certain nombre de produits grand public donnant des données pertinentes sur le mouvement tout en restant accessibles financièrement. Cependant, comme décrit dans la partie 2.2.1.a sur la nature du geste, le choix des descripteurs relève de l'activité et des besoins finaux. La caractérisation peut être quantitative, qualitative ou une combinaison des deux. De plus, ces descripteurs ont une influence sur la qualité de la reconnaissance. Dans la partie suivante, nous détaillons les principales méthodes de reconnaissance du geste.

2.2.2.c *Classification et reconnaissance du geste*

La phase de classification ou de reconnaissance permet de transcrire les vecteurs de primitives en symboles. Ces symboles sont associés à des classes de gestes prédéfinis qui sont le vocabulaire de gestes de l'application.

Il existe plusieurs méthodes pour modéliser les données sous forme de classes. La constitution des classes se fait généralement par un apprentissage. Celui-ci consiste à donner au système de reconnaissance un ensemble de vecteurs de primitives qui sera associé à chaque classe. Fournir au système plusieurs vecteurs de primitives d'un même geste permet d'intégrer les variations éventuelles, évitant d'avoir une classe trop restrictive. Il est possible d'associer les vecteurs de primitives aux classes par un apprentissage supervisé en deux étapes :

- Dans une première phase d'apprentissage, c'est par exemple l'utilisateur qui fournit l'ensemble des vecteurs de primitives au système en fonction des classes et symboles désirés en sortie.
- Dans la seconde phase de décision ou de reconnaissance, le système prédit la classe à partir d'une nouvelle donnée du vecteur de primitives, par comparaison à la classe élaborée pendant la phase d'apprentissage.

Il existe plusieurs méthodes d'apprentissage supervisé. Dans le domaine de la performance musicale par le geste, (Bevilacqua et al. 2010; Françoise et al. 2013) détaillent plusieurs méthodes. Ils proposent notamment un apprentissage par démonstration en deux temps. Dans la première phase dite d'apprentissage, l'utilisateur écoute un morceau ou un son. Ensuite, simultanément à la réécoute, il associe un geste au son entendu. Ce geste sera enregistré par le système et associé — c'est-à-dire couplé — au son. Cette opération peut être répétée pour plusieurs gestes et sons. Durant la deuxième phase dite de performance, l'utilisateur reproduit les gestes et le système traduit les modulations des gestes en modulations sonores.

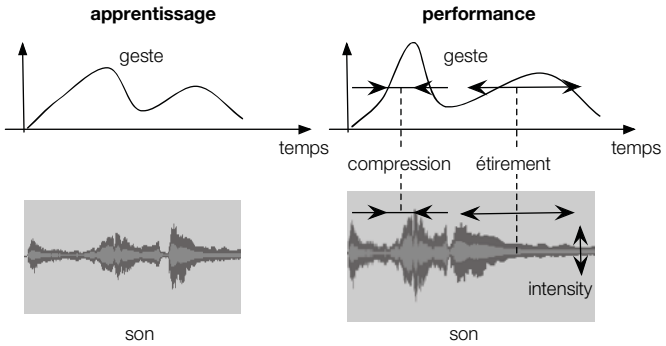


Figure 2-12 : Association geste et son à l'apprentissage et à la performance (Bevilacqua et al. 2010)

Il existe d'autres méthodes d'apprentissage artificiel que l'apprentissage supervisé. (Françoise et al. 2013) distinguent l'apprentissage par renforcement qui consiste à ne pas disposer d'exemples apriori, mais d'acquérir l'exemple durant l'interaction. Si l'utilisateur intervient dans l'apprentissage en fournissant des exemples a priori et pendant l'interaction, ils parlent d'apprentissage actif. Enfin, si les exemples sont fournis pendant l'interaction et sans exemple

apriori, ils parlent de façonnage interactif. Dans le cas d'apprentissage interactif du couplage geste-son, les auteurs soulèvent trois enjeux :

- Le premier est d'apprendre à partir de très peu d'exemples ;
- Le second est d'apprendre à partir de données acquises pendant l'interaction ;
- Enfin, le troisième enjeu est d'adapter de manière interactive l'apprentissage à l'utilisateur.

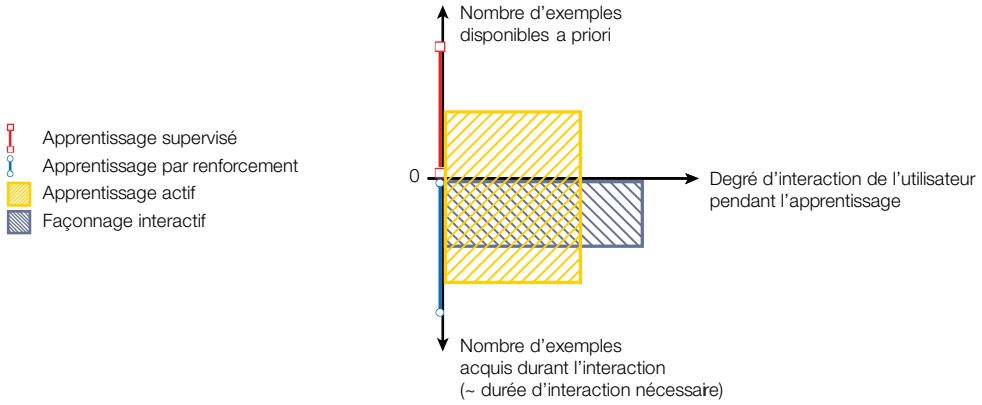


Figure 2-13 : Relations du façonnage interactif à d'autres champs de recherche (Françoise et al. 2013)

Il existe d'autres techniques d'apprentissage artificiel. Dans le cas d'apprentissage supervisé probabiliste, plutôt que de donner une classe, c'est une probabilité d'appartenance qui est attribuée à chacune des classes prédéterminées. Dans un apprentissage non supervisé, le système effectue l'apprentissage à partir de la masse de données reçues et extrait lui-même les différentes classes selon des critères paramétrés à l'avance. L'apprentissage semi-supervisé combine à la fois l'extraction de classes de façon supervisée et automatique. Le but est d'améliorer la qualité de l'apprentissage. Il existe enfin l'apprentissage par renforcement qui permet de donner plus de poids à certains échantillons, de façon manuelle ou automatique.

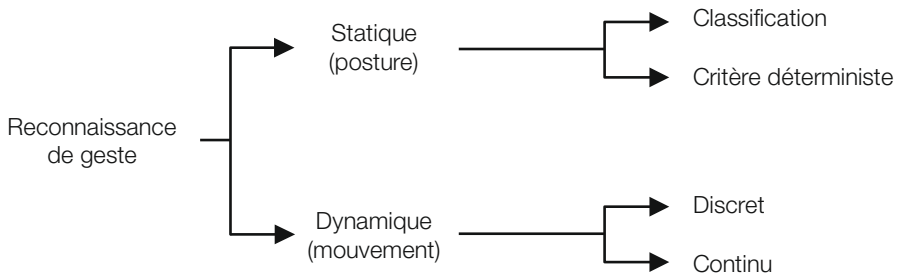


Figure 2-14 : Techniques possibles de reconnaissance de gestes

Différents algorithmes permettent d'effectuer l'apprentissage ou la classification. L'état de l'art de (Rautaray et al. 2012) propose une classification exhaustive de techniques de reconnaissance de la main par vision. Ces techniques sont d'ailleurs connexes à celles utilisées notamment pour la reconnaissance de l'écriture de la parole ou de l'écriture (Rabiner 1989; Wobbrock et al. 2007). Elles peuvent naturellement s'étendre à la détection et la reconnaissance de gestes produits par différentes parties motrices du corps. En se basant sur la

classification de (Rautaray et al. 2012), nous présentons quelques principales méthodes de reconnaissance de geste (Figure 2-14). Nous les détaillons dans les prochains paragraphes en distinguant en premier lieu reconnaissance de geste statique et dynamique.

(a) Classification de gestes statiques ou postures

Pour la détection des gestes statiques ou postures, nous retrouvons essentiellement dans la littérature les classifieurs linéaires, non-linéaires et la comparaison de prototype (*template matching*). Ils permettent d'associer les classes de gestes aux propriétés similaires. (Rautaray et al. 2012) proposent une description détaillée ainsi qu'une liste des avantages et des limites de ces classifieurs. Parmi les principaux, citons les algorithmes de partitionnement des *k-moyennes* (*k-means*), la méthode des « k plus proches voisins » (« *k-nearest neighbors* ») ou le *mean shift clustering*. Il existe les *Séparateurs à Vaste Marge* (aussi appelés *SVM*, *machines à vecteurs de support* ou *Support Vector Machine*). Il existe aussi des techniques de *boosting* qui permettent d'optimiser les performances des classifieurs.

Il est également possible de reconnaitre des gestes par des modèles déterministes. Ils peuvent notamment faire l'objet d'un simple seuillage (*threshold*). Par exemple, (Manitsaris 2010) propose de reconnaitre les doigtés d'un pianiste en temps-réel par caméra. La reconnaissance du doigté se met en œuvre en déterminant le seuil d'appui effectué par chaque doigt sur les touches du clavier. Dans le domaine de la réalité virtuelle, Cabral et al. (Cabral et al. 2005) détecte un ensemble de geste pour le contrôle d'application, pour la sélection et pour la manipulation dans un dispositif de réalité virtuelle. Les postures de l'utilisateur ont été prédéfinies par seuillage de positions des mains relatives à celle de la tête de l'utilisateur.

(b) Reconnaissance de gestes dynamiques par apprentissage

Dans le cas d'une reconnaissance de gestes dynamiques, les classes représentent des gestes dynamiques continus, le plus souvent segmentés au niveau du *stroke*. Une fois que le modèle a acquis les classes (par exemple trois classes dans l'illustration (a) de la Figure 2-15), le geste est reproduit, arrive en entrée du système de reconnaissance (b). Celui-ci donne en sortie par comparaison les différentes classes du geste (c). Nous allons voir qu'il existe principalement des algorithmes de reconnaissance discrète de gestes dynamiques. Mais il existe également des algorithmes de reconnaissance de gestes en continu.

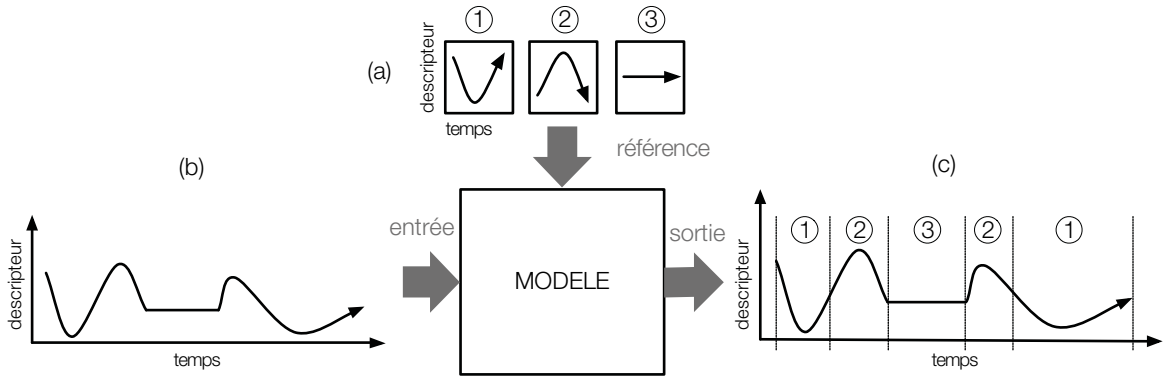


Figure 2-15 : Association de classe de gestes a un geste continu, extrait de (Françoise 2011). (a) apprentissage de trois gestes de référence. (b) reproduction d'un geste en continu en entrée du modèle. (c) extraction des classes de gestes par le système de reconnaissance.

Algorithmes de reconnaissance discrète de gestes dynamiques

Pour la reconnaissance de gestes dynamiques, il existe de nombreuses techniques. Les techniques de classification citées précédemment peuvent également permettre la reconnaissance de gestes dynamiques. Par exemple (LaViola et al. 2011) ont comparé l'implémentation des classifieurs linéaires et la technique d'A*daBoost* pour la reconnaissance de gestes avec une W*iiimote.* Pour un vocabulaire de 25 gestes, des taux de reconnaissance obtenus sont supérieurs en moyenne à 90% pour toutes les techniques utilisées.

Il existe d'autres approches. En se référant à l'état de l'art de (Rautaray et al. 2012), nous distinguons les quatre plus courantes : les approches connexionnistes, les automates à états finis, la déformation temporelle dynamique et les approches basées sur les Modèles de Markov Cachés.

- Les approches connexionnistes :

Il est possible de classifier des gestes par des réseaux de neurones aussi appelés réseaux connexionnistes. Ils sont constitués de neurones c'est-à-dire de cellules regroupées en couches. Chaque cellule d'une couche est reliée à toutes les cellules précédentes. Dans le cas d'un réseau de neurones simple (perceptron multicouche), celui-ci est entrainé par un algorithme de propagation arrière qui détermine les valeurs des liens entre les neurones (Martin 2000). Il y a deux principales limites. La première est le temps de calcul nécessaire à la convergence du réseau pour l'entrainement. Le second est l'aspect « boite noire » du système qui fait que l'entrainement ne permette pas toujours une convergence optimale. (Rautaray et al. 2012) note que les réseaux de neurones de type TDNN (Time Delay Neural Network) sont principalement utilisés. Ils ont l'avantage de permettre un apprentissage rapide, mais manquent cependant de robustesse pour la reconnaissance dans des cas d'application utilisant des gestes de la main.

- Les automates à états finis :

Il existe des méthodes comme l'approche par *FSM* (automate à états finis). Le geste est segmenté en séquences de plusieurs configurations de la main par exemple. Chaque état correspond à une partie du mouvement ou une configuration de la main. (Martin 2000) note que les états intermédiaires correspondent aux configurations intermédiaires et comprennent par conséquent les coarticulations. La transition entre états ne s'effectue que lorsque la configuration est connue. Si ce n'est pas le cas, elle se fait vers un état intermédiaire. L'arrivée à l'état final signifie que le geste est reconnu. Les *FSM* ont l'avantage d'être rapides à implémenter et peu couteux en temps de calcul. Mais là aussi, il est difficile de les rendre robustes. En revanche, pour un nombre élevé de gestes, le système se complexifie, rendant la construction de l'automate plus difficile.

- La déformation temporelle dynamique :

L'algorithme de déformation temporelle dynamique (de l'anglais *Dynamic Time Warping*, noté *DTW*) permet notamment d'apparier de façon optimale deux séries temporelles qui varient au cours du temps. Par exemple, il est possible de mettre en correspondance deux trajectoires en considérant le décalage possible entre les points à mettre en correspondance. Cet alignement autorise des étirements/réductions locales des trajectoires. Le *DTW*

principalement utilisé en reconnaissance de la parole ou comparaison de formes écrites s'emploie également en reconnaissance de geste (Martin 2000).

- Les Modèles de Markov Cachés :

Il existe les Modèles de Markov Cachés (ou *HMM* de l'acronyme anglais *Hidden Markov Models*). Ces derniers se sont généralisés depuis les années 1990 pour la reconnaissance de mots et plus récemment de l'écriture et l'interaction gestuelle. Le principe est une extension des chaînes de Markov. Celles-ci sont comparables aux *FSM*, mais dans ce cas, c'est une probabilité qui est assignée à chaque transition d'état. Les chaînes de Markov imposent cependant d'avoir seulement une transition pour chaque sortie. Les *HMM* sont considérés comme une généralisation des chaînes de Markov sans cette restriction (Rautaray et al. 2012). Le but est d'estimer la similarité entre le geste effectué et les gestes modélisés. On reconnait le geste effectué par obtention du « maximum de vraisemblance ».

Il existe de nombreuses mises en œuvre. (Starner et al. 1998) proposent un système de reconnaissance de geste de la langue des signes basé sur des *HMM* par caméra. Ils obtiennent des taux de reconnaissance élevés, supérieurs en moyenne à 90 %. (Morrison et al. 2002) ont implémenté un système pour permettre à des personnes à mobilité réduite d'interagir par des gestes, même de faible amplitude. (Schlömer et al. 2008) ont développés un algorithme *open source WiiGee* permettant l'apprentissage et la reconnaissance de gestes du bras à partir des données d'accélération d'une manette *Wiimote.*

À noter cependant que les *HMM* restent complexes à implémenter du fait du grand nombre de paramètres à définir. Cependant, un temps de calcul et un nombre d'échantillons élevé demeurent nécessaires à l'apprentissage. Enfin, la plupart de ces systèmes permettent de reconnaitre les gestes a posteriori. Aussi, un minimum de mouvements transmis permet la reconnaissance, rendant une reconnaissance en continu difficile.

Algorithmes de reconnaissance en continu de gestes dynamiques

Il existe des possibilités de reconnaissance de geste en continu. Dans ce cas, la mesure de vraisemblance du geste n'est plus une probabilité, mais une densité de probabilités. Quelques implémentations existent (Alani et al. 1994; Manitsaris 2010) qui associent *HMM* et Modèles Gaussiennes Mixtes (*GMM* de l'acronyme anglais *Gaussian Mixture Models*). Les *GMM* servent à modéliser le geste sous forme d'un mélange de distributions gaussiennes.

(Bevilacqua et al. 2010) proposent un algorithme basé *HMM* qui donne en temps-réel et en continu des mesures de similarité entre le geste reproduit et geste original (*following*). La reconnaissance se faisant en continu, elle donne une vraisemblance pour les distributions. Peu d'échantillons de données suffisent, ce qui permet une reconnaissance dite précoce. Un exemple d'application est la reconnaissance de gestes pour la synthèse sonore. Le système permet d'obtenir des mesures de similarité de geste en cours d'exécution parmi un vocabulaire de gestes enregistrés, et cela en temps-réel et en continu. De plus, l'algorithme donne certaines caractéristiques du geste reconnu par rapport à son modèle, tels que la vitesse du geste en cours ou l'avancement (pourcentage de complétion). L'intérêt est de pouvoir reconnaitre des gestes avant leur complétion, ce qui rend éventuellement possible une anticipation de l'action de l'utilisateur. Enfin, à noter que l'apprentissage de chaque geste peut se faire avec un seul échantillon, aussi appelé *one-shot learning* (Guyon et al. 2012).

→ Les modèles *HMM* ont montré leurs avantages dans la reconnaissance de gestes. En revanche, ils ne distinguent principalement qu'une probabilité d'appartenance du geste à une classe. L'approche de (Bevilacqua et al. 2010) donne une densité de probabilité en temps-réel et en continu. Il est possible d'obtenir certaines caractéristiques supplémentaires comme le pourcentage de complétion ou la vitesse d'exécution par rapport au geste original. Concernant la phase d'apprentissage par l'utilisateur, il existe plusieurs méthodes, la plus classique étant l'apprentissage supervisé. Il reste à définir comment coupler la reconnaissance du geste à l'activité en virtuel.

2.2.2.d *Interprétation du geste et rétroaction*

À l'issu du processus de classification ou reconnaissance, les éléments symboliques « en sortie » sont interprétés en fonction du contexte. Il peut s'agir d'une probabilité d'appartenance à une classe de geste ou d'une similarité. De plus, certaines caractéristiques du geste peuvent être transmises en cours de son suivi comme la vitesse ou le pourcentage d'avancement par rapport au geste original. Chaque geste reconnu peut être associé à des commandes afin de déclencher des évènements ou des contrôles plus fins pour réaliser des PCV décrites précédemment qui, rappelons-le, sont :

- L'observation ;
- Le déplacement ou la navigation ;
- L'action sur le monde virtuel ;
- La communication avec autrui ou avec l'application.

La notion de contexte est importante pour plusieurs raisons. Pour l'utilisateur cela permet de réutiliser le vocabulaire gestuel adéquat dans la situation appropriée. Côté système, le contexte peut être un moyen de renforcer l'utilisation des données "en sortie" pour l'interaction. En d'autres termes, cela permet d'optimiser la reconnaissance du geste au regard de l'activité. Par exemple, (X. Cao et al. 2005) proposent un moteur de décision qui prend en compte le contexte d'application (comme la zone de l'environnement virtuel ou une tâche précise serait à effectuer) et l'historique des commandes ou leur prédiction.

Parmi les nombreux exemples d'interaction gestuelle par reconnaissance, le prototype de (Sowa et al. 2002) permet notamment de désigner un objet de la scène par des gestes iconiques. L'objet est associé à un geste qui symbolise les parties saillantes de la forme de l'objet. Dans le cas d'une activité, l'interaction gestuelle comporte donc un retour qui permet au sujet de comprendre son avancement dans l'activité. Il demeure important de lui renvoyer des informations relatives à la complétion de la tâche, par des rétroactions aussi appelées *feedbacks*.

Dans le cas d'une interaction gestuelle sonore, le geste peut accompagner ou diriger le son (*following*) (Bevilacqua et al. 2010). Dans ce cas, le système suivant le geste est capable de s'adapter aux variations temporelles

→ Dans le cas d'une reconnaissance, la notion de couplage geste et action est capitale. Elle intègre la notion sous-jacente de rétroaction. Par conséquent le couplage est inhérent à la phase d'apprentissage. Nous proposons de détailler les notions de

rétroaction et d'explorer les possibilités d'apprentissage supervisé en vue d'une reconnaissance continue.

2.2.3. Considérations sur l'usage de l'interaction gestuelle

2.2.3.a *Interaction et rétroaction*

Dans la phase d'interprétation du geste, il est important de donner une information en retour à l'utilisateur, appelée rétroaction ou *feedback*. L'implémentation du geste en réalité virtuelle a souvent recours à un *feedback* visuel plutôt qu'haptique. Pour comprendre les processus mis en jeu, nous proposons détailler les notions de rétroaction et interaction. Selon (Cadoz 2006), rétroaction et interaction sont des notions distinctes bien qu'elles peuvent parfois se recouper. L'interaction correspond à une influence réciproque de deux entités, en l'occurrence ici l'homme et le système en réalité virtuelle, sans pour autant considérer que l'un est agissant, l'autre subissant. L'interaction se rapproche des principes systémiques du monde (comme les écosystèmes, naturels ou artificiels) et est à considérer comme des systèmes imbriqués avec relation et échange comme la matière, l'énergie, ou l'information, etc. L'auteur définit la rétroaction comme « un système orienté avec une entrée et une sortie, puis une réinjection, additionnée à l'entrée avec un certain coefficient, de ce qui est produit en sortie ». La rétroaction correspond aux informations en retour sur les résultats d'une performance.

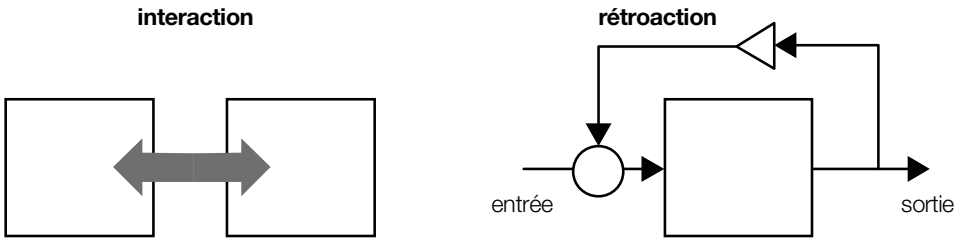


Figure 2-16 : Interaction (gauche) et rétroaction (droite), extrait de (Cadoz 2006)

Feedback intrinsèque et extrinsèque chez l'homme

Selon la théorie du contrôle moteur (Adams 1971; Rigal 2003), l'homme dispose de plusieurs informations pendant une action. Celles-ci sont liées à la proprioception, aux perceptions mnémoniques :

- Il y a avant le mouvement la préaction, c'est-à-dire une copie efférente du mouvement qui « préparerait les centres sensitifs aux conséquences de l'action » ;
- L'information concomitante au mouvement qui compare passé et présent. C'est une forme de connaissance de l'exécution : « on sent que le mouvement est correct » ;
- Il y a des éléments postérieurs au mouvement :
  - Le *feedback* intrinsèque correspond aux différentes afférences d'origine perceptive. Chez l'homme, il est interne au sujet et correspond aux informations d'origine visuelle, auditive, proprioceptive, etc. et cette boucle se révèle la plus rapide en traitement de régulation de l'action. Elle correspond au déroulement de l'action et de son résultat

- Le *feedback* extrinsèque correspond à une information extérieure supplémentaire. En théorie du contrôle moteur, Adams définit deux sous-*feedbacks* extrinsèque : la connaissance du résultat et la connaissance de la performance, mais ces boucles de rétroaction se révèlent plus longues, et donc plus lentes.

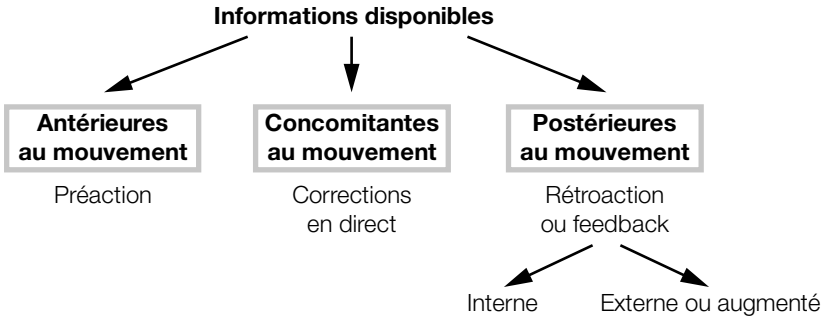


Figure 2-17 : Informations disponibles lors d'un mouvement (interaction) et *feedback* (Rigal 2003)

Dans la relation geste et musique, (Cadoz 2006) considère l'existence d'une première boucle d'interaction avec couplage geste-instrument-son. Une seconde boucle est en jeu et correspond à la relation du geste à l'œil. Celle-ci s'ajoute à la manipulation physique de l'instrument et au son entendu. Il estime que ces « différentes boucles, de l'action gestuelle respectivement aux perceptions tactiles, auditives et visuelles ne sont pas indépendantes : ce qui se produit selon l'une des trois boucles peut influencer le résultat cognitif correspondant aux autres ». Ceci complexifie davantage la notion de *feedback*.

Cas particulier de la rétroaction visuelle sans retour haptique

En réalité virtuelle, il arrive qu'on ne puisse pas simuler complètement une rétroaction comme le retour haptique pour des raisons physiques ou techniques. Il existe un cas particulier de la relation geste-image sans retour haptique, où différentes modalités perceptives entrent en jeu ou se substituent. (Mine et al. 1997) proposent de mieux considérer la proprioception et la kinesthésie pour compenser le fait de manipuler des objets principalement par retour visuel. Pour rappel, la proprioception correspond à la perception des parties du corps (leurs positions, orientations…) et la kinesthésie renvoie à la sensation de mouvement provoquée par leurs déplacements. Ils proposent de privilégier la manipulation directe. Celle-ci fait appel à la colocalisation d'un objet notamment dans la main de l'utilisateur. La proprioception permet à l'utilisateur de mieux estimer la position de l'objet virtuel rendant l'interaction plus précise. Dans le contrôle d'une tâche de manipulation, (Hand 1997) remarque que la perception haptique, qui combine tactile et kinesthésie, jouerait un rôle plus important que la vision. Faute d'un système haptique adéquat, il propose de faire appel aux rétroactions multi sensorielles (traduit de « *feedbacks* cross modaux »). (Fuchs, Berthoz, et al. 2006) utilisent plutôt le terme de métaphores avec substitution sensorimotrice. Par exemple, il est possible de signaler une collision par un retour visuel ou auditif. Ceci se révèle généralement signifiant pour l'utilisateur à condition que la métaphore soit suffisamment explicite.

2.2.3.b *Ergonomie du geste en interaction*

L'utilisation du geste en interaction n'est pas sans préconisation, notamment en termes de fatigue à l'usage, et de troubles éventuels liés à la répétition du geste.

Syndrome « bras de gorille »

En ergonomie du geste, il existe le syndrome « bras de gorille » (traduction de *gorilla arm syndrom*) qui se retrouve de plus en plus en interaction gestuelle. En effet, ce syndrome apparait lorsqu'un membre est maintenu en position contre la gravité, et spécialement au-dessus du cœur. Cela se traduit par une fatigue musculaire et l'impression d'avoir les bras qui gonflent d'où l'analogie au gorille. Ce syndrome a plusieurs facteurs : il dépend de l'amplitude, de la durée de maintien de la contraction musculaire sans interruption. En effet, on observe une réduction de l'apport de sang dans les muscles devant toujours fournir un effort. Ce syndrome est caractéristique des interfaces gestuelles nécessitant des mouvements amples, de longue durée et positionnant le bras au-dessus du cœur. (Baudel et al. 1993) notent que la communication gestuelle sollicite davantage de muscles qu'une interaction avec appui comme la souris. Par conséquent une interaction gestuelle doit être concise et rapide afin de minimiser l'effort. En particulier, l'implémentation de contrôle gestuel doit éviter des gestes requérant une grande précision pendant une longue période. Dans l'application de contrôle par posture en réalité virtuelle de (Cabral et al. 2005), les auteurs notent que la grande amplitude des mouvements des bras entraine effectivement une certaine fatigue. Elle a d'ailleurs été ressentie par 40 % des participants de leur étude.

Répétition du geste et troubles musculo-squelettiques

La répétition du geste, notamment d'un même geste sur une période prolongée peut amener à certains troubles appelés troubles musculo-squelettiques. Ces troubles se retrouvent essentiellement dans le travail à répétition sollicitant les mêmes parties du corps. Les principaux maux liés à la fatigue et l'usage intensif des gestes sont les lésions et troubles musculo-squelettiques. Ils correspondent aux troubles mécaniques du corps liés à des mouvements inappropriés ou répétitifs (Simoneau et al. 1996). On distingue ici la fatigue de répétition plutôt que de garder le bras en l'air. Outre la fatigue musculaire localisée et persistante, plusieurs symptômes sont recensés comme les courbatures, des picotements ou engourdissements lors d'atteintes nerveuses et la difficulté à réaliser des efforts.

La répétition d'un même geste d'interaction entrainera fatigue ou lassitude de l'utilisateur qui tendra probablement à simplifier son geste ou à l'écourter. Cela peut par conséquent modifier le geste dans sa topologie et sa temporalité. Afin de mettre en place d'un système de reconnaissance de geste, il semble nécessaire de caractériser ces variations. Nous proposons plusieurs méthodes dans le paragraphe suivant.

2.2.3.c *Variabilité et comparaison de gestes*

En plus des considérations ergonomiques, il semble approprié d'évaluer la modification du geste à l'usage, aussi bien à des fins d'observation que d'aide à la mise en place d'un système d'apprentissage artificiel. Il est possible de pouvoir mesurer les écarts entre les occurrences d'un même geste par différents algorithmes comparaisons de gestes « hors ligne ». Ces

algorithmes comparent les gestes selon plusieurs critères, il convient de choisir les bons descripteurs au regard de ces critères.

Variabilité de forme du geste

Pour comparer la répétition de geste d'écriture de formes sur surfaces tactile, (Stößel et al. 2010) proposent notamment de comparer la précision du geste reproduit selon des paramètres de stabilité de la forme et la stabilité de direction du geste. Selon un geste transcrit en lignes ou circulaires, les auteurs comparent les variations angulaires ou l'excentricité d'une ellipse se superposant au mieux aux données. Les données sont ensuite comparées de façon statistique par une analyse de variance (*ANOVA*). Il existe d'autres méthodes, comme la comparaison de proportions de formes dans la connaissance de gestes d'écritures de (Kristensson et al. 2004). Il est possible d'étendre cette méthode à la comparaison de geste dans un espace 3D.

Variabilité des trajectoires et rotations

Afin de comparer des mouvements en trois dimensions, (Moreau et al. 2008) proposent de comparer les distances euclidiennes en position et orientations par la technique des « quaternions pures ». Celle-ci consiste à transformer les orientations d'un geste en une trajectoire 3D dimensions sur la peau d'une sphère (Figure 2-18 gauche). Dans une approche similaire, (Cifuentes et al. 2013) comparent les orientations selon leurs longueurs d'arc par une technique que les auteurs nomment *Dynamic Arc-Length Warping algorithm*. Au lieu de faire une comparaison de distance euclidienne de trajectoires, il est possible de procéder à une comparaison par la distance de Mahalanobis, qui pallie certains cas de mauvais appariements en euclidien (Figure 2-18 droite). Cette méthode est habituellement utilisée en « reconnaissance de formes », mais (Coster et al. 2010) l'ont implémentée pour la comparaison de gestes dessinés sur tablette.

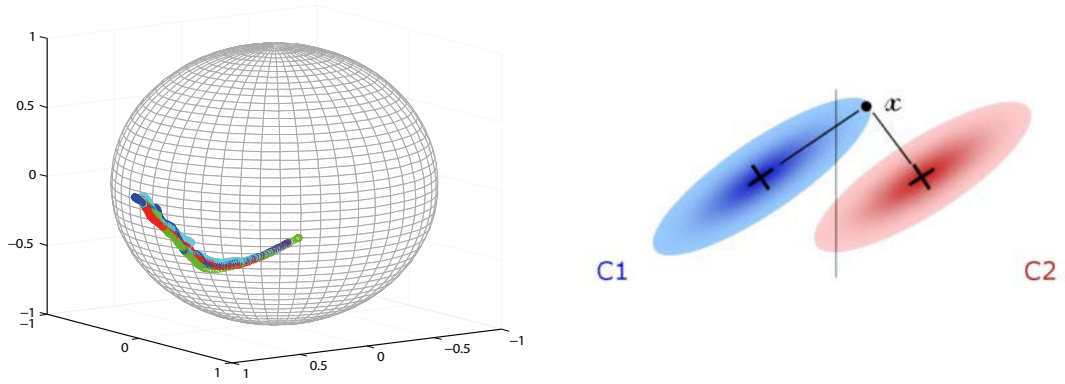


Figure 2-18 : À gauche, représentation de trajectoires angulaires de différents gestes experts dans l'espace des quaternions (Moreau et al. 2008) ; à droite, distance de Mahalanobis (Coster et al. 2010). En distance euclidienne, x appartient à la classe C1alors qu'avec la distance de Mahalanobis, x appartient à la classe C2

Variabilité de vitesse du geste

Pour des comparaisons de mouvements de la main, (Gold et al. 2008) emploient une méthode similaire. Ils comparent les profils de vitesse de flexion de doigts dans le cas d'imitation de gestes de la main par une analyse de variance.

Variabilité de la durée

Une méthode de comparaison possible est de mesurer des écarts temporels. Ces écarts peuvent s'observer par simple comparions de chronométrage du geste. Cela peut s'avérer un indicateur de performance, notamment en chirurgie (C. G. L. Cao et al. 1996).

2.2.4. Synthèse

Nous avons vu qu'il est possible de procéder à une reconnaissance des gestes pour interagir en temps-réel dans le cadre d'une interaction centrée-utilisateur. L'acquisition du geste au moyen de dispositifs grand public et à bas coût est possible. Il est d'ailleurs envisageable de les combiner et de les modifier. Le geste peut être caractérisé de façon quantitative ou qualitative en fonction de l'activité en environnement virtuel.

En ce qui concerne l'apprentissage, différentes techniques existent. Là aussi leur choix relève de l'application finale et de l'activité. Nous retenons cependant que les modèles basés sur les *HMM* ont certains avantages en reconnaissance discrète et continue du geste. L'apprentissage des gestes peut se faire de façon supervisée ou interactive, et dépend aussi de la rétroaction ou du couplage mis en jeu.

À noter qu'il y a certaines considérations dans l'usage du geste en environnement virtuel aussi bien cognitives qu'ergonomiques. En effet, la rétroaction est importante dans l'exécution du geste et celle-ci est à la fois dépendante du sujet, de son ressenti, de la fatigue, et de ce que renvoie l'environnement. Le geste peut par conséquent varier dans l'interaction selon plusieurs facteurs, physiques ou selon les influences de la rétroaction. Il existe différents moyens d'observer et de comparer les variations des gestes.

Nous constatons qu'une interaction gestuelle définie par l'utilisateur prend en compte beaucoup de paramètres et suscite des interrogations. Le chapitre suivant détaille le concept et les problématiques soulevées.

CHAPITRE 3
CONCEPTS SUR L'APPROCHE OÙ L'UTILISATEUR DÉFINIT L'INTERACTION GESTUELLE

Nous avons choisi d'orienter l'adaptation vers l'approche qui consiste à laisser l'utilisateur choisir et définir ses gestes pour interagir. Ses besoins et ses spécificités sont directement pris en compte. Nous présentons dans ce chapitre notre concept où l'Utilisateur Définit l'Interaction gestuelle (UDIg) qui semble répondre au mieux aux besoins des utilisateurs. La mise en place de cette approche amène à un certain nombre de problématiques. Nous proposons de les détailler selon quatre critères : la répétabilité, la mémorisation, la rétroaction et les considérations techniques.

3.1. L'UTILISATEUR DÉFINIT L'INTERACTION GESTUELLE

Les cas d'application de cette thèse demandent une adaptation. Notre idée forte est de permettre une activité en environnement virtuel à partir d'une gestuelle non imposée. Pour cela nous proposons une approche où l'homme choisi lui-même ses gestes pour interagir. La gestuelle n'est plus imposée par le système, mais initiée par l'utilisateur et dirigée vers l'action. Cette approche que nous nommons « Utilisateur Définit l'Interaction gestuelle » s'inspire de l'approche « User-Derived Interaction » de (Good et al. 1984). Celle-ci s'adresse à tout utilisateur, car elle permet à chacun d'enregistrer ses propres gestes dans le système de façon totalement libre au regard de l'activité. Les usages, les besoins et les spécificités sensorimotrices sont ainsi directement pris en compte.

La démarche de conception inclusive que nous avons vue (2.1.2.c) détermine, à partir des gestes d'un panel d'utilisateurs, celui qui semble le plus adapté à l'action à réaliser. Elle reste donc une démarche de consensus ou de standardisation contrairement à notre démarche UDIg qui est adaptée à chaque utilisateur. En revanche, ces études démontrent qu'il est possible d'exposer l'utilisateur à différentes activités et de lui laisser la liberté de proposer des gestes d'interaction. Nous avons aussi parcouru quelques exemples d'intégration de cette approche notamment utilisée à des fins d'usage personnalisé en informatique. Cependant, à notre connaissance, il n'existe pas d'évaluation d'usage avancée sur l'utilisation de cette approche en réalité virtuelle.

Dans l'optique d'une adaptation personnalisée de l'application à l'utilisateur, nous envisageons une mise en œuvre en deux étapes :

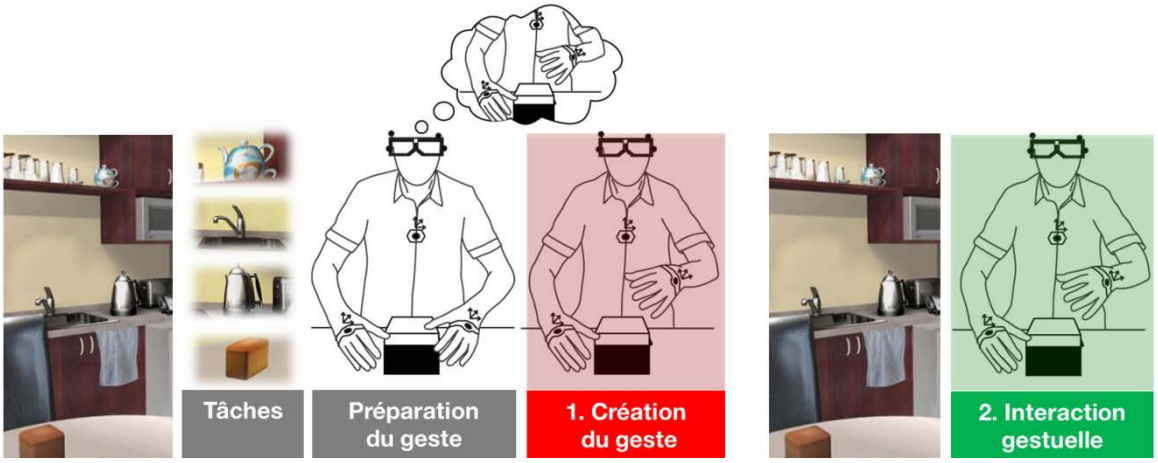


Figure 3-1 : Notre approche UDIg en deux étapes

1. Dans la phase de **création du geste**, l'utilisateur est exposé à l'objet qu'il devra manipuler. Si l'objet a une animation particulière, celle-ci doit se lancer afin que l'utilisateur puisse l'observer et comprendre l'activité. En premier lieu, il prépare son geste en l'imaginant par exemple, puis le propose ensuite au système. Si nécessaire, le geste peut être répété afin que l'utilisateur s'approprie l'interface, modifie, réajuste ou affine son geste. Cette opération doit être effectuée intégralement pour chaque objet nécessaire à la complétion de l'activité. Dans le cas d'une interaction avec des effets visuels ou d'animation d'un avatar, l'utilisateur peut également associer ses propres mouvements ou poses.

2. La deuxième étape est la phase d'**interaction par les gestes** où l'utilisateur peut réutiliser à sa guise les gestes appris au système afin d'effectuer différentes tâches dans l'environnement virtuel.

3.2. PRINCIPALES PROBLÉMATIQUES SOULEVÉES PAR L'APPROCHE UDIG

La mise en place d'une telle approche soulève un certain nombre d'interrogations dont l'acceptation par l'utilisateur, mais aussi les contraintes à la fois côté utilisateur et côté système. L'interrogation principale reste la mise en place de l'approche UDIg. Comment faire en sorte que l'utilisateur puisse proposer un geste au système ? Comment le système l'enregistre-t-il ? Lors de l'interaction, comment faire comprendre l'avancement de l'activité ?

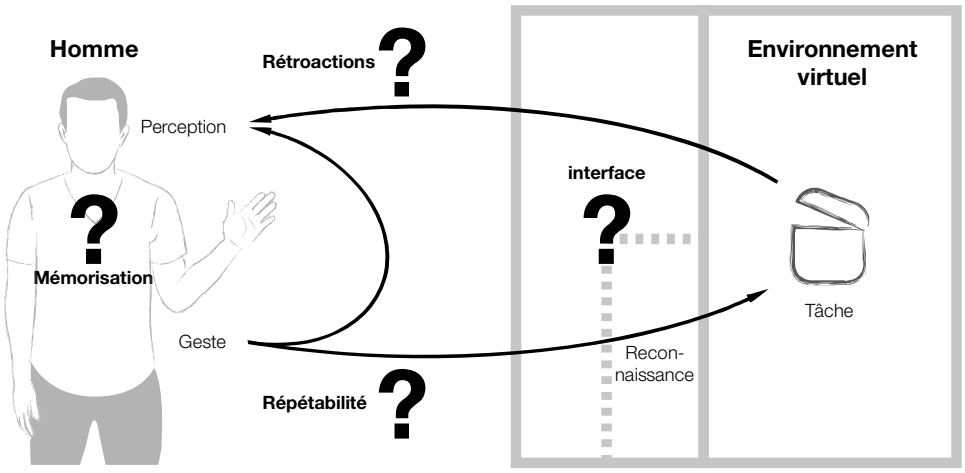


Figure 1-6 : Positionnement des problématiques soulevées par une approche où l'Utilisateur Définit l'Interaction gestuelle

Nous proposons de regrouper les interrogations sous-jacentes à cette problématique en quatre catégories principales (Figure 1-6) que nous détaillons dans les paragraphes suivants :

- La répétabilité du geste ;
- La mémorisation du geste et sa charge cognitive ;
- Les rétroactions et le couplage;
- Les considérations techniques sur l'interfaçage.

3.2.1. Répétabilité du geste et variabilité

Nous avons pour objectif que l'utilisateur puisse proposer et réutiliser des gestes sans l'obliger à reproduire exactement le même geste. Or l'utilisateur s'impose un geste. Cela nous amène à considérer ce qu'est la variabilité du geste et la tolérance d'un système aux variations. Comment permettre une interaction fonctionnelle tolérante à la variabilité du geste ? Que signifie « refaire le même geste » ? L'utilisateur doit-il s'astreindre à refaire le même geste ? Comment faire en sorte que l'utilisateur propose un geste reproductible ?

La fatigue joue a priori un rôle dans l'évolution d'un geste répété ou dans l'utilisation d'un vocabulaire de gestes pendant une durée prolongée. L'état de l'art sur l'ergonomie (0) suggère d'éviter les mouvements à répétition et certaines postures. Par exemple, interagir en gardant le bras en l'air sans support physique génère de la fatigue. De plus, avoir le bras au-dessus du cœur est déconseillé (syndrome des « bras de gorille »). Les résultats des études de conception

inclusive (cf. Table 2-2) montrent que certains gestes proposés sont difficilement réutilisables et peuvent causer une fatigue d'usage prématurée.

→ L'expérimentation du chapitre 4 propose d'étudier la répétabilité du geste. Dans l'expérimentation du chapitre 5, nous tâchons d'apporter des réponses à la question sur la variabilité à l'usage et le rôle de la fatigue. De plus, des solutions permettant de limiter la fatigue seront proposées et évaluées.

3.2.2. Mémorisation et charge cognitive

3.2.2.a *Charge cognitive et coordination motrice*

En informatique, les deux mains sont parfois utilisées pour interagir. Selon (Hand 1997), il existe deux approches : l'approche parallèle où les deux mains participent à la même action et l'approche asymétrique, où chaque main a une fonction. Nous avons vu précédemment (0) que (Cabral et al. 2005) ont implémenté un système de contrôle du déplacement et de manipulation par geste dans un système CAVE[3]. La main gauche permet de définir le contexte tandis que la main droite manipule. Outre la fatigue à l'usage, le problème soulevé est celui d'une augmentation de la charge cognitive. Dans le cas d'une interaction gestuelle bimanuelle, il faut bien évidemment prendre en considération la complexité de l'utilisation asymétrique. Cependant, dans le cas d'activités moins concrètes, l'utilisateur peut sans le savoir proposer un geste asymétrique qui peut se révéler complexe à réutiliser. Il nous semble préférable d'orienter l'interaction vers une approche symétrique limitant ainsi la charge cognitive.

3.2.2.b *Mémoire à court terme et empan mnésique*

Laisser à l'utilisateur la possibilité d'improviser des gestes lui donne une grande liberté, mais peut avoir quelques limites. Une interrogation liée à l'utilisation d'un vocabulaire de gestes créé par l'utilisateur est de savoir combien de gestes il peut se remémorer.

Étant donné que nous ne souhaitons pas contraindre l'utilisateur à se remémorer des gestes, nous nous adressons à une mémoire dite de travail. Le concept théorique de mémoire de travail présume que l'information est stockée et accessible temporairement dans la mémoire. Cela permet de supporter des charges mentales courtes et sert de tampon entre perception, mémoire à long terme et action (Baddeley 2003). Il est admis que cette mémoire de travail est limitée dans sa capacité. (Miller 1956) estime qu'un individu peut en moyenne se remémorer une suite de sept chiffres, à plus ou moins deux chiffres, son empan mnésique serait donc de 7±2. (LaViola et al. 2011) soulignent que l'un des problèmes en interaction gestuelle est la mémorisation du vocabulaire de gestes. Les auteurs présument qu'un utilisateur peut mémoriser sept gestes et soulignent qu'un usage avancé fera appel à des gestes plus complexes et plus nombreux. Cela posera des problèmes aux utilisateurs novices notamment. En fait, des études récentes (Farrington 2011) montrent que selon les conditions d'utilisation et la charge cognitive à l'usage, cette mémoire de travail serait plutôt limitée à trois ou quatre éléments.

[3] Le CAVE pour « Cave Automatic Virtual Environment » est une salle immersive cubique dans laquelle l'utilisateur est en immersion et interagit.

Une étude qui associe geste et remémoration a été conduite par (Wagner et al. 2004). Les auteurs s'interrogent sur les représentations mentales sous-jacentes au geste qui accompagne la parole en étudiant les mémoires de travail visuospatiales et verbales. Ils étudient l'impact de l'association d'un geste sur la remémoration d'une chaîne de lettres ou de symboles visuels. Au préalable, ces suites d'éléments doivent être mémorisées, puis les participants doivent se remémorer ces symboles tout en résolvant un problème mathématique verbalisé oralement avec ou sans gestes. Les auteurs notent que les participants se remémorent plus d'éléments lorsqu'ils effectuent des gestes. De plus, le nombre d'éléments remémorés dépend du sens véhiculé par le geste augmentant sensiblement lorsque ces gestes correspondent à la même information propositionnelle. Les gestes semblent ainsi participer à la remémoration d'éléments à condition que ces gestes soient corrélés au contenu. Cela corrobore l'étude de (Goldin-Meadow et al. 2013) décrite au chapitre 2. Cette étude montre que dans la communication orale, le geste semble refléter les pensées et le non verbal, et il peut réciproquement influencer la pensée de celui qui parle.

→ Nous constatons que le nombre de gestes à se remémorer peut être limité. Est-il possible de connaitre cette limite ? Quels seraient les facteurs qui aident ou limitent la remémoration ? Nous proposons dans le chapitre 4 d'étudier la remémoration.

3.2.3. Rétroactions et couplage

3.2.3.a *Le choix de la rétroaction visuelle*

L'état de l'art sur l'interaction et les rétroactions montre que ces notions se recoupent (2.2.3.a). Dans le cas d'une interaction gestuelle avec un dispositif informatique, le système donne un *feedback* sur l'action en cours. Comment choisir la rétroaction qui permet d'aider à la complétion de l'activité ? Il existe plusieurs façons d'interagir, entre commande discrète ou contrôle continu :

- La commande discrète correspond par exemple au déclenchement d'une action, pendant ou après l'exécution du geste ;
- Le contrôle continu correspond au couplage geste-action en continu ou par paliers pour une séquence d'actions.

→ Par conséquent, faut-il donner le *feedback* après le geste ou pendant le geste ? Si ce *feedback* intervient pendant le geste, a-t-il une influence sur le geste ? Le *feedback* peut-il impliquer davantage l'utilisateur dans la complétion de l'activité ? Nous tâchons de répondre à ses interrogations par une expérimentation décrite dans le chapitre 5.

3.2.3.b *L'interaction en continu*

Le contrôle continu amène un certain nombre d'interrogations. Par exemple, comment est perçu le couplage geste-action ? Nous savons que l'homme a, en parallèle, une connaissance a priori du résultat grâce son *feedback* intrinsèque et ses connaissances (2.2.3.a). Il se pose alors la question de la concordance entre le retour visuel et l'information proprioceptive de l'utilisateur. Il semble important de considérer le *feedback*, aussi bien intrinsèque

qu'extrinsèque par rapport à ce que renvoie l'environnement virtuel et ce qu'interprète l'utilisateur. Quelle est l'influence du couplage geste image sur le geste ?

Enfin, il est à noter que l'implémentation de différents *feedbacks* notamment continus peut se révéler complexe à mettre en œuvre ou plus coûteuse. Or nous souhaitons rendre l'approche accessible. Il semble pertinent de comparer les différences dans l'interaction selon des rétroactions plus ou moins complexes.

→ Quelle rétroaction donner à l'utilisateur ? Comment et quand faut-il mettre en œuvre cette rétroaction ? Quelle est l'influence du couplage ? Les expérimentations décrites dans le chapitre 5 donnent des solutions.

3.2.3.c *Colocalisation*

En fonction des types de capteurs, la colocalisation d'entités 3D peut se révéler difficile. Cela dépend notamment de la qualité du suivi de point de vue ou du relief. Aussi, les dispositifs utilisés peuvent limiter le « volume d'interaction ». Celui-ci est déterminé par l'espace de captation du geste dépendant du matériel.

→ Une situation non colocalisée pour interagir avec des gestes en environnement virtuel pose-t-elle des problèmes de compréhension ? A-t-elle une influence sur le geste ? Nous explorons cette problématique dans les chapitres 4, 5 et dans les perspectives du chapitre 7.

3.2.4. Considérations techniques pour la mise en œuvre de l'UDIg

Pour la mise en place technique, nous proposons de détailler notre approche d'apprentissage artificiel supervisé en deux étapes. Nous y incluons des considérations sur l'acquisition, la segmentation des données pour l'étape de classification et l'interprétation du geste avec des dispositifs peu performants (Figure 2-9).

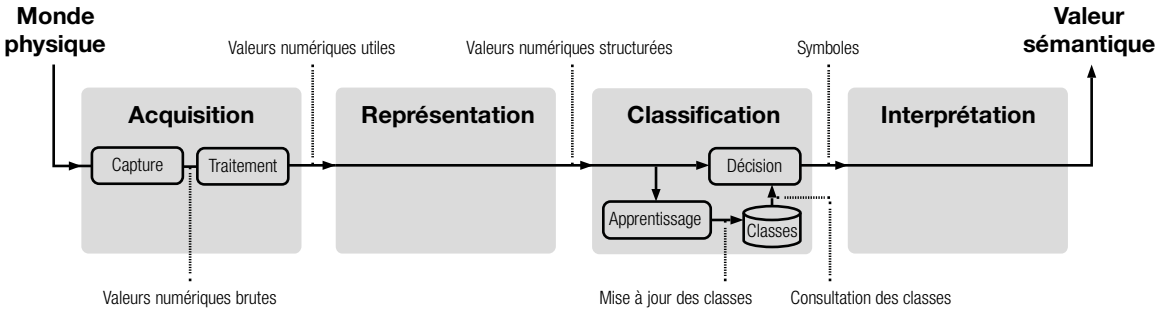


Figure 2-9 : Schéma général d'un système de reconnaissance de gestes (Braffort et al. 2006)

3.2.4.a *Acquisition avec des systèmes peu performants*

Variabilité du geste

Comment prendre en compte la variabilité du geste ? Si nous prenons l'exemple des modèles d'apprentissage basé *HMM*, ceux-ci tirent profit de la variabilité du geste. En effet, l'apprentissage d'un même geste se fait par répétition. Les systèmes de reconnaissance de type

HMM et *DTW* ont par conséquent une certaine tolérance aux variations spatiales du geste et à sa dynamique, sa temporalité, etc. (2.2.2.c). Cependant, un geste trop différent entre sa création et sa réutilisation dans l'interaction risque d'être mal ou non reconnu et l'interaction ne sera pas utilisable. Dans le cas d'un apprentissage à un échantillon (*one shot learning*), en revanche, le système doit permettre une certaine tolérance à la variabilité du geste reproduit par rapport au seul geste de référence appris. Dans ce cas, comment avoir un geste représentatif ? Nous envisageons des solutions dans les chapitres 4 et 5.

Transparence de l'interface

Au niveau de l'acquisition certains auteurs soulèvent la nécessité d'une réelle transparence de l'interfaçage (Fuchs et al. 2003; Norman 1989a). Comment permettre une transparence de l'interfaçage pour l'interaction ? Celle-ci concerne aussi bien le caractère non intrusif des capteurs que la marge de manœuvre pour l'interaction. Si l'utilisateur se déplace dans la pièce, le système de reconnaissance doit permettre la reconnaissance du geste. Nous explorerons plusieurs stratégies dans les chapitres 4 et 5.

Isoler la partie signifiante du geste

À noter qu'il faut considérer la découpe de la partie signifiante du geste pour la classification L'idée est d'en isoler le *stroke*. Mais comment procéder ? Il existe différentes approches. Il est possible de l'isoler à l'apprentissage par une découpe automatique, ou éventuellement par une action manuelle de l'utilisateur lui-même ou par un opérateur. L'implémentation du système dans les expérimentations des chapitres 4 et 5 donnera des préconisations sur la découpe du *stroke*.

3.2.4.b *Interprétation avec des systèmes peu performants*

« Syndrome de l'immersion »

Nous avons vu en interaction gestuelle le « syndrome de l'immersion » (2.2.2.). Par conséquent il est nécessaire de pouvoir isoler les gestes non intentionnels des gestes d'interaction comme des gestes intempestifs de grattage ou des gestes coverbaux. Il y a également des gestes provoqués par des pathologies comme les tremblements ou les myoclonies. Dans l'art de la scène, des zones d'interactions précises peuvent être définies afin de permettre des moments d'improvisation hors du cadre d'interaction prévu. Nous proposerons des solutions dans les chapitres 4 et 5, ainsi que dans les perspectives.

Latence

La minimisation de latence reste un enjeu constant en réalité virtuelle. Et chez les fabricants de matériel ou chez les concepteurs de jeux vidéo (McMahan et al. 2010). L'idée est de gagner du temps de traitement aussi bien au niveau des relevés de mesure que dans le traitement de l'information.

Mais mesurer précisément la latence reste une problématique ouverte. Par conséquent, l'utilisation d'interfaces à bas coût peut fortement limiter les conditions d'immersion et d'interaction selon leur qualité. Il existe des limitations inhérentes aux degrés de liberté disponibles, à la fréquence et à la latence, etc. Par conséquent, il se révèle difficile de donner une rétroaction de haute-fidélité (Vercher 2006). (Regan 1995) souligne qu'une latence trop

importante modifie le comportement « naturel » de l'utilisateur qui ralentira ses mouvements. Dans le cas d'un couplage geste-image, la latence peut perturber le couplage à cause d'une animation arrivant en retard par rapport à ce qui est attendu. Tout l'enjeu reste de minimiser ces artefacts, voire de les prévoir et de les pallier. Dans ce cas, la question est de savoir comment compenser les effets de la latence. Nous envisageons des solutions dans les chapitres 5 et 7.

Couplage geste-animation

Dans le cas d'une interaction entre le geste et une rétroaction avec un couplage, en plus du problème du découpage du geste, il y a celui de l'association du geste à l'animation. Certains systèmes de reconnaissance permettent une synchronisation optimale, en s'affranchissant notamment des variations temporelles (Figure 3-2). Indépendamment de la qualité de l'interfaçage, il semble toutefois difficile de coupler le geste et l'image de façon optimale, car l'animation peut avoir de l'avance ou du retard. La latence est un facteur déterminant. Il semble nécessaire d'envisager des solutions pour pallier ces problèmes éventuels de couplage et rendre l'interaction acceptable et fonctionnelle. Nous proposons aux chapitres 5 et 6 des solutions afin de permettre une interaction optimale.

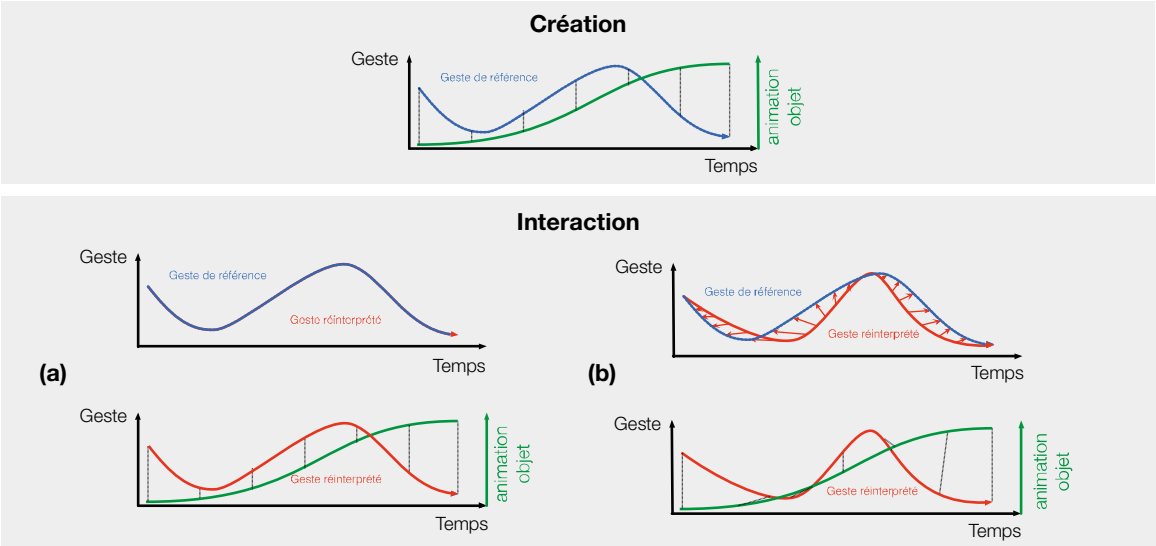


Figure 3-2 : Couplage geste et animation d'objet aux phases de création et d'interaction du geste (a). À noter en (b) que l'invariance aux variations temporelles permet de conserver la synchronisation du mouvement d'un geste à l'animation. Adapté de (Françoise et al. 2013)

CHAPITRE 4
EXPÉRIMENTATION SUR LA REMÉMORATION DE GESTES PROPOSÉS PAR L'UTILISATEUR

Afin de répondre aux problématiques soulevées, nous conduisons une première expérimentation. Celle-ci permet de tester la mise en œuvre de l'approche et d'observer son appropriation par les utilisateurs. La question de recherche principale de cette expérimentation est centrée sur la proposition et la mémorisation d'un vocabulaire de geste. Nous allons isoler les hypothèses centrées sur le schème et les vérifier par une expérimentation.

4.1. QUESTION DE RECHERCHE ET HYPOTHÈSES

Cette expérimentation a deux enjeux principaux. Le premier est la mise en œuvre de l'approche et l'évaluation de son appropriation par les utilisateurs. Le second est orienté sur la mémorisation. Si l'utilisateur a la possibilité de créer un vocabulaire de gestes pour interagir, sera-t-il en mesure de se remémorer ses gestes ? L'état de l'art montre que l'utilisateur a une mémoire de travail limitée. Est-il possible de connaitre cette limite ? Dans ce cas, quels seraient les facteurs qui aident ou limitent la remémoration d'un vocabulaire de gestes défini par l'utilisateur ?

Notre état de l'art a permis d'isoler plusieurs types de connaissances de l'utilisateur (2.1.1.b). Les approches possibles sont le schème, les métaphores et les affordances. Nous proposons de nous focaliser sur le schème qui permet d'éviter un apprentissage, car celui-ci correspond à une réutilisation de connaissances acquises. On pourrait penser que la solution idéale serait de faire en sorte que l'utilisateur ait systématiquement recours à des schèmes pour interagir. Or, (Fuchs et al. 2003) notent que toutes les activités en virtuel ne peuvent être basées sur des schèmes, et qu'il est possible d'employer des métaphores. C'est par exemple le cas d'un déplacement en environnement virtuel sans pouvoir bouger physiquement le corps. Il est possible d'utiliser des commandes de déplacement au moyen périphérique isométrique comme un joystick ou un *wand*[4]. On peut également faire appel à différentes techniques d'interaction (voir (Mine 1995) pour une liste d'exemples). L'utilisation de gestes du bras pour se déplacer en virtuel peut avoir une signification beaucoup plus abstraite pour l'utilisateur qu'un schème de déplacement du corps plus écologique. Il semble pertinent de comparer la mémorisation d'éléments concrets et basés sur des schèmes par rapport à la mémorisation d'éléments plus symboliques ou imagés.

En nous reposant sur le choix du schème comme aide à la mémorisation, nous souhaitons savoir ce qui le caractérise. Nous constatons qu'en situation réelle, le geste basé sur un schème est induit par les affordances. Celles-ci sont des indices visuels d'utilisation de l'objet (Gibson 1979). Aussi, un schème de manipulation s'effectue de façon colocalisée directement sur l'objet en étant soumis à des contraintes de l'environnement physique. (Norman 1989a) parle de contraintes naturelles ou symboliques. Ces concepts auraient-ils un rôle dans la remémoration des gestes ? Nous proposons de détailler ces concepts dans les paragraphes suivants.

4.1.1. Hypothèse A : affordances des objets

L'un des moyens pour susciter un schème pour l'interaction avec des objets ou scènes virtuels est de donner à l'utilisateur suffisamment d'indices. Ces indices lui permettent d'anticiper l'utilisation d'objets ou l'action. Comme décrit au paragraphe 2.1.1.b, ces indices visuels (mais aussi tactiles ou auditifs) sont appelés affordances. À noter qu'un article de (McGrenere et al. 2000) proposent de clarifier ce concept selon le point de vue de Gibson et celui de Norman (Table 4-1) et notent une différence fondamentale entre les définitions. Pour Gibson, l'affordance est la possibilité d'action elle-même tandis que pour Norman c'est à la fois la possibilité d'action et la façon dont cette possibilité d'action est transmise ou rendue visible à l'acteur. Il demeure important qu'elles soient explicites et visibles pour guider l'utilisateur.

[4] Équivalent du joystick utilisé dans des dispositifs de réalité virtuelle, souvent sous forme de télécommande.

Table 4-1: Comparaison des copropriétés des affordances selon Gibson et selon Norman (McGrenere et al. 2000)

Selon Gibson	Selon Norman
▪ Possibilités d'action de l'environnement vis-à-vis des possibilités d'action d'un acteur ; ▪ Indépendante de l'expérience, de la connaissance, de la culture ou de la capacité à percevoir de l'acteur ; ▪ Existence binaire : une affordance existe ou elle n'existe pas.	▪ Suggestions ou indices sur la façon d'utiliser ; ▪ Peut être dépendante de l'expérience, des connaissances ou de la culture de l'acteur ; ▪ Propriétés perçues qui peuvent ou ne peuvent pas exister réellement ; ▪ Peut rendre une action difficile ou facile.

À noter que les affordances existent en informatique principalement pour donner des indices d'utilisation à l'utilisateur afin de rendre l'interface accessible. On les retrouve dans le principe de bureau virtuel des systèmes d'exploitation. Dans cette même optique (Mullaly 1998) propose la conception d'interfaces de logiciels aux apparences visuelles similaires à celles d'objets réels, en leur conférant un rendu « réaliste » et des indices visuels d'utilisation explicites. Il a conçu par exemple un logiciel d'agenda rappelant visuellement un agenda réel. Il existe aussi des extensions issues de ces affordances appelées skeuomorphismes[5]. Ceux-ci rendent les interfaces plus familières et aident donc explicitement et implicitement à la prise en main par des utilisateurs novices en faisant appel à leurs connaissances. Ils adopteront ainsi un comportement à l'utilisation du logiciel proche du comportement habituel devant un objet similaire. En revanche, il n'est pas étonnant de voir certains utilisateurs experts considérer ces artifices visuels superflus.

Dans le cas d'une tâche de manipulation, (Rizzolatti et al. 1997) définissent les affordances comme les éléments nécessaires à la construction de possibilités motrices de saisie, dépendant de la position relative de l'objet par rapport au corps. Dans le domaine de la manipulation en réalité virtuelle, (Kallmann et al. 1999) proposent les *Smart Objects* : la main virtuelle de l'utilisateur est entourée de deux mains-aides donnant les possibilités d'interaction de différentes parties de l'objet. (Smets et al. 1994) montrent que l'association des affordances et de la manipulation directe est un moyen d'améliorer le design d'objet en environnement virtuel. Cela suggère que la colocalisation pourrait avoir un rôle dans la perception des affordances, dans la construction du schème et, par conséquent, dans la remémoration.

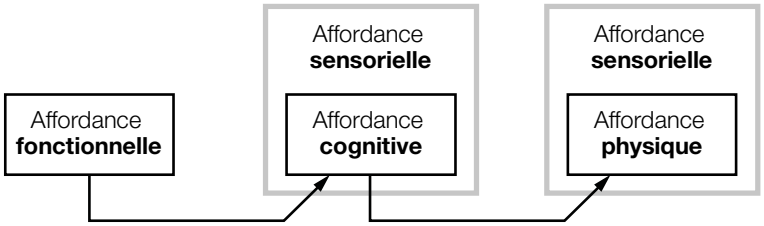


Figure 4-1 : Affordances présentes à trois niveaux : physique, cognitif et fonctionnel selon (Hartson 2003)

[5] Notons par exemple la sonnerie d'un smartphone imitant la sonnette mécanique d'un combiné, le son d'un déclencheur mécanique sur un appareil photo numérique, ou encore le dessin du relief des touches de clavier F et J comme aide à la frappe qui persiste encore sur les claviers virtuels de certaines tablettes numériques.

À noter que (Hartson 2003) propose de prendre en considération les affordances à plusieurs niveaux dans la conception d'une interaction (Figure 4-1). Ces niveaux coïncident avec les trois niveaux du modèle de conception $3I^2$ de (Fuchs et al. 2003) détaillés au chapitre 2. À partir de ces éléments, nous émettons une première hypothèse :

→ **Hypothèse A :** la présence d'affordances visuelles explicites facilite la remémoration d'un vocabulaire de gestes. Réciproquement, des gestes non basés sur des affordances visuelles sont difficilement remémorés.

4.1.2. Hypothèse B : contraintes et colocalisation

Dans le cas de la manipulation d'un objet réel, un schème est aussi caractérisé par le fait que ledit objet soit manipulé de façon directe, ou colocalisée. Si la manipulation se fait par la main, l'utilisateur ressent cet objet réel, sa position, sa forme grâce aux informations tactiles et proprioceptives. Le geste est alors dirigé par la perception de l'environnement physique (Gibson 1979). En réalité virtuelle, la colocalisation, aussi appelée manipulation directe, est une situation où l'espace de manipulation et l'espace de visualisation sont superposés. Une situation où l'espace de manipulation et le rendu visuel ne le sont pas correspond à une manipulation non colocalisée ou à distance. Il est fréquent de manipuler des objets colocalisés dans un environnement virtuel, notamment sur les dispositifs appropriés comme les systèmes Holobench[6] ou les systèmes CAVE.

Cependant, nous savons que dans une application de réalité virtuelle le retour sensoriel sur les aspects fins de la manipulation manque souvent (manipulation fine avec les doigts, retour tactile, haptique ou colocalisation, etc.). Cela oblige à manipuler « dans les airs » et nous avons vu que l'utilisateur compensera le manque d'information haptique perçue, par la vision notamment (cf. 2.2.3.a). (Mine et al. 1997) estiment que la prise en compte des informations proprioceptives et kinesthésiques permettrait également de compenser dans une certaine mesure l'absence de retour haptique. Ces informations relatives au corps donnent ainsi la possibilité de manipuler presque « les yeux fermés ». Les auteurs recommandent de permettre une interaction de préférence dans un volume accessible par les bras, c'est-à-dire dans le volume de travail naturel de l'utilisateur. Cela permet un couplage direct entre le mouvement de la main et celle de l'objet, avec plus de précision. Dans le cas d'une tâche de pointage, (Paljic et al. 2002) montrent que la manipulation directe et proche de l'objet est plus efficace qu'une manipulation à distance (avec ici une distance supérieure à 40 cm). La vitesse du curseur 3D et les indices visuels associés, comme un rayon symbolisant le décalage entre le pointeur et l'objet, diminueraient les performances. On peut interpréter que l'environnement virtuel et les objets ne doivent pas comporter d'indices visuels artificiels ou superflus et notamment pour ne pas perturber l'évocation du schème. Nous émettons une seconde hypothèse sur le rôle de la colocalisation :

→ **Hypothèse B :** la colocalisation en environnement virtuel, existante en situation réelle, contribue à la remémoration du geste. À l'inverse, la manipulation distante baisserait la performance de remémoration des gestes.

[6] Le Holobech est un dispositif de type CAVE à échelle réduite et souvent à deux faces. Il permet une manipulation bimanuelle d'objets en relief.

4.1.3. Hypothèse C : mémorisation et empan mnésique

La question de la variété des gestes et de leur mémorisation est centrale à notre approche. Comme précisé dans les problématiques soulevées, la mémoire de travail permet de stoker un nombre limité d'informations (cf. 3.2.2.a). Dans notre expérimentation, nous souhaitons observer si effectivement l'utilisateur peut se remémorer plusieurs gestes. (Farrington 2011) montre que selon les conditions et la charge cognitive à l'usage, cette mémoire de travail serait limitée à trois ou quatre éléments. Et si l'utilisateur n'a pas pour consigne explicite de mémoriser ses gestes, mais simplement d'associer un geste d'interaction pour chaque élément, nous supposons que ce nombre sera effectivement très limité.

→ **Hypothèse C :** on note une diminution de la performance de remémoration dans le cas de gestes non basés sur des affordances explicites. Selon l'état de l'art et les conditions de charge cognitive, cette diminution se constate à partir de trois éléments à se remémorer.

4.2. L'EXPÉRIMENTATION

4.2.1. Tâche

Nous proposons une expérimentation afin de valider la mise en œuvre de l'approche et nos hypothèses. Le principe de l'expérience est d'inviter les participants à proposer des gestes de leur choix pour ouvrir une série de boites. Les boites affichées présentent soit des indices visuels explicites —c'est-à-dire des affordances— sur leur mécanisme d'ouverture, soit elles sont simplement différenciées par leur couleur. Pour les boites avec affordances comme pour les boites de couleur, il est expliqué aux participants qu'ils devront proposer le geste de leur choix. La seule consigne donnée est de proposer des gestes différents pour chaque boite. Les gestes proposés sont enregistrés dans la phase de création puis reconnus par le système pendant la phase d'interaction. Nous détaillons ce protocole dans le paragraphe suivant.

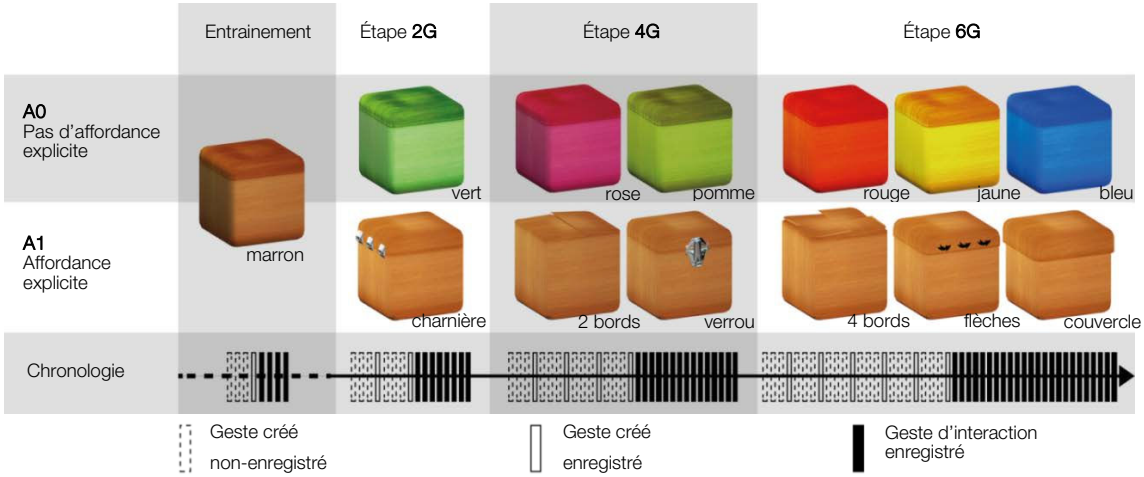


Figure 4-2 : Étapes de l'expérimentation, liste des boites, conditions d'affordance, et chronologie de l'expérimentation

4.2.2. Protocole

Avant de commencer l'expérimentation, chaque participant précise sa latéralité (gaucher, droitier ou ambidextre). Les participants sont invités à proposer des gestes spontanés pour ouvrir chaque boite. Nous précisons qu'il n'y pas de retour haptique : en situation colocalisée, les objets apparaissent entre l'écran physique et le sujet, à portée de main. Le but de l'étude n'a pas été exposé aux participants ni le nombre d'étapes ni le nombre de boites à ouvrir. Durant l'expérimentation, les participants ne savent pas s'ils effectuent le bon geste pour chaque boite. Cela permet d'éviter qu'ils ne procèdent par élimination pour la remémoration.

- Pendant la phase de **création du geste**, une série de boites virtuelles modélisées en 3D est affichée sur une table devant l'utilisateur. Il est demandé aux participants d'approcher leur main de la boite et de proposer un geste différent pour ouvrir chaque boite. Les participants doivent répéter quatre fois leur geste, seule la quatrième répétition est enregistrée (nous verrons au paragraphe 4.2.5 que les prétests ont permis d'estimer que le geste est déjà stabilisé à sa quatrième répétition). De plus, les utilisateurs vont découvrir d'eux-mêmes qu'il s'agit d'une tâche de remémoration et s'imposeront ou non la

mémorisation du geste qui correspond à chaque boite. Il leur est cependant imposé de ne jamais proposer de gestes similaires sur l'ensemble des boites.

- Pendant la phase **d'interaction par le geste**, pour chaque condition, toutes les boites sont affichées tour à tour dans un ordre aléatoire. L'utilisateur doit effectuer le même geste que celui qu'il a proposé dans la phase de création de gestes au regard de chaque boite.

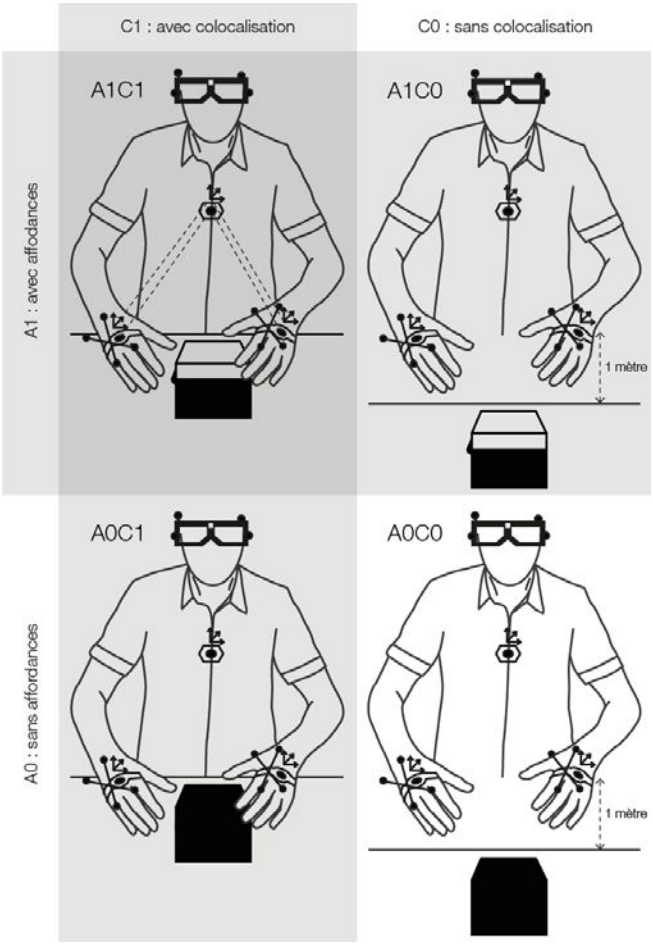


Figure 4-3 : Les boites sont présentées à l'utilisateur selon quatre combinaisons de conditions d'expérimentation. A0: pas d'affordance explicite, A1: affordance explicite. C0: manipulation distante, C1 manipulation colocalisée

4.2.2.a *Conditions expérimentales*

La Figure 4-3 décrit les quatre conditions d'expérimentation qui concernent l'affordance et la colocalisation. Les deux variables indépendantes (affordances et colocations) sont croisées pour donner 4 conditions. Chaque sujet effectue les 4 conditions (*within subject design*). Les boites sont présentées à l'utilisateur selon quatre combinaisons :

- A0 : pas d'affordance explicite ;
- A1 : affordance explicite ;
- C0 : manipulation distante ;
- C1 : manipulation colocalisée.

La Figure 4-3 indique également le placement des capteurs pour cette manipulation. À noter que pour la condition d'interaction C0, la boite apparait à un mètre plus loin qu'à la condition C1. L'utilisateur ne peut alors pas interagir directement sur la boite présentée dans le cas C0. Dans cette condition, il a pour consigne d'interagir au même endroit que lorsqu'il a créé son geste en situation colocalisée.

Les trois niveaux de difficulté sont proposés et représentés sur la Figure 4-2. Au début de chaque étape, pendant la phase de création du geste, les utilisateurs doivent proposer quatre répétitions du geste (barres en pointillés). Seule la quatrième occurrence du geste est enregistrée (barres blanches) dans le système d'apprentissage artificiel. Les gestes d'interaction sont ensuite effectués et répétés (barres noires) et comparés à ceux proposés dans la phase de création). L'expérimentation dure environ 20 minutes par participants. Au total, chaque sujet effectue 104 gestes.

4.2.2.b *Dispositif expérimental*

L'expérimentation a lieu dans une salle immersive équipée d'un écran large rétroprojeté de 3.1m de long par 1.7m de haut. Sa résolution est de 1920 par 1080 pixels à 120Hz, ce qui permet un rendu stéréoscopique actif[7]. Afin d'avoir une perception du relief optimale, la distance interoculaire (DIO) du participant est mesurée grâce à un pupillomètre. Cette distance permet de calibrer correctement les caméras de rendu stéréoscopique du moteur 3D temps-réel.

Figure 4-4 : Un utilisateur pendant l'expérimentation et détail du dispositif

Acquisition précise pour l'immersion et la mesure des gestes

Deux caméras infrarouges[8] permettent le suivi du point de vue et la capture de mouvements de la main dominante de l'utilisateur. Le *headtracking* n'a pas été fait avec des capteurs à bas coût délibérément, car il s'agissait dans un premier temps de tester l'approche UDIg et la mémorisation des gestes au regard des affordances. Un suivi du point de vue de qualité permet de s'assurer que la perception des boites virtuelles par l'utilisateur est meilleure qu'avec un système à bas coût. Les positions et les orientations absolues de chaque main ont été enregistrées et horodatées à 30Hz au moyen du système de capture *A.R.T.II*.

[7] Rendu avec carte graphique *nVidia Quadro FX 5600*
[8] Deux caméras infrarouges *A.R.T.II.* de *Advanced Real-time Tracking GmbH* et trois cibles optiques

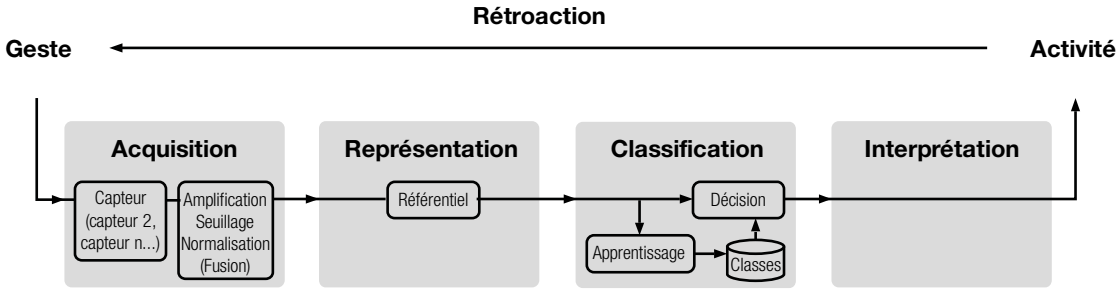


Figure 4-5 : Schéma du système de reconnaissance de gestes utilisé pour l'UDIg

Acquisition pour la reconnaissance en temps-réel

Le système de reconnaissance basé sur des Modèles de Markov Cachés (ou HMM) choisi prend en entrée des positions et des orientations normalisées comme descripteurs 6 degrés de liberté (DOF) du mouvement des deux mains. Pour ce faire, nous avons combiné des capteurs à bas coût. Les prétests ont montré que le placement de chaque capteur sur les mains fournit plus de variations de degrés de liberté que sur les poignets ou sur les avant-bras. Cela se révèle aussi moins intrusif que de le placer sur les doigts. Afin de capter les mouvements de la main indépendamment des mouvements du haut du corps de l'utilisateur, les coordonnées des mains sont calculées dans le référentiel du torse grâce à un capteur posé sur le torse.

Choix des capteurs

À propos des descripteurs gestuels, nous optons pour l'utilisation de données de capteur normalisées qui correspondent à des primitives directes (cf. 2.2.2.b). Cela semble satisfaisant pour un usage simple et permet d'éviter des traitements en cascade qui peuvent ajouter de la latence et des implémentations techniques supplémentaires. À noter que l'avantage technique de procéder à un apprentissage par l'utilisateur à partir de données brutes est de ne pas avoir à placer les capteurs de façon précise et de ne pas demander à l'utilisateur de faire une calibration préalable. Cela permet de faire abstraction de la configuration du capteur. Pour cette expérimentation, notre choix s'est porté sur deux capteurs, la caméra *Kinect* de *Microsoft* et les centrales inertielles *MotionPods* de *Movea*. Ceux-ci fournissent des données partielles, mais qui se complètent afin d'obtenir 6 DDL pour chaque main en positions et orientations absolues. Nous procédons à une fusion de données entre différents capteurs si certains ne donnent que des DDL partiels.

En détail, les positions des mains et du torse sont fournies par une caméra *Kinect* et les données sont transmises par le logiciel *FAAST* (Suma et al. 2011) et par le protocole réseau *VRPN*. Nous avons choisi de normaliser les positions 3D de chaque main dans une amplitude qui correspond à une sphère 3D de rayon de 1m partant du torse. Trois centrales inertielles *Movea MotionPods* et un serveur de *VRPN* maison permettent d'avoir les orientations du torse et des deux mains. Un *MotionPod* est placé près de l'encolure et les deux autres sont placés sur le métacarpe de chaque main (Figure 4-4). À propos de la précision et du bruit des appareils à faibles coûts utilisés, les angles des capteurs *MotionPod* se révèlent suffisamment précis (< 0,01 degré, avec très peu de dérives). La caméra *Kinect* a en revanche une précision métrique plus approximative (dérive jusqu'à 3 cm) et du bruit. Par conséquent, afin de compenser le faible bruit et les artefacts éventuels, nous appliquons des filtres tels que

la moyenne des 5 dernières valeurs combinées à un seuillage. Cela crée une légère latence, mais les prétests ont permis d'obtenir des taux de reconnaissance satisfaisants.

Invariance aux déplacements

Afin de compenser la variabilité due aux déplacements du corps dans l'espace, il est possible de calculer les coordonnées des capteurs dans le référentiel du corps. Nous détaillons cette méthode : concrètement, les données peuvent être représentées dans un référentiel différent de celui du capteur. À propos des référentiels, (Berthoz 1997b) en distingue deux principaux :

- Le référentiel environnemental dit allocentrique qui décrit les mouvements par rapport à des repères extérieurs. Dans ce cas la position absolue du capteur est représentée par rapport à l'espace de l'environnement, de la pièce ;
- Le référentiel égocentrique est définit dans la relation au corps. Dans ce cas il est possible de calculer la position d'un membre dans le référentiel d'un autre.

Afin de permettre une invariance aux mouvements de l'utilisateur qui peut se pencher ou se tourner, nous proposons de calculer les positions et orientations des mains dans le référentiel du torse de l'utilisateur plutôt que dans le référentiel absolu du la pièce.

Découpe du *stroke*

La partie signifiante du geste (*stroke*) peut être isolée de façon automatique ou manuelle par un opérateur par un système d'embrayage et de débrayage. La segmentation automatique correspond par exemple à la découpe par un algorithme, ou par la position du geste en référence à une boite englobante. La découpe manuelle peut se faire en utilisant un bouton. Afin d'automatiser la découpe pendant l'activité de l'utilisateur, nous optons pour le principe de la boite englobante (photo Figure 4-6).

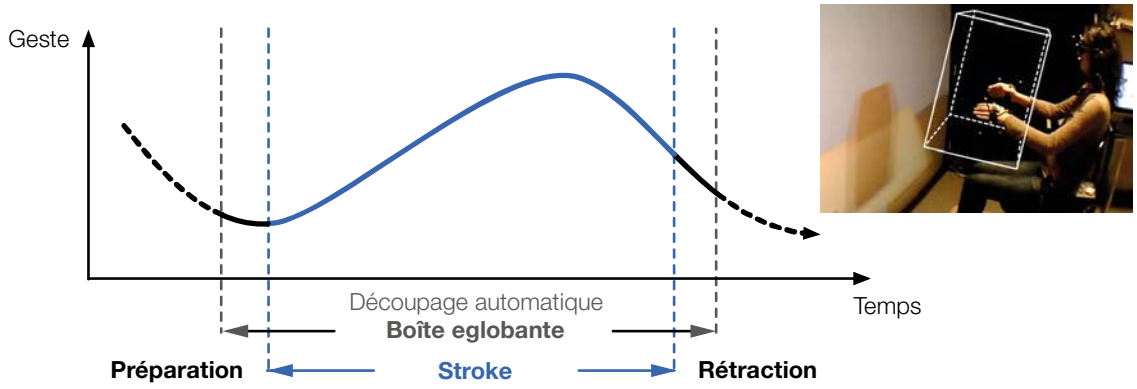


Figure 4-6: Notre approche du découpage d'une geste basé sur (Kendon 1980). À droite, un utilisateur effectuant un geste de manipulation dans une boite englobante

Certes la découpe reste approximative pour isoler le *stroke* de façon optimale, mais elle permet une segmentation satisfaisante et sans mise en œuvre complexe ou sans l'intervention d'un opérateur. À noter que ce découpage est contextuel. En effet, c'est la proximité à l'objet qui

définit la découpe. Il serait aussi possible de procéder à l'inverse : la segmentation du geste débute lorsque la main quitte sa zone de repos.

Dans cette expérimentation, les participants sont invités à poser leurs mains sur leurs cuisses au début puis à la fin de chaque répétition des gestes. Cela permet d'une part de se relaxer entre deux gestes comme recommandé par (M. Nielsen et al. 2004). D'autre part, cela permet de segmenter le geste avec une phase de repos intermédiaire. L'enregistrement et la reconnaissance des gestes ne se font que lorsque la main dominante de l'utilisateur traverse un volume invisible englobant la boite. Cela permet de supprimer automatiquement les phases de préparation ou rétraction (arrivée et retrait de la main) tout en conservant une partie proche du *stroke*. L'utilisateur est informé de l'enregistrement de son geste par deux sons : un premier indique le commencement de l'enregistrement du geste lorsque la main traverse la boite englobante et un second signale la fin lorsqu'elle quitte le volume.

4.2.3. Métriques de comparaison

L'évaluation de la mesure de similarité entre plusieurs occurrences d'un geste n'est pas triviale. Nous avons vu qu'il existe plusieurs méthodes pour comparer les gestes selon différents critères. Il est possible d'avoir recours à des algorithmes de classification et d'apprentissage (2.2.2.c) ou à des méthodes de comparaison a posteriori (2.2.3.c). Par conséquent, afin d'isoler les paramètres spatiaux et temporels, nous avons sélectionné trois critères qui nous semblaient les plus pertinents durant les prétests. Les comparaisons ont été faites par des algorithmes décrits dans l'état de l'art. Les critères pris en compte sont :

- Le temps de réflexion. Il correspond au temps chronométré dès l'apparition de la boite à la fin du geste d'interaction (hors de la *boite englobante*). À ce temps est soustrait le temps du geste créé à la phase d'apprentissage lorsque la main est dans la *boite englobante* ;
- La distance euclidienne moyenne des trajectoires en position de la main dominante entre le geste créé et le geste d'interaction. Celles-ci sont mesurées avec un système de capture de mouvements précis. Toutes les trajectoires des gestes ont été échantillonnées à 200 points dépendamment de l'horodatage.
- La mesure de similarité (probabilité) de reconnaissance du geste relatif au plus haut taux de complétion du geste détecté par le système. Nos prétests ont permis de déterminer que cela s'avère plus représentatif que de moyenner les taux de probabilité de reconnaissance du geste en continu.

4.2.4. Population

L'étude a été menée avec 21 participants (2 gauchers et 19 droitiers), âgés de 21 à 34 ans (M = 26.5) avec une distance interpupillaire moyenne de 63.9 mm (amplitude de 56 mm à 70 mm). L'acuité stéréoscopique de tous les participants a été mesurée au moyen du *test de Wirt*, M = 82.3 % (amplitude de 10 % à 100 %). Aucune des personnes n'est daltonienne.

4.2.5. Prétests

Afin de limiter les problèmes et les biais, des prétests ont été conduits avec un groupe de participants indépendant de l'étude. Ils ont permis de vérifier que l'approche en deux phases est acceptée par les participants. Ces prétests ont également permis d'établir que quatre gestes suffisent à l'obtention d'un geste stable et une reconnaissance satisfaisante. En effet, le

système d'apprentissage *Gesture Follower* (Bevilacqua et al. 2010) implémenté permet un apprentissage à un seul échantillon. Cela pose la question de la capture d'un geste « stabilisé ». Nous avons observé de façon empirique que la première occurrence d'un geste est hésitante. Cela vient du temps d'adaptation de l'utilisateur au système et surtout du fait qu'il manipule dans les airs sans support. Cette condition sans retour haptique peut être perturbante pour l'utilisateur habitué à reposer son bras sur l'objet à manipuler. Par conséquent, la deuxième répétition est nettement moins approximative puis une certaine aisance apparait dès la troisième répétition. Une quatrième répétition permet ainsi de valider la stabilité du geste. Enfin, ces prétests ont permis de déterminer les trois métriques de comparaison parmi une liste d'options.

4.3. RÉSULTATS

Les graphiques des Figure 4-7, Figure 4-8 et Figure 4-9 synthétisent les résultats pour respectivement le temps de réflexion en millisecondes, les distances euclidiennes des trajectoires moyennées entre le geste original et sa reproduction, et les taux de probabilité de reconnaissance du geste. Les différents niveaux de difficulté de mémorisation 2G (pour 2 gestes à se remémorer), 4G (pour 4 gestes) et 6G (pour 6 gestes), sans la partie entrainement, sont représentés par colonnes. L'écart type est indiqué sous la forme de barre d'erreur. Les moyennes des tests sont analysées par des analyses de variance (ANOVA) à deux facteurs et les résultats chiffrés et détaillés sont en Annexe A. Nous rappelons que les boites sont présentées à l'utilisateur selon quatre combinaisons : A0 (pas d'affordance explicite), A1 (affordance explicite), C0 (manipulation distante) et C1 (manipulation colocalisée).

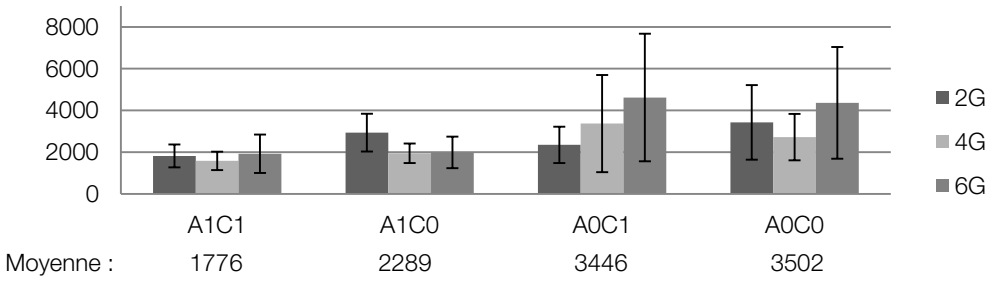


Figure 4-7 : Temps de réflexion en millisecondes. Les valeurs basses représentent une meilleure performance

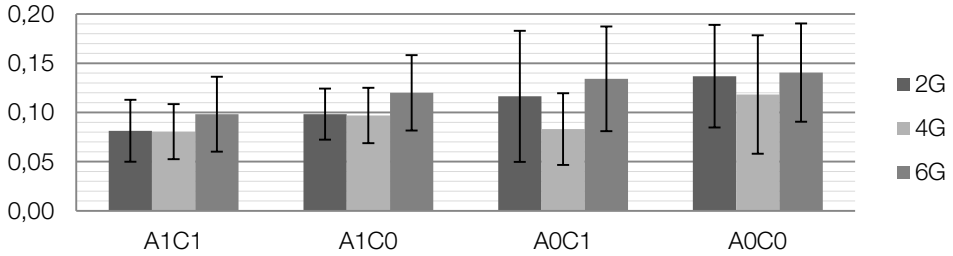


Figure 4-8: Distances euclidiennes moyennées des trajectoires entre le geste original et sa reproduction. Les valeurs basses représentent une meilleure performance (en mètres)

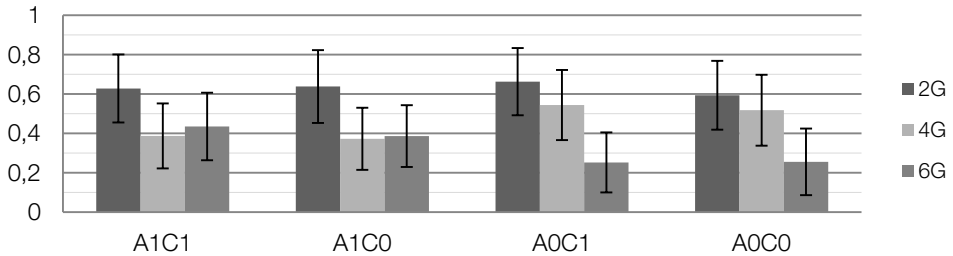


Figure 4-9 : Taux de reconnaissance du geste (1 correspond à 100%). Les valeurs hautes représentent une meilleure performance

La Figure 4-10 représente la comparaison entre les trajectoires du geste créé (en noir) et du geste reproduit (en couleur) à l'interaction (référentiel absolu). Nous décrivons ces résultats dans les paragraphes suivants.

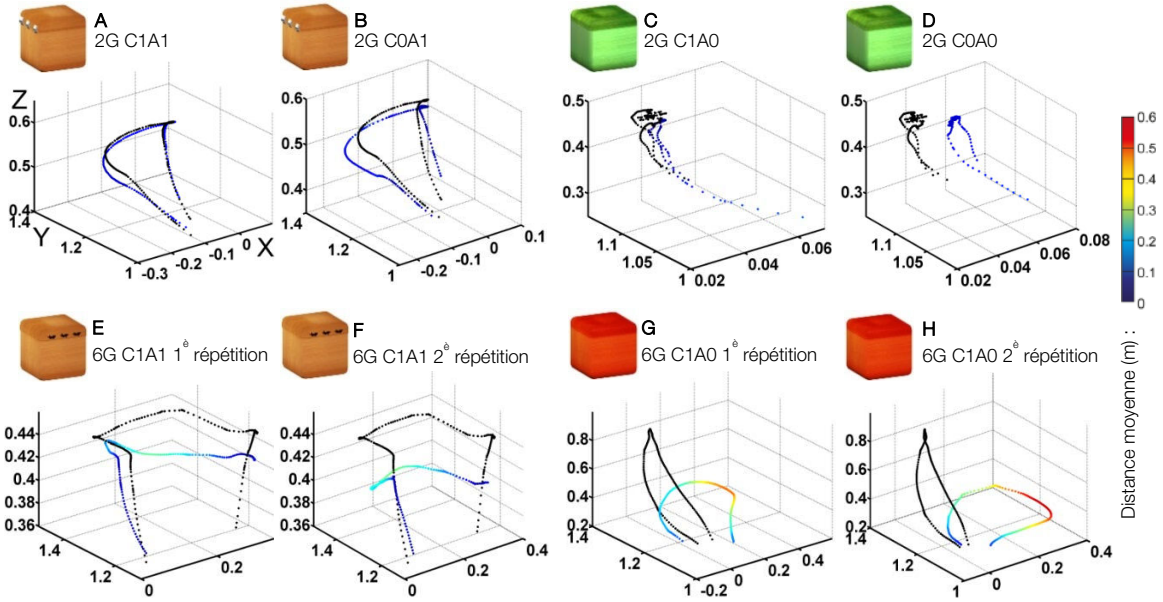


Figure 4-10 : Comparaison entre les trajectoires du geste créé et du geste reproduit à l'interaction (référentiel absolu). La courbe noire représente le geste original. La courbe colorée représente la distance à la courbe originale.

4.3.1. Résultats hypothèse A : influence de l'affordance

Les affordances explicites des objets aident à la remémoration des gestes proposés. Lorsqu'elles sont présentes, le temps de réflexion reste court quel que soit le nombre de gestes à se remémorer. L'analyse des variances entre les conditions avec affordances A1 et sans affordances explicites A0, indépendamment de la colocalisation et pour toutes les séries de gestes, se révèle significative : $F(1,20) = 33,34$; $p < 0,001$. Cela est également confirmé dans la mesure de la fidélité spatiale où l'ANOVA est significative $F(1,20) = 4,88$; $p = 0,039$ entre les distances euclidiennes moyennées.

Les graphiques A et C de la Figure 4-10 montre une remémoration correcte du geste, celui-ci étant fidèlement répliqué dans l'espace pour la condition avec affordances explicite (A1). Les graphiques E et F expriment un résultat similaire, mais la reproduction du geste est moins précise. Ceci est sans doute dû à la fatigue, voire à l'aisance ou à la simplification du geste à la fin de l'expérimentation. Les graphiques G et H montrent clairement deux réplications incorrectes du geste : l'utilisateur ne s'est pas remémoré son geste original.

4.3.2. Résultats hypothèse B : influence de la colocalisation

Nous observons de façon globale que les performances semblent indépendantes de la condition de colocalisation. Pour le temps de réflexion, la différence n'est pas significative $F(1,20) = 2,22$; $p = 0,152$, cependant l'ANOVA sur la distance euclidienne est significative $F(1,20) = 25,74$; $p < 0,001$. Ceci résulte d'un décalage spatial quasiment systématique entre

le geste original et sa reproduction (voir graphiques B et D de la Figure 4-10). Cela vient du fait qu'en manipulation non colocalisée (C0), malgré la consigne de faire le geste au même endroit qu'à la création du geste, les utilisateurs ont tendance à vouloir interagir en atteignant la boite distante. Le fait qu'elle soit inaccessible influence le geste répété et celui-ci se retrouve spatialement légèrement au-dessus du geste original. Le geste se révèle ainsi clairement décalé vers la boite, ce qui montre l'influence des contraintes physiques, même perçues par la vision et non par le toucher.

4.3.3. Résultats hypothèse C : influence du nombre de gestes

Le premier constat, les performances des utilisateurs décroissent à mesure que le nombre de gestes à se remémorer augmente, mais seulement dans la condition sans affordance explicite (A0). Cela se constate pour le temps de réflexion (voir A0, Figure 4-7), la précision spatiale (A0, Figure 4-8), et les taux de reconnaissance (A0, Figure 4-9). Secondement, nous notons que les performances ne baissent pas lorsqu'il y a présence d'affordances. Le temps de réflexion et les taux de reconnaissance montrent la même tendance pour les conditions C1A1 et C0A1 des Figure 4-7 et Figure 4-9. Les valeurs moyennes du temps de réplication des gestes avec affordances sont inférieures à 2289 ms, tandis que sans affordances, ces moyennes sont nettement supérieures à 3446 ms (voir Figure 4-7).

Sans affordance explicite (A0), les performances baissent clairement lorsqu'il y a plus de gestes à se remémorer. Cela se vérifie notamment pour le temps de réflexion et les taux de reconnaissance. Il semblerait que l'absence d'affordances explicites impacte la remémoration des gestes entre les étapes 2G et 6G, qui présentent respectivement 1 et 3 boites de couleurs. Les ANOVAs ne révèlent pas de différences significatives entre les étapes 2G et 4G, pour le temps de réflexion $F(1,20) = 0,108$; $p = 0,746$ et le taux de reconnaissance $F(1,20) = 0,63$; $p = 0,437$. Cependant, il y a une différence significative entre les étapes 4G et 6G sur le temps $F(1,20) = 8,89$; $p = 0,007$. La même tendance s'observe pour les taux de reconnaissance qui diminuent de moitié (A0, Figure 4-9) et la différence se révèle significative $F(1,20) = 52,38$; $p < 0,001$. Cela montre que l'affordance est une aide directe à la remémoration des gestes à partir de l'étape 4G où deux gestes sont à se remémorer.

4.3.4. Limites et biais possibles

Différentiation des boites et mnémoniques

Nous avons observé durant les prétests que certains utilisateurs ont tenté d'utiliser des mnémoniques pour la création de gestes au regard des boites sans affordance. En fait, dans un premier prototype, nous avions différencié les boites sans affordance par un chiffre inscrit sur le devant. Ainsi, certains utilisateurs créaient des gestes métaphoriques en se basant sur ces chiffres, en les redessinant littéralement dans les airs ou même leur première lettre (comme un « U » pour le chiffre « Un »). Cela nous a orienté vers une différenciation par des couleurs afin d'éviter des moyens mnémotechniques trop évidents. Malgré ce protocole, nous avons demandé via le questionnaire si les utilisateurs adoptaient des stratégies de mémorisation. Certains testeurs ont associé la boite rouge à un gros bouton poussoir. D'autres testeurs perçoivent une simulation qui incite à frapper la boite. Cela signifie que le choix d'un bon visuel pour suggérer un geste non signifiant reste une question ouverte.

Taux de reconnaissance

En ce qui concerne la reconnaissance par le système d'apprentissage artificiel, on pourrait penser que les taux de reconnaissance auront tendance à décroître à cause du nombre de gestes à différencier. Ceci est une limitation inhérente aux algorithmes d'apprentissage et au choix des descripteurs gestuels. Cependant, on observe clairement que le taux de reconnaissance se stabilise pour un nombre de gestes à reconnaitre supérieur à quatre en présence d'affordances visuelles (Figure 4-9). Ce taux ne varie pas au-delà de l'étape 4G. Ce résultat signifie que l'origine d'une meilleure performance provient vraisemblablement des affordances.

Pour certains utilisateurs, les taux de reconnaissance globaux sont très faibles. En effet, leurs gestes demeurent difficiles à reconnaitre car proches au niveau des descripteurs gestuels choisis (données brutes normalisées) ou de trop faible amplitude. Cela vient notamment de notre approche qui laisse une totale liberté à l'utilisateur dans le choix de ses gestes, en se réservant de les orienter dans le protocole. Cela montre aussi qu'un choix pertinent de descripteur est essentiel pour maximiser les taux de reconnaissance, mais il n'existe pas de descripteur générique idéal. Il arrive aussi que la caméra de profondeur transmette des données erronées dues à l'occultation ou à la non-détection des bras. Cependant, pour cette expérimentation, les moyennes restent représentatives malgré de grandes variances. Aussi, le décalage entre les trajectoires observé dans les conditions C0 semble avoir un impact limité sur le taux de reconnaissance (voir graphiques B et D de la Figure 4-10).

Choix des métriques de comparaison

Concernant les métriques de comparaison, précisons que le système de reconnaissance est invariant au temps (voir Figure 2-12). Afin d'avoir une confirmation plus juste des résultats, il aurait fallu de comparer les distances euclidiennes moyennes des trajectoires des gestes avec invariance temporelle (sans intégrer l'horodatage à l'échantillonnage). Les résultats obtenus restent toutefois représentatifs. Enfin, dans le but de comparer des gestes proches en positions mais différents en orientations, il aurait été envisageable de mesurer en plus les écarts en orientation comme le proposent (Moreau et al. 2008). Nous envisageons d'utiliser cette méthode dans l'expérimentation suivante.

Biais dans les conditions

Nous proposons de discuter de biais éventuels sur une telle expérimentation. Nous présentons les boites dans un ordre aléatoire tout au long de l'expérience. Cependant, la chronologie de l'expérimentation suit toujours le même ordre pour les étapes 2G, 4G et 6G, où respectivement 1, 2 et 3 gestes par conditions d'affordances sont à proposer et à remémorer. Ainsi, même après la phase d'entrainement, à l'étape 2G, certains participants demandent confirmation sur le déroulement de la tâche lorsque la boite apparait à distance. Cela donne un temps de réflexion supérieur pour les conditions C0 comme l'indique la Figure 4-7.

À propos du rôle du nombre de répétitions du même geste pendant la phase de création, nous observons que le geste se stabilise, mais qu'implique le nombre de répétitions au niveau de la mémorisation ? Nous avons demandé aux participants leur avis dans le questionnaire ; aucun ne pense que des répétitions supplémentaires auraient contribué à la mémorisation.

4.4. DISCUSSION ET CONCLUSION

Nous avons mis en place une approche qui permet aux utilisateurs de proposer leurs gestes en vue d'une interaction dédiée. L'étude conduite a pour but d'explorer le rôle des affordances visuelles et de la colocalisation sur la mémorisation des gestes. Les résultats montrent le rôle capital des affordances pour la création et la remémoration des gestes. Nous avons sélectionné six différents systèmes d'ouverture de boite avec affordances visibles : des charnières, des verrous et des couvercles avec des flèches ou des rebords. Il est envisageable qu'il existe des niveaux d'affordance qui pourraient avoir une influence sur la performance, notamment si les indices visuels sont peu explicites ou difficiles à discerner. Nous n'avons pas observé par ailleurs que des utilisateurs oublient leurs gestes dans le cas de boites avec affordance explicite.

À propos du caractère signifiant des affordances, une étude connexe dans le domaine de la création musicale été menée par (Caramiaux et al. 2011). Celle-ci propose de comparer des gestes produits par des utilisateurs à l'écoute de sons signifiants ou concrets, à des gestes produits par rapport à des sons abstraits. Les auteurs notent que les sons concrets induisent essentiellement des gestes iconiques qui miment l'action qui a produit le son. Les représentations iconiques font cependant appel à des références propres à chaque participant. Par ailleurs, quand le son n'est pas —ou trop difficilement— identifiable, le geste effectué est plutôt métaphorique. Le geste correspond ainsi à l'aspect acoustique du son et fait essentiellement référence au son lui-même. Il est clair que le caractère concret ou abstrait du son modifie le type de geste et les références sollicitées. Le geste serait basé sur des indices issus du son et des connaissances de l'utilisateur. Cela corrobore nos observations : nous avons aussi constaté que les indices (ici visuels) concrets induisent le geste.

En ce qui concerne la colocalisation, le résultat principal est qu'elle n'a pas d'effet sur la mémorisation. À noter cependant que les contraintes de la boite distante modifient la directivité du geste. Les contraintes physiques et la perception des objets ont une incidence sur la forme et le positionnent du geste. Aucun utilisateur ne semble être affecté par l'incohérence d'une manipulation d'objets virtuels sans retour haptique. Il n'y a pas eu de problèmes de compréhension apparents.

Au niveau de l'empan mnésique, il s'avère difficile de se remémorer plus de deux gestes non signifiants ou abstraits sur un vocabulaire de quatre gestes. Cela est directement lié au manque d'affordances explicites. En revanche, grâce aux affordances, les utilisateurs sont en mesure de se remémorer rapidement l'ensemble des gestes proposés. Il semble bien entendu possible de mémoriser plus de gestes, mais combien ? Les situations de la vie quotidienne suggèrent que cette remémoration de gestes pourrait être illimitée dans le cas du schème, car il fait appel à l'expérience acquise et aux connaissances de l'utilisateur. Ainsi, la limite pour une application interactive serait la capacité du système à enregistrer et différencier les gestes d'un grand vocabulaire.

Nous n'avons pas évalué l'impact du retour visuel et de la rétroaction. Dans cette expérience, nous avons choisi de ne pas fournir de retour visuel puisque nous voulions nous concentrer dans un premier temps sur le rôle des affordances visuelles liées à l'apparence de l'objet. L'intégration d'une animation visuelle comme l'ouverture progressive d'une boite couplée au geste ferait appel à un autre type d'affordances « dynamiques ». Le prolongement naturel de ce travail est l'étude de l'effet d'animations visuelles en rétroaction à l'exécution du geste.

CHAPITRE 5
EXPÉRIMENTATION SUR LA RÉPÉTABILITÉ DU GESTE ET LES RÉTROACTIONS

Dans l'expérimentation précédente, nous avons proposé l'approche UDIg basée sur une création de gestes par l'utilisateur en vue d'interagir avec différents objets. Nous nous sommes focalisés dans un premier temps sur les affordances visuelles statiques liées à l'apparence de l'objet dans la phase de création du geste. Nous avons ainsi constaté que les schèmes sont principalement induits par la perception des affordances. Dans cette première expérimentation, à la phase d'interaction par le geste, nous n'avons pas mis en place de retour visuel sur l'activité, et par conséquent, pas de rétroaction, notamment visuelle. Nous souhaitons aller plus loin dans la compréhension du geste à l'interaction et des processus mis en jeu dans un schème de manipulation. Nous proposons de nous focaliser sur le rôle de la rétroaction visuelle et la répétabilité du geste à l'interaction. L'utilisateur sera exposé à des objets animés dans la phase de création, il aura un retour visuel relatif à son geste dans la phase d'interaction. Nous souhaitons observer le comportement de l'utilisateur selon différentes conditions de rétroaction et de proposition du geste.

5.1. QUESTIONS DE RECHERCHE ET HYPOTHÈSES

Dans la première expérimentation, nous nous sommes intéressés à l'apparence de l'objet à travers les affordances « statiques » dans le but d'inspirer la création du geste. En effet, les objets ne disposaient pas d'animation suggérant leur fonctionnement ou leur évolution dans le temps. Dans la seconde partie de ce travail, nous nous focalisons sur un autre type d'affordances que nous appelons affordances « dynamiques ». Il s'agit d'une animation visuelle, telle qu'une boite qui s'ouvre progressivement pendant l'exécution d'un geste. Une simulation idéale, ou en tout cas le plus proche de la réalité en l'état actuel des techniques de simulation, demeurerait très complexe à mettre en place. Par exemple, une manipulation d'objets virtuels précise avec la main requiert un suivi des doigts à haute fréquence et un moteur physique sur des représentations géométriques précises. Les mains et l'environnement virtuel doivent être correctement colocalisés, ce qui implique un système de rendu en relief bien calibré, un retour visuel avec latence minimale, des animations avec calage, etc. Mais faut-il nécessairement une correspondance exacte avec une situation écologique ? Est-il possible d'extraire des éléments suffisamment représentatifs qui permettent à l'utilisateur d'effectuer tout de même l'activité ?

Il semble que cela dépende de l'activité désirée. Nos cas d'application peuvent nécessiter des rétroactions non écologiques ou pas nécessairement continues dans la phase d'interaction. En effet, l'activité à vocation créative ou à des fins de réadaptation peut faire appel à des rétroactions non réalistes, non écologiques ou simplifiées. Celles-ci peuvent varier du symbolique à l'écologique, être plus ou moins explicites et concrètes, entre commande simple en tout ou rien et contrôle continu de l'action. Dans le cas d'une rééducation cognitive, une tâche peut être faite de plusieurs façons : un geste peut déclencher une action sur un objet ou un déplacement, etc. Par exemple, il est envisageable d'attribuer une animation déterministe sur différents objets virtuels selon un scénario souhaité. Il est aussi possible de déclencher l'animation dès la reconnaissance du geste, sans la synchroniser avec le déroulement du geste. À l'inverse, il est envisageable de synchroniser l'animation et le geste. Dans la pratique artistique, le geste incarné de l'artiste peut déclencher un effet visuel, que l'artiste peut contrôler et moduler. Par conséquent, quel type de rétroaction doit-on donner à l'utilisateur dans l'activité ?

Au chapitre précédent, nous avons vu les avantages du schème pour l'aide à la proposition et la remémoration des gestes. Nous choisissons d'explorer le schème à l'interaction. Nous proposons de le décomposer en étudiant le rôle de différentes propriétés visuelles. De plus, il semble important de prendre en compte les préférences subjectives et la fatigue des utilisateurs à l'usage.

Par ailleurs, nous situons toujours nos travaux dans l'éventualité que les interfaces limitent les conditions d'immersion et d'interaction. En fonction de l'interfaçage et du positionnement des capteurs, les données peuvent être bruitées ou partielles. Par conséquent, les différents degrés de qualité de restitution visuelle (entre couplage, continuité et réalisme) ont-il une influence sur le comportement de l'utilisateur ?

Nous proposons d'orienter nos recherches selon deux axes :

- Le premier se focalise sur la rétroaction visuelle. À quel moment faut-il donner la rétroaction ? Quel est le rôle du couplage à l'interaction ? De quelle nature doit être la rétroaction ?
- Le deuxième axe concerne la répétabilité du geste. Nous savons que la répétition des gestes entraine une certaine fatigue. Un geste basé sur un schème, mais simplifié ou de plus faible amplitude, est-il toujours assimilable ? Enfin, quel est le rôle de la fatigue dans la répétition d'un geste ?

Afin de répondre à ces interrogations, nous proposerons plusieurs hypothèses regroupées dans les deux axes susnommés.

5.1.1. Rétroactions visuelles

Selon l'activité, différents aspects visuels de rétroactions peuvent apparaitre à l'utilisateur. Or, nous ignorons quelle rétroaction est la plus appropriée pour aider ou guider l'utilisateur dans son activité.

Nous avons vu que les affordances visuelles et la perception de celle-ci par l'utilisateur influencent la proposition du geste. En donnant un retour visuel à l'interaction, nous souhaitons savoir si celui-ci influence la reproduction du geste à l'interaction. (McGrenere et al. 2000) proposent un continuum qui lie les degrés d'affordances (la facilité d'appréhension) en fonction du degré de clarté des informations perçues (Figure 5-1). Plus l'information est explicite, plus l'affordance est facile à appréhender.

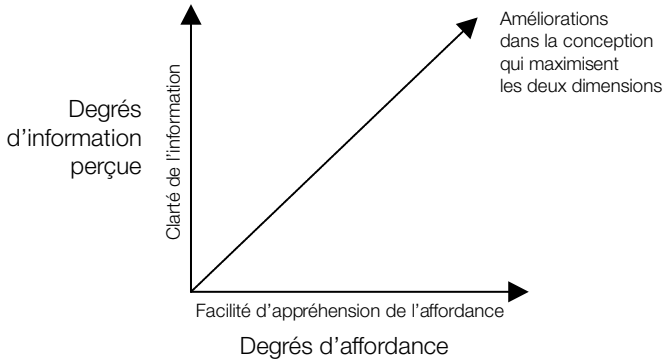


Figure 5-1 : Continuum représentant les affordances et les informations qui définissent ces affordances (McGrenere et al. 2000)

Les auteurs constatent que l'affordance d'une interface est perçue différemment selon les utilisateurs. Avec des utilisateurs novices ou experts, McGrenere et al. comparent l'usage d'interfaces en lignes de commande (*CLI*) avec celui d'interfaces graphiques (*GUI*) (voir 0 pour les définitions). Les *CLI* fournissent très peu d'information sur leurs fonctionnalités tandis que les *GUI* fournissent des informations explicites. Malgré ces informations explicites, les utilisateurs experts ayant les commandes en tête préfèrent les *CLI*, car elles se révèlent plus rapides à utiliser au clavier que d'utiliser un périphérique isotonique comme la souris. Les *CLI* leur paraissent plus « affordantes » que le *GUI* qui leur parait moins efficient en comparaison. En revanche, pour un utilisateur novice, les informations visuelles et la souris facilitent l'usage sans avoir à apprendre les commandes. Cela confirme que les affordances - qui font appel aux connaissances acquises - sont perçues subjectivement selon l'expérience de

l'utilisateur. Il semble important de prendre en compte la perception subjective des utilisateurs dans l'évaluation de l'affordance dynamique.

5.1.1.a *Hypothèse A : influence du couplage de la rétroaction sur le geste à l'interaction*

Nous proposons que l'utilisateur puisse créer un geste au regard d'un objet ayant une animation ou un comportement préétabli. À l'interaction, il est possible de synchroniser l'animation au geste d'interaction. Ceci correspond au principe de couplage préconisé par (Norman 1989b). Cette rétroaction couplée permet de donner la possibilité à l'utilisateur de comprendre son avancement dans l'activité. Dans le cas d'objet virtuel comportant une animation, lorsque l'utilisateur interagit, l'animation suit son geste. Si l'utilisateur l'interrompt, l'animation s'interrompt également. Ce couplage pose un certain nombre de problématiques d'usage que nous avons décrit au chapitre 3.2. Par conséquent, nous proposons d'explorer dans un premier temps le rôle du couplage.

Dans le cas d'une interaction par geste continu, la création d'un retour visuel en tout début de geste revient à ne pas prendre en compte le couplage. Si le résultat final de la tâche est donné au moment où le geste débute, il ne semble plus utile de coupler le geste au reste de la tâche. Nous proposons d'isoler le rôle du couplage geste et action. Nous souhaitons dans un premier temps observer si le couplage geste-animation serait perçu par l'utilisateur. Quels seront ses comportements avec ou sans couplage ? Si nous n'observons pas de différence significative avec et sans couplage, une interaction plus simple à mettre en place pourrait convenir. Par exemple, un déclenchement de l'animation par simple proximité physique de la main à l'objet.

→ **Hypothèse A** : Le couplage geste-animation est plus incitatif à exécuter une tâche. Avec ce couplage l'utilisateur effectue un geste d'interaction conforme à l'original. En revanche, s'il n'y a pas de couplage, mais un déroulement déterministe d'une animation représentant le comportement de l'objet, l'utilisateur n'est plus obligé à reproduire son geste en entier.

5.1.1.b *Hypothèse B : influence sur le geste du moment d'apparition de la rétroaction couplée*

Dans le cas d'une rétroaction couplée au geste pendant son exécution, cette rétroaction peut arriver pendant l'exécution du geste ou a posteriori. En se référant au continuum McGrenere et al. (Figure 5-1), nous estimons qu'une rétroaction présentée au cours de l'exécution du geste permet d'aider l'utilisateur à reproduire son geste. Si la rétroaction n'apparait qu'à la fin du geste, le degré d'affordance est plus faible et induit plus de variabilité dans la reproduction du geste.

→ **Hypothèse B** : Une rétroaction créée au fur et à mesure de l'exécution du geste présente des affordances qui guident l'utilisateur dans l'exécution de l'activité. Une rétroaction a posteriori ne présente des affordances qu'à la fin du geste, ce qui ne permet pas de guider le geste et induit plus de variabilité dans la reproduction du geste.

5.1.1.c *Hypothèse C : influence sur le geste des degrés d'informations perçues en continu*

Nous souhaitons comparer l'effet de différentes rétroactions dynamiques à l'apparence visuelle plus ou moins explicite. Nous souhaitions initialement faire une distinction entre le caractère imagé ou littéral du rendu visuel comme à la première expérience. Or, il semble difficile de dissocier l'aspect visuel de l'aspect dynamique dans la perception de l'affordance. En effet, la qualité de la continuité confère à l'affordance un caractère explicite différent entre statique et dynamique. Nous avons vu précédemment que la perception et l'action sont liées (Gibson 1966; Neisser 1976; Varela et al. 1991). En conséquence, il parait difficile de comparer des rétroactions dynamiques en dissociant l'apparence visuelle d'une rétroaction de sa continuité. Par exemple, une rétroaction écologique non fluide dans sa continuité paraitra plus artificielle à l'utilisateur qu'une rétroaction plus symbolique, mais fluide. Par conséquent, comme le proposent (McGrenere et al. 2000), nous préférons raisonner par degrés d'informations disponibles pendant la rétroaction. Nous estimons que plus le degré d'informations visuelles est important à la rétroaction, plus le geste sera guidé.

→ **Hypothèse C :** les gestes reproduits par l'utilisateur sont influencés par la rétroaction. Une rétroaction présentant un degré d'information élevé permet au geste d'interaction d'être guidé. À contrario, une rétroaction donnant des informations moins explicites ne permet pas un guidage du geste.

5.1.2. Répétabilité

5.1.2.a *Hypothèse D : existence d'un geste de faible amplitude basé sur un schème*

L'étude précédente a démontré que la colocalisation ne semblait pas influencer la remémoration du geste. En revanche, l'utilisateur influencé par la perception de l'objet distant modifie son geste en essayant de se rapprocher inconsciemment de l'objet. La question est de savoir si des gestes issus de schèmes, mais étant réduits en amplitude permettent une interaction équivalente à une interaction avec un geste d'amplitude réaliste.

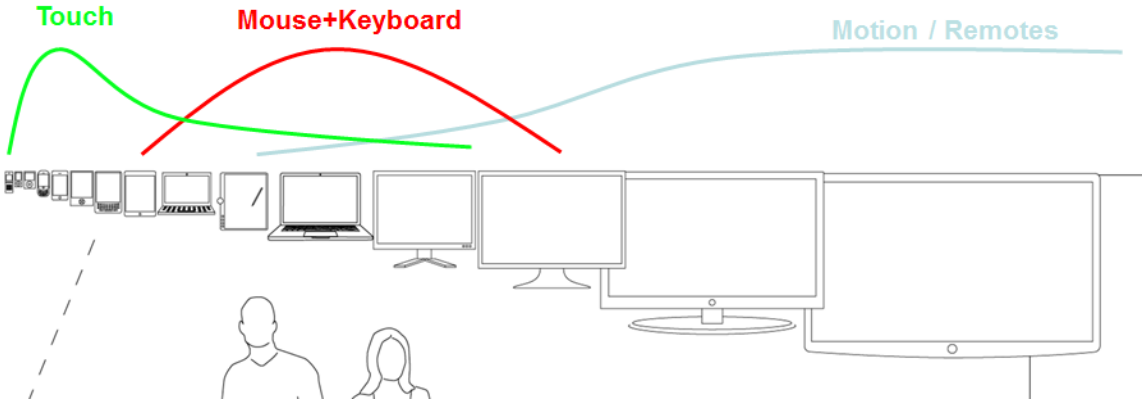


Figure 5-2 : Schéma technocentré des modalités d'interaction selon les dispositifs (Lee 2011)

(Lee 2011) note que l'échelle de l'affichage du contenu, c'est-à-dire la condition d'immersion, dimensionne le geste. En effet, l'auteur constate que les dispositifs plus immersifs peuvent obliger l'utilisateur à se déplacer physiquement ou à effectuer des gestes amples pour interagir. Ceux-ci se révèlent plus fatigants à l'usage. Par conséquent, il est possible d'avoir recours à des modalités plus commodes comme le contrôle gestuel à distance et l'usage de télécommandes (Figure 5-2). Par ailleurs, nous avons vu dans l'étude de conception inclusive de (Voida et al. 2005) que les utilisateurs privilégient la manipulation distante plutôt que de se déplacer lorsque l'écran n'est pas accessible directement.

Notons que certains jeux vidéo de la *Nintendo Wii* proposent des interactions issues de schèmes. Cela a notamment été évalué par (McMahan et al. 2010). Or, certains joueurs sont conscients que les manettes *Wiimote* utilisent des accéléromètres. Ils feignent alors un geste réduit par rapport au geste réel évoqué par l'activité ludique, mais ayant une signature d'accélération équivalente. Prenons par exemple un jeu vidéo de tennis sur console. L'utilisateur est incité à jouer debout en frappant la balle virtuelle avec un mouvement du bras de grande amplitude inspiré d'une situation écologique. Mais il est aussi possible de jouer assis en tournant simplement le poignet (Lange et al. 2011).

Cette modalité de manipulation ou d'interaction à distance peut faciliter l'usage des utilisateurs qui ne souhaitent pas ou qui ne peuvent pas interagir de façon colocalisée. Un utilisateur ayant des spécificités motrices ne peut pas toujours faire un geste de grande amplitude. Son geste peut cependant demeurer précis. Il peut par contre faire un geste équivalent à celui qu'il ferait pour une activité écologique, mais d'amplitude plus réduite. Du côté des arts numériques de la scène, un acteur peut, au cours d'une pièce, produire un geste en manipulation non colocalisée comme la télémanipulation d'un objet virtuel sur scène. Mais un geste réduit en amplitude basé sur une activité écologique reste un geste différent d'un geste écologique de grande amplitude. L'utilisateur sera-t-il en mesure de l'utiliser et de le répéter ? Cela facilite-t-il l'usage ?

Dans le cas d'une interaction gestuelle définie par l'utilisateur, il semble possible de proposer un geste réduit en amplitude pour une interaction à distance. Enfin, nous estimons qu'un geste plus réduit requiert en principe moins d'efforts pour être effectué.

→ **Hypothèse D** : un geste de faible amplitude basé sur un schème permet une interaction similaire à une interaction avec un geste d'amplitude réaliste. Ce geste qui requiert moins d'effort est moins sujet à la fatigue et subit moins de variations à la répétition et reste conforme au geste proposé à la phase de création.

5.1.2.b *Hypothèse E : influence de la répétition sur le geste à l'interaction*

On peut considérer que dans le cas d'une activité à répétition, l'homme tend naturellement à limiter la fatigue à l'usage. On s'attend par conséquent à ce que l'utilisateur produise un mouvement plus efficient, c'est-à-dire qui requiert moins d'effort pour compléter la tâche.

→ **Hypothèse E** : Toutes rétroactions confondues, au fur et à mesure des répétitions, la fatigue entrainera un geste plus efficient que le geste proposé à la phase de création et requérant un travail moindre.

5.2. L'EXPÉRIMENTATION

5.2.1. Tâche

Le participant est invité à effectuer une série de trois tâches de manipulation dans une cuisine virtuelle (Figure 5-3). Les tâches consistent à ouvrir trois objets : une porte, une boite, et un store. Dans **la phase de création** du geste, ces objets préanimés sont exposés à l'utilisateur. L'animation démarre dès que la main de l'utilisateur s'approche de l'objet. L'utilisateur doit proposer un geste pendant l'animation d'ouverture, en tâchant autant que possible, de se caler spatialement et temporellement sur l'animation. Ce geste est enregistré par le système. Puis dans **la phase d'interaction**, les objets sont présentés à l'utilisateur, et celui-ci reproduit son geste. Une rétroaction signale la complétion de l'activité.

Figure 5-3 : Cuisine réelle (gauche) et reproduction 3D temps-réel[9] (droite)

L'idée générale est de permettre à l'utilisateur de proposer des gestes issus de schèmes. Nous influençons la répétition du geste en modifiant les degrés de perception de l'affordance visuelle à l'interaction. Nous envisageons des rétroactions de types différents, donnant plus ou moins d'indices par rapport à l'animation réaliste d'origine. Au niveau des contraintes induites par l'interaction, nous souhaitons également comparer les résultats avec et sans couplage geste-animation. Nous souhaitons observer si une interaction gestuelle avec un geste faible en amplitude influence la complétion de l'activité. Enfin, nous tâcherons d'isoler le rôle de la fatigue créée par ce type d'usage.

[9] Modèle 3D de la cuisine fourni gracieusement par le laboratoire de Cyberpsychologie de l'Université du Québec en Outaouais et créé pour l'étude suivante : (Bouchard et al. 2012).

5.2.2. Protocole

L'expérimentation dure approximativement une heure. Elle comprend une phase d'équipement des capteurs et d'explication des consignes (5 min) et une session d'entrainement sur deux objets (5 min). Il y a ensuite deux sessions d'interaction correspondant aux conditions d'interactions par gestes d'Amplitude Réaliste (**AR**) et par gestes d'Amplitude Faible (**AF**) (deux fois 25 min). Il y a une courte pause entre chacune des sessions. Chacune d'elle comprend 6 sous-sessions correspondant aux 6 types de rétroactions choisies, Table 5-1.

- Pendant **la phase de création** du geste, l'utilisateur doit réaliser une série de 3 tâches à effectuer dans une cuisine virtuelle : ouvrir une porte, ouvrir une boite et ouvrir un store. Il est demandé aux participants de garder le bras reposé sur une tablette, d'approcher leur main de l'objet et de proposer un geste pour interagir. Comme mentionné au chapitre précédent (4.2.5), afin d'avoir un geste stabilisé pendant la phase de création, les participants sont invités à répéter trois fois leur geste. La dernière répétition est enregistrée par le système.
- Pendant **la phase d'interaction** par le geste, pour chaque type de rétroaction étudiée, les 3 objets sont présentés tour à tour, 5 fois chacun et dans un ordre aléatoire. L'utilisateur doit effectuer le même geste qu'il a proposé dans la phase de création du geste au regard de chaque objet. Pour chaque type de rétroaction, nous obtenons 15 gestes à effectuer. À la fin de chaque série de 15 gestes par type de rétroaction, les utilisateurs donnent leur avis sur la qualité de l'interaction perçue (questionnaire en Annexe C).

Le but de cette expérimentation n'est pas de faire une comparaison intergeste. L'apprentissage artificiel sert ici à coupler le geste de l'utilisateur à une rétroaction. Nous avons détaillé le fonctionnement de ce couplage et les problématiques au Chapitre 3.2.4.b. Par conséquent, nous utilisons, par geste, la fonction de suivi de sa complétion (voir *following* (Bevilacqua et al. 2010)). Le principe est de présenter différents types de rétroactions à l'utilisateur. Pour cela ce taux de complétion est couplé à l'animation de l'objet, de façon plus ou moins continue. Il peut aussi être représenté de façon plus symbolique pendant l'interaction. À noter que lors du mouvement du geste, nous ne donnons pas la possibilité à l'utilisateur de revenir en sens inverse sur la trajectoire de son geste. S'il souhaite refermer l'objet ou la porte au lieu de continuer à l'ouvrir, celui-ci reste figé.

Les deux variables indépendantes (2 conditions d'amplitude du geste **AR** er **AF**, et 6 rétroactions) sont croisées pour donner 12 conditions. Chaque sujet effectue les 12 conditions (*within subject design*). Pour chaque condition de rétroaction, les utilisateurs répètent 15 gestes (pour 3 objets et 5 fois par objet). Nous moyennons les résultats des mesures obtenues des gestes dans l'ordre de leur apparition à la répétition. Cela permet d'obtenir 5 mesures moyennes pour les 5 répétitions.

5.2.2.a *Choix des rétroactions*

Parmi les rétroactions envisagées dans l'hypothèse, certaines nous semblent très similaires. Afin de limiter la complexité du test, nous nous sommes arrêtés à quatre catégories de rétroaction principales (Table 5-1 : Discret, par palier, continu, continu non-couplée). Les trois premières sont des catégories de rétroactions couplées au geste. Pour la dernière dite non couplée, l'animation se déroule jusqu'à la fin, quoique fasse l'utilisateur. Parmi les trois

catégories de rétroaction couplée, la première est une rétroaction discrète. La seconde comprend deux rétroactions s'affichant par paliers (l'une symbolique et l'autre au rendu visuel plus écologique). La troisième catégorie comprend deux rétroactions continues avec un rendu évoquant l'évolution dynamique de l'objet plus proche d'une situation écologique.

À propos de la qualité des rétroactions continues, nous avons vu qu'elle peut être impactée par le découpage du geste ou la qualité de l'interfaçage (3.2.4). Nous savons que les systèmes de réalité virtuelle ont toujours une certaine latence due aux capteurs, aux traitements de l'information et aux protocoles de communication. Nous supposons qu'une plus grande tolérance du système pallie les effets de la latence. Pour ce faire, il nous est possible de paramétrer un système de reconnaissance avec un gain de tolérance à la variabilité du geste. Cette tolérance et le caractère précoce de la reconnaissance permettent d'animer l'objet avec moins de retard sur le mouvement de l'utilisateur. En revanche, une trop grande tolérance provoque un déclenchement intempestif : le taux de complétion du geste peut avancer de lui-même jusqu'à 100% malgré le peu de mouvement produit par l'utilisateur. Cela donne l'impression d'une interaction sans contrôle et sans couplage. Nous ne le souhaitons pas, car le geste et sa réalisation effective sont au cœur des usages proposés.

Nous avons vu qu'il est préférable de raisonner par degrés d'informations perçues subjectivement par l'utilisateur. Par conséquent nous proposons d'ordonnancer par degrés d'informations pouvant être perçue selon le continuum Figure 5-1. Nous avons choisi différentes rétroactions ordonnancées par degré de proximité à la situation de référence Table 5-1. Nous avons sélectionné ces 6 rétroactions :

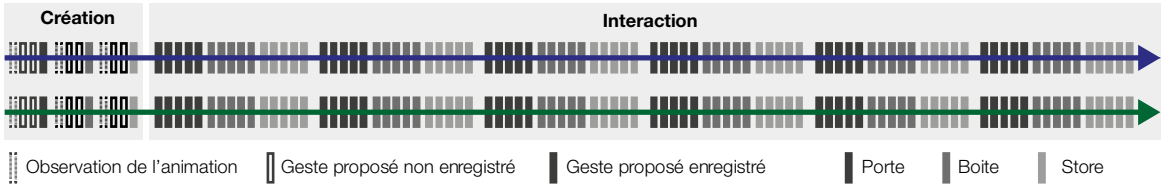
- La première rétroaction notée **Bool** est une commande discrète. L'objet reste fermé jusqu'à ce que le geste d'interaction soit complété à 85% par rapport au geste proposé à la phase de création. Lorsque ces 85% de complétion sont atteints, l'objet est présenté totalement ouvert. Cette rétroaction s'apparente plus à une interaction par commande binaire que par contrôle progressif ;
- **Pa%**, pour rétroaction par paliers exprimée en pourcentage, affiche un texte donnant le taux de complétion du geste menant à l'ouverture totale de l'objet. L'objet reste inerte, seul le texte change par palier de 20% (de 0 à 100%) ;
- **PaAn** pour rétroaction avec animation par paliers, reprend les paliers de 20% de la précédente, mais ici l'objet est animé et s'ouvre par étape ;
- **AnCo** pour Animation Couplée correspond à l'animation temps-réel couplée directement au geste ;
- **AnC+** pour Animation Couplée correspond à l'animation temps-réel couplée directement au geste avec une légère avance. Ici le système de suivi du geste est paramétré avec un gain de tolérance à la variabilité (+ 25%). Le suivi plus précoce confère à l'animation de l'objet une petite avance sur le geste. La limite dans l'utilisation de ce gain est un risque de faux positif, car le taux de complétion risque d'avancer légèrement même si l'utilisateur interrompt son geste ;
- Enfin, **TANC**, pour Témoin avec Animation Non-Couplée, correspond à la condition témoin. L'animation est lue automatiquement dès que l'utilisateur approche sa main de l'objet. Il n'y a pas de couplage geste-animation, cette dernière se fait automatiquement comme dans la phase de création du geste.

Ces rétroactions choisies sont classées par degrés d'information disponible en continu, vis-à-vis de l'animation présentée lors du geste original (hors condition témoin non couplée).

Table 5-1 : Choix des rétroactions présentées à l'utilisateur classées par dynamique et chronologie de l'expérimentation

Conditions						
	Bool Booléen	Pa% Paliers en pourcentages	PaAn Paliers Animation	AnCo Animation Couplée	AnC+ Animation Couplée + gain	TANC Témoin Animation Non Couplée
Quand ?	À la fin du geste	Pendant le geste	Pendant le geste	Pendant le geste	Pendant le geste	Au début du geste
Comment ? (continuité)	Discret	Continue par paliers de 20%		Continu		Discret
rendu visuel statique	écologique	symbolique	écologique	écologique	écologique	écologique
animation	fermé puis ouvert	(pourcentage affiché par paliers de 20%	animation affichée par paliers de 20%	animation continue	animation continue	Animation continue déclenchée par proximité à l'objet
Conformité de l'information perçue par rapport au geste créé	partiel ⟷ conforme					Témoin

Chronologie de l'expérimentation (hors entrainement) :



AR Amplitude Réaliste (bleu), **AF** Amplitude Faible (vert)

5.2.2.b *Choix des objets à manipuler*

Nous avons animé cinq objets différents pour la tâche de manipulation, dont deux objets supplémentaires pour la phase d'entrainement. Durant la phase d'entrainement, les participants doivent enregistrer deux gestes différents pour deux tâches : tourner le couvercle d'une boite pour l'ouvrir et ouvrir la porte d'un réfrigérateur. Chaque objet est présenté pendant la phase d'interaction avec deux rétroactions sélectionnées aléatoirement pour chaque condition de geste d'Amplitude Réaliste (AR) et de geste d'Amplitude Faible (AF). À noter que cette phase d'entrainement sert à vérifier que le geste est correctement reconnu et suivi par le système. Si les paramètres prédéfinis de gain ne permettent pas la détection du geste de l'utilisateur, ce gain est majoré.

Nous obtenons pour la phase d'entrainement :

- Un réfrigérateur avec animation d'ouverture de la porte. Elle induit un geste de préhension de la porte puis de traction ;
- Une boite avec animation du couvercle qui se dévisse et se pose à droite sur la table. L'utilisateur est donc incité à dévisser le couvercle et le poser à droite ;

Pour le test, trois objets sont à manipuler:

- Une boite avec charnière à droite dont le couvercle est animé. Elle induit la préhension du couvercle et une ouverture vers la droite ;
- Un store qui s'ouvre en tirant sur une corde. Il induit la préhension d'une corde et une traction vers le bas ;
- La porte dont la poignée est animée, puis s'ouvre vers l'extérieur. Elle induit trois gestes coarticulés : la préhension de la poignée, une rotation et une poussée de la porte vers l'extérieur.

À noter qu'aucune autre indication que l'aspect visuel et l'animation n'est donnée à l'utilisateur. Ceci permet d'éviter d'influencer le geste proposé. L'utilisateur garde donc une liberté de proposition de geste.

5.2.2.c *Dispositif expérimental*

L'expérimentation a lieu dans une salle immersive équipée d'un écran large rétroprojeté de 3.1m de long par 1.7m de haut, d'une résolution de 1280 par 720 pixels à 120Hz, permettant un rendu stéréoscopique actif[10]. Puisque nous souhaitons évaluer la rétroaction visuelle en temps-réel, nous avons limité au maximum la possibilité de latence du système. Les prétests indiquent que cette résolution suffit pour une perception correcte des objets et de leur animation.

Capteurs et positionnements

Afin d'obtenir une reconnaissance optimale et en continu du geste, nous avons choisi de ne pas utiliser la caméra de profondeur *Kinect* pour cette expérimentation. En effet, les occultations et les données parfois approximatives peuvent perturber la rétroaction. Nous avons orienté notre choix vers trois centrales inertielles *Movea*. Un capteur est positionné sur le bras, les deux autres sur le métacarpe et sur le pouce de la même main. Cela permet d'obtenir davantage de données sur le geste de la manipulation des objets virtuels (voir Figure 5-4). Plutôt que de calculer les coordonnées dans le référentiel du torse, nous avons choisi de les calculer dans le référentiel des bras (humérus). Les prétests ont permis de noter une meilleure invariance aux mouvements qu'un calcul dans le référentiel du torse. Mais, comparativement, cette configuration limite davantage les degrés de liberté qui décrivent le geste.

[10] Rendu avec carte graphique *nVidia Quadro 4000*

Figure 5-4 : Un utilisateur interagissant avec la porte virtuelle ; à gauche, condition geste d'Amplitude Réaliste (AR) ; à droite, condition geste d'Amplitude Faible (AF)

Descripteurs gestuels

Nous utilisons dans cette expérience les données normalisées des orientations du pouce et de la main. Nous avons constaté dans l'expérience précédente que les quaternions peuvent donner des valeurs discontinues. Ainsi, pour certaines rotations, des composantes du quaternion peuvent passer de la valeur -1 à la valeur 1. Afin d'éviter cette discontinuité pouvant nuire à un suivi en continu du geste, nous avons converti chaque axe de rotation du capteur en cosinus et sinus.

Découpe du *stroke*

Pour cette expérience, nous avons choisi de découper manuellement le geste afin de l'isoler plus précisément (Figure 5-5). Un autre avantage est de pouvoir s'assurer de la qualité des données enregistrées.

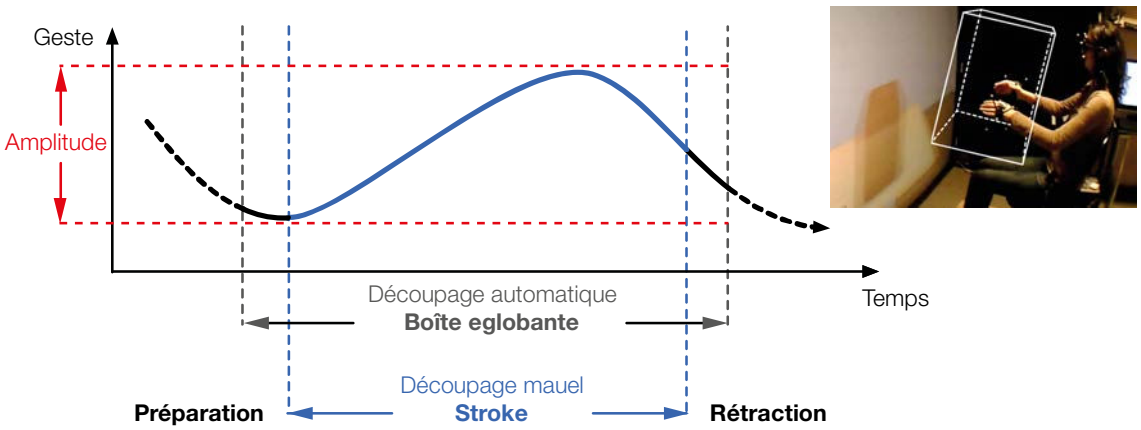


Figure 5-5: Notre approche du découpage d'une geste basé sur (Kendon 1980). À droite, un utilisateur effectuant un geste de manipulation dans une boite englobant

5.2.2.d *Conditions d'ergonomie*

L'utilisateur repose son bras sur un support situé à droite du siège. Le coude et l'avant-bras reposent ainsi sur la tablette et le poignet sur un coussin. L'utilisateur est assis en face des objets à manipuler et est incité à effectuer plutôt des gestes de grande amplitude (condition AR, voir Figure 5-4). Le bras se retrouve en général au niveau ou au-dessus du cœur selon la taille de l'utilisateur. Le but est d'induire une certaine fatigue à l'usage afin d'observer

l'évolution du geste dans le cas d'une interaction à répétition. De plus, les animations des objets d'une durée de quatre secondes obligent l'utilisateur à fournir un effort plus longtemps.

5.2.3. Métriques de comparaison

Pour cette expérimentation, nous avons évalué la rétroaction à travers quatre critères principaux :

- Nous mesurons les écarts spatiaux entre les trajectoires de la main dominante en position et en orientation. Cela correspond à la distance euclidienne moyenne entre les trajectoires enregistrées aux phases de création et d'interaction. Les trajectoires mesurées avec un système de motion capture précis[11] (le marqueur est placé sur le métacarpe) sont comparées avec l'algorithme de (Moreau et al. 2008). Toutes les trajectoires des gestes ont été échantillonnées à 200 points indépendamment de la temporalité du geste (la distance correspond à la distance vis-à-vis des points 1 à 200) ;
- Le rapport de travail fourni par le geste d'interaction sur le travail fourni à la phase de création du geste. Nous approximons le travail mécanique par la distance parcourue. C'est une approximation très rudimentaire, mais elle permet une au moins première estimation des variations. Ce travail correspond à la longueur de la trajectoire du geste dans un espace 3D euclidien calculée à partir des positions, et également à partir des orientations selon l'algorithme de (Moreau et al. 2008), 2.2.3.c ;
- Le rapport de temps d'exécution du geste d'interaction sur celui du geste original. La segmentation des gestes se fait automatiquement par une *boite englobante*.
- Le plus haut taux de complétion du geste détecté par le système de suivi du geste pendant l'interaction. La segmentation du geste lors de sa création est faite manuellement par un opérateur. À l'interaction, elle se fait automatiquement par la *boite englobante*.

5.2.4. Population

L'étude a été menée avec 20 participants (3 gauchers et 17 droitiers), âgés de 24 à 34 ans (M = 27 ; SD = 2.1) avec une distance interpupillaire moyenne de 64,6 mm (SD = 1.8 ; amplitude de 59 mm à 71 mm). L'acuité stéréoscopique de tous les participants a été mesurée au moyen du test de *Wirt*, M = 88,5 % (SD = 14 ; amplitude de 40 % à 100 %).

Dans cette expérience, nous exposons les utilisateurs à des objets virtuels dont les affordances et les contraintes visuelles incitent à les manipuler avec la main droite. Par conséquent, les capteurs ont été placés exclusivement sur le bras droit des participants. Bien que trois personnes gauchères aient participé, cela ne pose pas de problème particulier à l'interaction. En effet, nous avons observé que des personnes valides gauchères emploient habituellement leur main droite face à des objets s'utilisant par la droite (comme ici une porte ou une boite ayant leur charnière à droite, ou un store accessible directement par la main droite). Cela se révèle plus commode à l'utilisation, plutôt que le geste soit effectué avec une plus grande amplitude. Cela permet d'avoir aussi le même cadre expérimental, où les *boites englobantes* sont placées à droite de l'utilisateur pour éviter des biais éventuels dans les mesures.

[11] Deux caméras infrarouges A.R.T.II. de Advanced Real-time Tracking GmbH et cibles optiques

5.3. RÉSULTATS

Comme premier constat général, nous notons que les utilisateurs se souviennent systématiquement des gestes proposés quelles qu'en soient les conditions[12]. Cela confirme les résultats de la première étude sur le rôle de l'affordance. La Figure 5-6 détaille les résultats selon les différentes métriques et selon les conditions de geste d'Amplitude Réaliste (AR en bleu) et d'Amplitude Faible (AF en vert). Nous avons moyenné les distances euclidiennes en position et orientation pour la condition geste d'Amplitude Réaliste (AR). Les ANOVAs ne révèlent aucune différence significative ni entre les rétroactions $F(4,595) = 0,164$; $p = 0,976$ ni entre les répétitions $F(4,595) = 0,057$; $p = 0,994$. Il en est de même pour la condition de geste d'Amplitude Faible (AF) : pour les rétroactions $F(4,595) = 0,068$; $p = 0,997$ et les répétitions $F(4,595) = 0,368$; $p = 0,831$. Cela s'observe également de façon globale dans le faible écart entre les distances euclidiennes moyennes des trajectoires du geste en position (M = 0,09 m ; SD = 0,02 m), pour le ratio de travail (pour les positions et les orientations moyennées : M = 1,10 ; SD = 0,20) et les taux de complétion moyens élevés (M = 0,92 ; SD = 0,10).

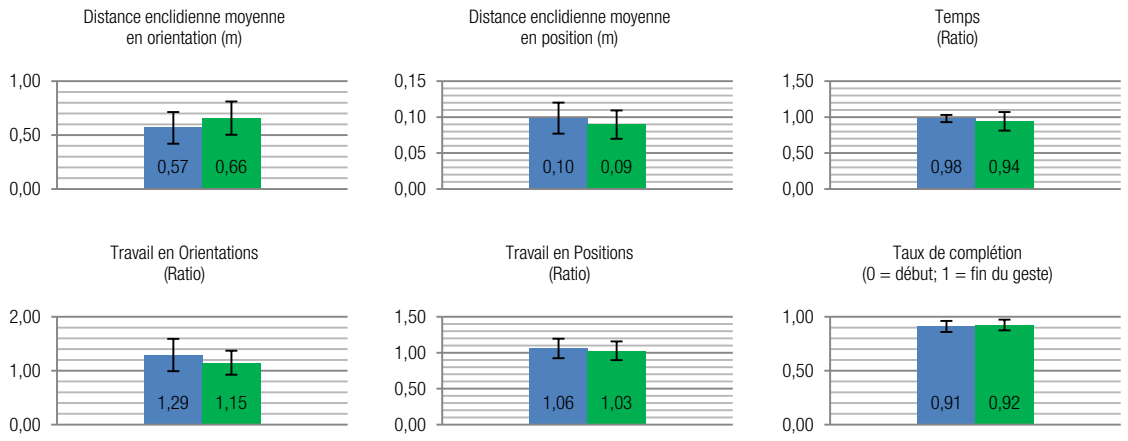


Figure 5-6 : Résultats globaux selon les métriques ; en bleu : condition geste d'Amplitude Réaliste (AR), en vert condition geste d'Amplitude Faible

Remarque : Les ANOVAs qui correspondent aux distances euclidiennes moyennées ne révèlent aucunes différences significatives. Nous n'utiliserons plus cette métrique dans les analyses suivantes. Par conséquent, nous estimerons la similitude entre le geste d'interaction par rapport au geste créé selon trois métriques :

- Le ratio de travail entre le geste de la phase de création et le geste d'interaction ;
- Le ratio de temps ;
- Le taux de complétion du geste détecté par le système de reconnaissance à la phase d'interaction par rapport au geste de la phase de création.

Pour la suite des analyses, dans un souci de lisibilité, les résultats moyennés qui concernent chaque hypothèse seront présentés au fur et à mesure. Le détail de ces résultats est consultable en Annexe C Matériel et résultats détaillés expérimentation 2.

[12] À noter que pour la condition sans couplage, les utilisateurs ont le choix de reproduire leur geste ou non, nous le détaillons dans le paragraphe 5.3.1.a

Dans le but de faciliter la compréhension de l'analyse, nous proposons de situer nos hypothèses et leur analyse sur le schéma de la Figure 5-7

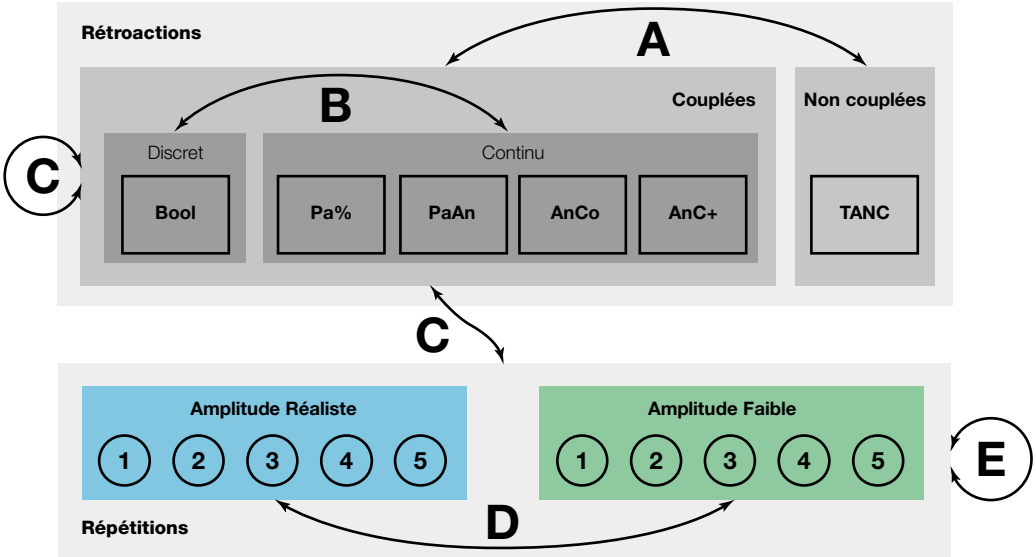


Figure 5-7 : Schéma de placement des hypothèses A à E dans l'analyse des conditions de rétroaction et de répétitions

5.3.1. Rétroactions visuelles

5.3.1.a *Résultats hypothèse A : influence du couplage dans l'interaction*

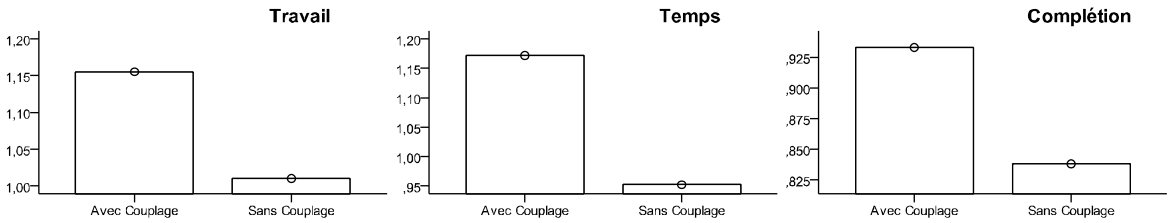


Figure 5-8 : Résultats des performances pour le Travail, le Temps et le Taux de complétion selon la condition de couplage des rétroactions

En comparant les conditions de rétroaction **Avec Couplage** (incluant toutes les rétroactions couplées) avec la condition **Sans Couplage** (qui correspond à la rétroaction **TANC**), nous observons que la plupart des utilisateurs perçoivent l'absence de couplage. En effet, **Sans Couplage**, il y a nettement moins de travail fourni, le temps d'exécution est plus court et le taux de complétion est moins élevé. Les résultats montrent qu'il y a des différences très significatives pour le travail fourni $F(1,398) = 23,02$; $p < 0,001$; pour le temps $F(1,398) = 50,71$; $p < 0,001$ et pour le taux de complétion $F(1,398) = 41,46$; $p < 0,001$. Nous avons constaté qu'après avoir avancé leur main vers l'objet, la plupart des utilisateurs remarquent que l'animation est lue automatiquement sans contrôle possible. Ils adoptent différents comportements. Certains respectent la consigne de l'expérimentation et reproduisent leur geste en se calquant sur l'animation. Certains utilisateurs attendent la fin de

l'animation sans pour autant effectuer de mouvement, puis repositionnent leur main sur la tablette. D'autres, en revanche, reposent leur main sur la tablette avant la fin de lecture de l'animation. Il en résulte un ratio temporel *gestes à l'interaction/geste créé* moyen inférieur à 1 (Figure 5-8). Chaque utilisateur étant libre de faire le geste ou d'attendre que l'animation se joue, on observe notamment de grandes variances pour le taux de complétion et le ratio de temps (détails en Annexe C). Enfin, les résultats des questionnaires quant à la qualité de l'interaction sont mitigés (Figure 5-11 page 100). Cette interaction « automatique » est notée comme très facile à utiliser.

→ L'absence de couplage est perçue par la plupart des utilisateurs. Ceux-ci adoptent un comportement différent et choisissent de reproduire le geste ou non. Le couplage geste-animation est plus incitatif à exécuter une tâche. L'hypothèse est vérifiée.

5.3.1.b *Résultats hypothèse B : influence sur le geste du moment d'apparition de la rétroaction couplée*

Pour l'analyse des résultats suivants, nous décidons d'écarter la condition témoin **TANC** (**Sans Couplage**) des ANOVA afin de comparer les rétroactions **Avec Couplage**.

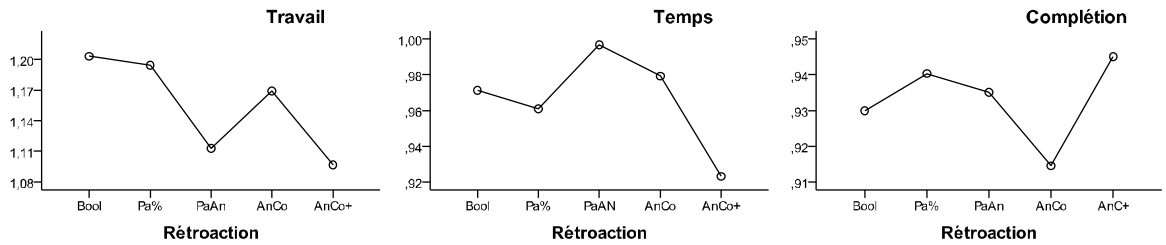


Figure 5-9 : Résultats des performances par rétroactions pour le Travail, pour le Temps et pour le Taux de complétion

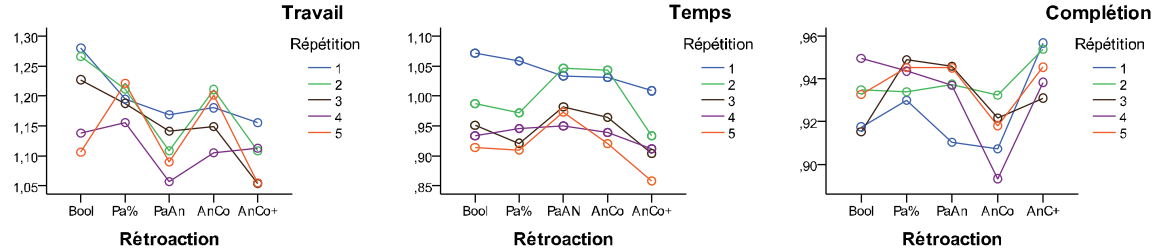


Figure 5-10 : Résultats des performances par répétitions en fonction des rétroactions pour le Travail, pour le Temps et pour le Taux de complétion

Dans la Figure 5-9 nous comparons les résultats des rétroactions **avec couplage**, toutes conditions d'amplitude et de répétition confondues. Globalement, il n'y a pas de différence significative pour le temps $F(4,995) = 1,549$; $p = 0,186$. En revanche, il y a des différences légèrement significatives pour le travail $F(4,995) = 2,726$; $p = 0,028$ et pour le taux de complétion $F(4,995) = 3,562$; $p = 0,007$. Pour le travail, le *post hoc* ne révèle pas de

différence significative, mais la probabilité la plus faible correspond à la comparaison des rétroactions **Bool** et **AnC+** (p = 0,076). Pour le taux de complétion, la différence se situe seulement entre les conditions **AnCo** et **AnC+** (p = 0,005), nous y reviendrons dans l'analyse de l'hypothèse C sur le rôle du gain de tolérance du système de suivi du geste.

À première vue, il parait difficile d'isoler l'effet de la rétroaction en continu par rapport à une rétroaction a posteriori. On s'attend à observer des courbes décroissantes pour le travail et le temps et croissantes pour le taux de complétion. On note cependant une tendance différence entre les conditions **Bool** et **AnC+** pour le travail. En observant plus en détail la répétition des gestes (Figure 5-10), on observe cette décroissance à la **1e** répétition du geste. La nature dynamique du retour visuel semble assister l'interaction.

→ L'utilisateur ayant des informations en continu effectue son geste plus rapidement avec moins de travail dès la **1e** exposition à la rétroaction (**1e** répétition). À contrario, lorsque le *feedback* apparait a posteriori (**Bool**) l'utilisateur semble effectuer un geste moins conforme à l'original qui requiert plus de travail. L'hypothèse est vérifiée.

5.3.1.c *Résultats hypothèse C : influence sur le geste des degrés d'informations perçues en continu*

Premier constat, la rétroaction a une incidence sur l'activité de l'utilisateur. Les graphiques de la Figure 5-9 montrent que pour les rétroactions **Bool** et **Pa%**, le travail fourni est supérieur aux autres rétroactions (M > 1,19). Ces rétroactions renvoient un faible degré d'informations. Par conséquent, l'utilisateur sans points de repère concrets en cours d'interaction effectue un geste plus rapide, mais avec plus de travail. À contrario, les rétroactions **PaAn** et **AnC+** requièrent moins de travail car l'utilisateur semble davantage guidé par le retour visuel dynamique et continu que par la rétroaction au degré d'informations plus faible. La rétroaction **AnCo** a demandé un travail supérieur aux rétroactions **PaAn** et **AnC+** qui ont pourtant un rendu visuel écologique similaire. De plus, la complétion est inférieur à la moyenne (M = 0,915 ; SD = 0,006). En effet, il est difficile d'atteindre un suivi du geste jusqu'à 100 % de complétion. Les utilisateurs fournissent un travail supplémentaire pour ouvrir l'objet jusqu'au maximum. En revanche, la rétroaction très similaire à **AnC+**, mais avec un gain supérieure à la variabilité du geste, requiert moins de travail et révèle un taux de complétion plus élevé (M = 0,945 ; SD = 0,006). Enfin, pour la rétroaction **AnCo**, nous avons constaté un léger retard du système. Celui-ci influence l'action gestuelle, nous y reviendrons dans le paragraphe suivant sur l'effet de la latence.

Rôle du gain de tolérance sur la rétroaction

Concernant la rétroaction **AnC+**, les tests *post hoc* ne renvoient pas de différences significatives pour le travail et le temps. En revanche, le taux de complétion est significativement plus élevé par rapport à **AnCo** (le *post hoc* renvoie p = 0,005). En effet, le gain supérieur du système permet logiquement une complétion plus rapide. Naturellement, les taux de complétion se révèlent supérieurs, car le système est plus tolérant (M = 0,945 ; SD = 0,006). À noter que par rapport à son geste original, le temps est plus court que les rétroactions précédentes (M = 0,92 ; SD = 0,026). Au niveau des résultats des questionnaires d'appréciation, les notes de la rétroaction **AnC+** sont globalement toutes

supérieures à celles des autres conditions (M = 3,9 ; SD = ,57). L'utilisateur note une facilité d'utilisation supérieure à la condition **AnCo**, mais la différence n'est pas significative U = 183, p = 0,63 (Figure 5-11). En résumé, le système étant plus réactif, l'utilisateur profite du gain et accélère son geste.

Effet de la latence

Malgré des différences en temps non significatives, nous étudions tout de même les mesures relatives au temps à la répétition. La première interrogation est de situer la latence dans notre système. Toutes rétroactions confondues, nous constatons au niveau du temps que la **1e** répétition du geste dure en moyenne 1,04 fois plus longtemps que le geste proposé au système à la phase de création. En revanche, lors de la **2e** répétition, on note que pour les rétroactions **PaAn** et **AnCo**, le temps est supérieur aux autres rétroactions, et même à celui de la **1e** répétition (voir graphique Temps Figure 5-10). Cela suggère que l'utilisateur a probablement perçu un ralentissement dès la **1e** répétition. Par conséquent, il semble avoir ralenti son geste à la **2e** répétition. Par ailleurs, lors des **3e**, **4e** et **5e** répétitions, l'utilisateur ne semble plus se soucier du phénomène de latence. La **2e** répétition servirait en quelque sorte de vérification. La consigne de l'expérimentation étant dans tous les cas d'ouvrir l'objet, une fois la rétroaction identifiée, les utilisateurs semblent continuer l'expérimentation sans se soucier du retard.

En effet, la rétroaction **PaAn** qui correspond à une animation par paliers de 20% peut être perçue comme crénelée, comme si le système « manquait de performance » pour afficher la rétroaction. Il en est de même pour la rétroaction continue **AnCo** qui, comme nos prétests le laissaient présager, pâtit d'un léger retard. Les utilisateurs ont tout de même apprécié l'usage de cette rétroaction (note globale M = 3,73 ; SD = 0,51).

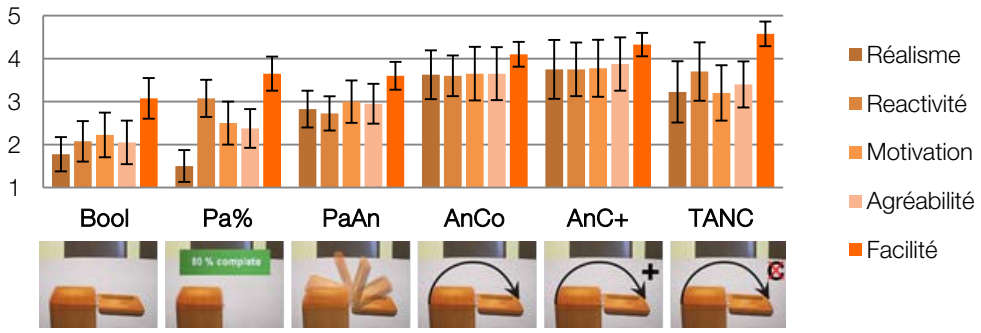


Figure 5-11 : Résultats des questionnaires subjectifs sur la qualité de l'interaction par rétroaction

Avis des utilisateurs sur les rétroactions

Au regard du critère d'appréciation, nous obtenons pour les rétroactions avec couplage :

- Rétroaction **Bool** (Booléen) : subjectivement moins appréciée et moins facile à utiliser que les autres rétroactions (note moyenne M = 2,24 ; SD = 0,47). La difficulté d'utilisation vient du fait que la rétroaction n'apparait qu'en fin de geste, laissant l'utilisateur sans « visibilité » sur son *feedback*. Comme le souligne (Norman 1989b) la visibilité est un critère essentiel à la compréhension de l'état du système et donc à l'avancement de l'activité ;

- Rétroaction **Pa%** (Palier %) : de nature symbolique (pourcentage de complétion du geste affiché littéralement). Elle est perçue comme plus réactive que la précédente (M = 3,08 ; SD = 0,43). Ici aussi, l'absence de visibilité semble être la cause de l'augmentation du travail fourni ;
- Rétroaction **PaAn** (Palier Animation) : paradoxalement, cette rétroaction a les mêmes paliers que la précédente. À noter qu'elle est perçue comme moins réactive que **Pa%** (la différence est presque significative U = 131,5 ; p = 0,058. La latence semble en être la cause. Par ailleurs, elle est perçue comme plus motivante à utiliser (M = 3,00 ; SD = 0,49). Cette rétroaction semble avoir un rôle de guide dans l'exécution du geste car le travail fourni est plus faible que la moyenne (M = 1,113 ; SD = 0,029) ;
- Rétroaction **AnCo** (Animation Couplée) : globalement les utilisateurs semblent satisfaits. Néanmoins, les moyennes indiquent que cette rétroaction est effectuée assez lentement en comparaison de l'ensemble. L'exécution du geste pour cette rétroaction demande un travail supérieur à celui fourni pour la rétroaction décrite précédemment. Au niveau des taux de complétion, on remarque qu'ils sont moins élevés (M = 0,915 ; SD = 0,006) ;
- Rétroaction **AnC+** (Animation Couplée + gain) : celle-ci est perçue comme plus réactive et bien appréciée dans l'ensemble (M = 3,9 ; SD = 0,57). On constate que le taux de complétion est plus élevé (M = 0,945 ; SD = 0,006). Ceci est expliqué par le fait que le système est paramétré avec une plus grande tolérance à la variabilité. On note un temps et un travail nécessaires nettement inférieurs aux autres rétroactions ;
- Nous ajoutons une précision sur la rétroaction témoin sans couplage **TANC**. La plupart des utilisateurs ont remarqué qu'il n'y avait pas de couplage, mais effectuaient tout de même le geste en « décalquant » l'animation comme à la phase de création. Par conséquent il y a moins de motivation à l'interaction (M = 3,2 ; SD = 0,64) qu'avec les deux rétroactions précédentes (M > 3,7). À noter que certains utilisateurs déclenchent simplement l'animation par proximité sans interagir jusqu'à la fin de l'animation. D'autres retirent rapidement leur main de la zone d'interaction une fois qu'ils ont compris le caractère automatique du déclenchement.

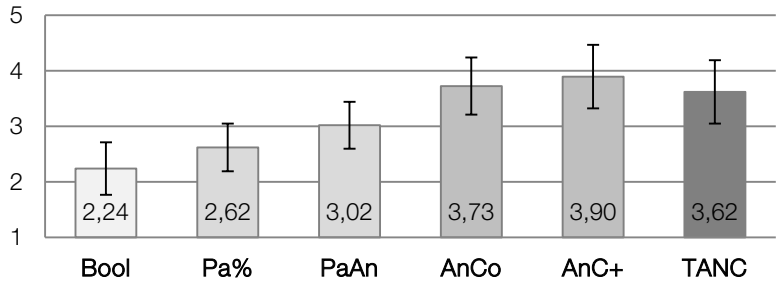


Figure 5-12 : Moyennes des résultats des questionnaires subjectifs sur la qualité de l'interaction en fonction de la rétroaction (détail des résultats sur la Figure 5-11)

Enfin, le graphique en Figure 5-12 correspond aux résultats moyennés de la Figure 5-11. À noter qu'une progression continue apparait dans l'appréciation d'utilisation. Les questionnaires montrent que la satisfaction générale augmente selon le degré de détails à l'interaction du retour visuel. À noter que l'on retrouve ce motif mais descendant pour la **1e** répétition du geste, en terme de travail et de temps (Figure 5-10). Il semblerait que l'utilisateur soit donc influencé par le caractère dynamique et concret de l'affordance dès la **1e** répétition du geste. Nous constatons que les préférences subjectives des utilisateurs suivent

les degrés d'informations perçus selon une progression croissante comparable au continuum de (McGrenere et al. 2000) décrit précédemment.

→ Le type de rétroaction influence le comportement de l'utilisateur à l'interaction, ce qui correspond à notre hypothèse. Par ailleurs, nous avons un certain nombre d'observations :

- La nature dynamique du retour visuel permet d'accompagner le geste à l'interaction. L'utilisateur ayant des informations en continu accélère son geste et diminue le travail dès la 1e exposition à la rétroaction. À contrario, un *feedback* a posteriori ne permet pas de guider le geste. Les rétroactions basées sur un faible degré d'informations fournies demandent un travail supplémentaire à l'utilisateur tandis qu'une rétroaction plus explicite semble guider le geste. Ce geste est complété en moins de temps et avec un travail moins important ;
- Un gain de tolérance du système permet une reconnaissance du geste sujet à une plus grande variabilité. Par conséquent, l'utilisateur fournit moins d'effort pour effectuer l'action et cela lui permet de compléter la tâche plus rapidement, ce qui est apprécié à l'usage ;
- Si une latence est perçue par l'utilisateur, celui-ci semble vérifier cette perception à la 2e répétition du geste. Par conséquent, il s'adapte et « attend le système ». Si le système le permet, l'utilisateur interagit plus rapidement ;
- Nous constatons enfin que les préférences subjectives des utilisateurs croissent selon l'augmentation des degrés d'informations perçus. Bien qu'il n'y ait pas de différence significative, les performances suivent cette préférence dès la 1e répétition du geste.

5.3.2. Répétabilité et fatigue

5.3.2.a *Résultats hypothèse D : existence d'un geste de faible amplitude basé sur un schème*

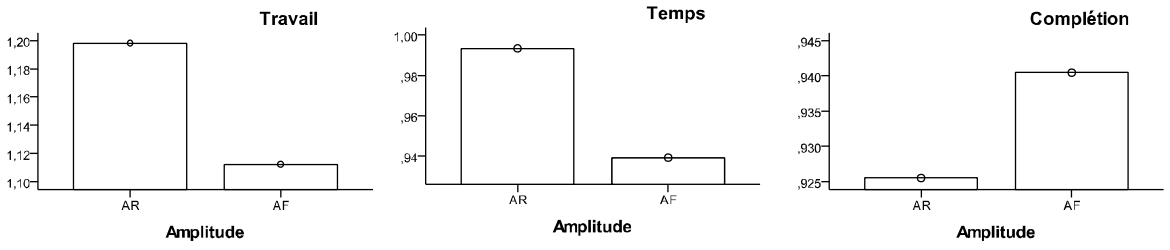


Figure 5-13 : Résultats des performances pour les conditions d'Amplitude pour le Travail, pour le Temps et pour le Taux de complétion

Premier constat, pour la condition d'**Amplitude Faible** (**AF**), les utilisateurs se remémorent leurs trois gestes et ont pu interagir de la même manière qu'avec la condition d'**Amplitude Réaliste** (**AR**). Le système est parvenu à détecter les gestes. Les taux de complétion sont en moyenne élevés (M = 94% ; SD = 0,4%). Les performances suivent globalement celles de la condition de gestes **AR**. On note une distance euclidienne moyenne faible

(M = 0,6 ; SD = 0,15) et inférieure à la condition **AR**. Cependant, on note une distance euclidienne supérieure en orientation (M = 0,09 ; SD = 0,02). Bien que l'utilisateur doive proposer un geste d'amplitude faible, il semble compenser la faible amplitude de mouvement dans l'espace par des rotations du poignet plus prononcées.

Toutes rétroactions et toutes répétitions confondues (Figure 5-13), le geste d'**Amplitude Faible** requiert significativement moins de travail F(1,998) = 10,92 ; p = 0,001. Il est exécuté plus rapidement F(1,998) = 7,56 ; p = 0,006 et est plus complet qu'un geste d'**Amplitude Réaliste** F(1,998) = 7,126 ; p =0 ,008.

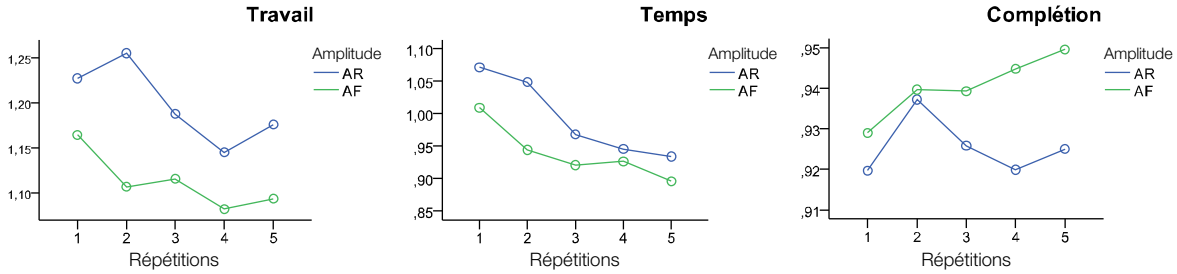


Figure 5-14 : Résultats des performances pour les conditions d'Amplitude en fonction des Répétitions pour le Travail, pour le Temps et pour le Taux de complétion

À l'usage, au fur et à mesure des répétitions (Figure 5-14), on n'observe aucune différence significative entre les deux conditions **AR** et **AF** quelles que soient les métriques : pour le travail F(1,998) = 0,384 ; p = 0,82 ; pour le temps F(1,998) = 0,530 ; p = 0,71 et pour la complétion F(1,998) = 0,610 ; p = 0,66. Ils suivent globalement les mêmes tendances. Bien que la différence ne soit pas significative, la condition **AF** semble faciliter l'usage (temps d'exécution plus court, travail moindre et taux de complétion supérieurs).

Nous avons soumis un questionnaire à chaque fin de test par rétroaction. L'avis des utilisateurs concernant les deux types de gestes a été analysé par un Test-U Mann-Whitney. L'application est considérée utilisable (note > 3 dans les deux cas : U = 171; p = 0,385). Cependant, parce que les conditions d'expérimentation contraignaient l'utilisateur à avancer son coude jusqu'à la *boite englobante*, on note un confort et une fatigue moindre, mais sans différence significative, respectivement U = 158,5; p = 0,234 et U = 161; p = 0,269. En revanche, dans l'impression de réalisme du mouvement par rapport à une situation réelle, il y a des différences significatives entre geste ample et réduit U = 91; p = 0,002, pour l'aisance U = 113,5; p = 0,016 et pour la spontanéité U = 102,5; p = 0,006.

À noter cependant que certains utilisateurs ressentent paradoxalement moins d'aisance et moins de confort dans la condition **AF**. Cela est dû aux contraintes d'expérimentation qui obligent les utilisateurs à parfois adopter une position inconfortable. En effet, il leur est imposé d'avancer leur main hors de la tablette vers la *boite englobante*, tout en gardant le coude reposé sur celle-ci. Cela n'a par ailleurs pas affecté l'utilisabilité de l'application (M = 3,85).

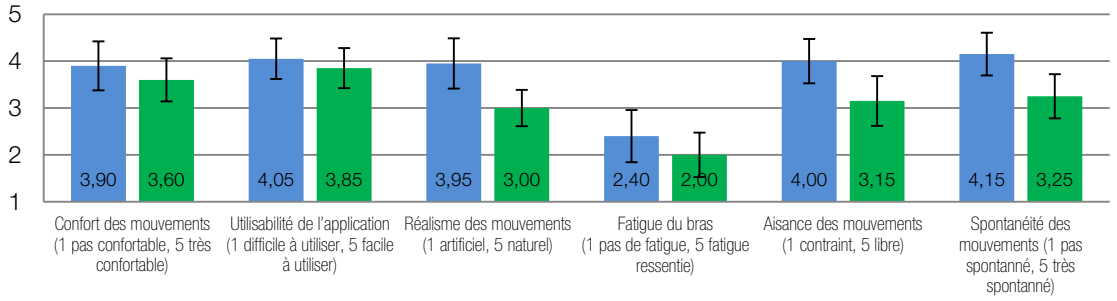


Figure 5-15 : Résultats du questionnaire subjectif d'utilisabilité entre conditions de gestes d'Amplitude Réalise (AR en bleu) et d'Amplitude Faible (AF en vert)

→ Le geste basé sur un schème mais d'amplitude faible permet d'interagir de la même façon qu'un geste d'amplitude réaliste. Il permet de fournir moins de travail et d'interagir dans un temps plus court. Les taux de complétion se révèlent même être supérieurs.

5.3.2.b *Résultats hypothèse E : influence de la répétition sur le geste à l'interaction*

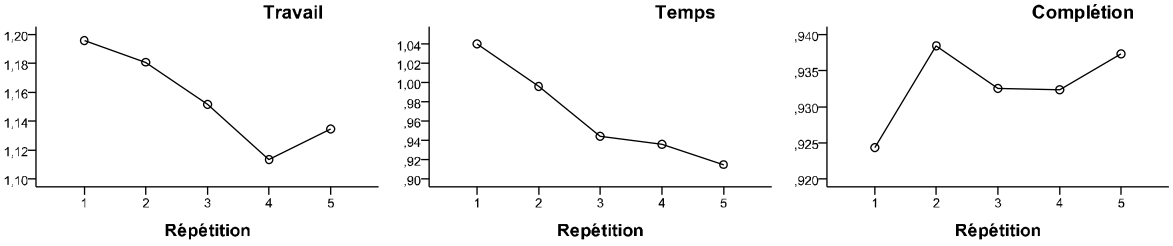


Figure 5-16 : Résultats des performances en fonction des Répétitions pour le Travail, le Temps et le Taux de complétion

Stabilité dans la répétition

Les ANOVAs ne sont pas significatives pour le travail $F(4,995) = 1,328$; $p = 0,258$ et le taux de complétion $F(4,995) = 0,796$; $p = 0,528$. En revanche, elles le sont pour le temps $F(4,995) = 5,36$; $p < 0,001$. Plus en détail, le test *post hoc* ne révèle pas différence entre la **1e** répétition et la **2e** répétition. La **2e** répétition servirait en quelque sorte à vérifier l'impression de la **1e** répétition. En revanche, il y a une différence significative entre la **1e** et les **3e**, **4e** et **5e** répétitions (le test *post hoc* renvoie respectivement $p = 0,018$; $p = 0,007$ et $p = 0,001$). En ce qui concerne la stabilité du geste lors de la répétition, on ne note aucune différence significative entre les trois dernières répétitions. Le geste se stabiliserait temporellement à la **3e** répétition. Cela rejoint les observations de prétests de nos deux expériences préconisant au minimum trois répétitions du geste à la phase de création.

Remarque sur le syndrome « bras de gorille »

En incitant l'utilisateur à interagir avec le bras au niveau ou au-dessus du cœur, nous nous attendons à observer plus de fatigue. Celle-ci modifiera probablement certaines

caractéristiques du geste. Nous avons vu que l'effet « bras de gorille » provoque une certaine fatigue à l'usage. Premier constat, dans le questionnaire les utilisateurs ne signalent pas une fatigue importante (voir fatigue du bras sur l'échelle Likert à 5 niveaux, $2 < M < 2,4$) et la différence subjective n'est pas significative $U = 161$; $p = 0,269$. Malgré le fait que l'utilisateur doive interagir avec le bras au niveau ou au-dessus du cœur, il n'y a pas de syndrome « bras de gorille ». Chaque geste d'interaction pouvant durer jusqu'à 4 secondes est répété 15 fois par type de rétroaction. Enfin, le bras en position de repos entre chaque interaction semble pallier la fatigue.

→ Contrairement à ce que l'hypothèse suggérait, il n'y a pas de baisse de l'efficience du geste à la répétition. Au contraire, la quantité de travail parait conservée afin de permettre une complétion plus grande de la tâche. La rétroaction couplée et dynamique contribuerait à faire en sorte que l'utilisateur s'oblige à terminer son action. En revanche, le geste est effectué plus rapidement. On constate une certaine accélération du geste entre sa 1e et sa 3e répétition. La 2e répétition jouerait un rôle de vérification de la perception de la réactivité du système à la 1e exposition à la rétroaction. Une fois ceci vérifié par l'utilisateur, le geste se stabilise dans sa vitesse d'exécution dès la 3e répétition.
La notion de fatigue et sa limitation ne sont pas liées au travail fourni mais au temps d'exécution du geste. Il semblerait que l'utilisateur tente de diminuer la fatigue en exécutant un geste plus rapide. La consigne étant d'ouvrir les objets, le travail ne peut être diminué pour que l'exécution se fasse.

5.3.3. Limites et biais possibles

Le peu de variation entre les résultats provient éventuellement du fait d'avoir moyenné les différents objets et répétitions. Il aurait été envisageable de ne comparer que les premières ou dernières répétitions du geste pour chaque objet et pour chaque rétroaction, mais cela démultiplie les combinaisons d'analyses. Nous avons choisi de grouper les rétroactions similaires par séries afin de recueillir plus facilement les avis. Les utilisateurs ont pour consigne générale de refaire le même geste, la découverte d'une nouvelle rétroaction n'apparait qu'à la première exposition de l'utilisateur au premier objet de cette série. Ils en déduisent que les objets suivants auront une rétroaction du même type. Présenter toutes les rétroactions de façon aléatoire plutôt que par groupe aurait éventuellement influencé le comportement des utilisateurs, mais il aurait été plus difficile de récolter leurs avis à la fin de l'expérimentation sur chaque type de rétroaction effectuée aléatoirement.

À propos des résultats globaux sur le travail moyen, il apparait en général plus élevé par rapport au geste proposé à la phase de création du geste ($M = 1,10$; $SD = 0,20$). En effet, nous constatons un travail mesuré en orientation ($M = 1,22$; $SD = 0,26$) supérieur au travail en position ($M = 1,04$; $SD = 0,13$). Nous constatons que ce travail en orientation a un ratio proche de 1. Ces résultats élevés pour le travail en orientation ne paraissent pas correspondre à un travail supplémentaire de l'utilisateur. Il semble qu'ils proviennent de la sensibilité de l'algorithme aux variations angulaires. Ils demeurent tout de même cohérents par rapport au travail fourni en position.

Le système de reconnaissance a été réglé avec une certaine tolérance pendant la phase d'entrainement afin de maximiser la reconnaissance du geste. Or cette tolérance pose des

problèmes inhérents de synchronisation geste-animation. Une tolérance insuffisante ne permet pas au système de reconnaitre le geste, tandis qu'avec un gain de tolérance trop élevée, on provoque des faux positifs. On observe lors un phénomène de glissement : le taux de complétion avance de lui-même jusqu'à 100% bien que le geste ne soit pas effectué.

De plus, les réglages de tolérance sont définis de façon globale, et non pas par geste. Il arrive dans certains cas que le système de reconnaissance ne parvienne pas à détecter le geste. Cela se traduit par une absence de rétroaction pour certains objets. Dans ce cas, l'utilisateur avait pour consigne de terminer le geste démarré et de reposer sa main au point de départ. Cela provient du fait que le geste n'a pas été correctement enregistré à la phase de création ou du choix les descripteurs gestuels Après vérification, la conversion en sinus et cosinus passe par des angles d'Euler. Les trois composantes de rotation sont calculées en cascade. Bien que le système ait toléré ces descripteurs, il serait pertinent d'envisager d'autres descriptions informatiques du geste. L'utilisateur a aussi bien pu modifier son geste en cours d'interaction par rapport au geste original, mais cette distinction ne se constate pas à l'œil nu. Il aurait été utile de prévoir dans le protocole un réenregistrement du geste, mais cela aurait aussi pu créer un biais si le geste avait évolué en cours d'expérimentation.

Aussi, nous avons fait le choix de découper le *stoke* à l'œil nu plutôt qu'automatiquement via une boite englobante comme lors de l'expérimentation précédente. Ce découpage n'étant pas parfait, il a également une influence sur le couplage geste-animation. Mais globalement, peu de décalages ont été perçus par les utilisateurs. Ce résultat est confirmé par les avis subjectifs, où les utilisateurs ont davantage préféré les rétroactions continues (AnCo et ANC+).

Enfin, l'un des objets choisi (la porte) incite à effectuer trois gestes coarticulés. Nous avons constaté qu'il est difficile de synchroniser les trois gestes coarticulés sur l'animation. Une étude indépendante permettrait de comprendre le détail des processus sous-jacents.

5.4. DISCUSSION ET CONCLUSION

Nous avons mis en place notre approche avec cette fois des objets animés. Les utilisateurs ont été en mesure de proposer des gestes au regard de l'affordance dynamique. Ces gestes ont été correctement remémorés et réutilisés à la phase d'interaction. Cela s'est vérifié à l'usage. Un vocabulaire de six gestes a été proposé vis-à-vis de trois objets différents dans une tâche de manipulation. Tous les utilisateurs ont pu se remémorer leurs gestes, que ce soit dans la condition au geste d'Amplitude Réaliste ou celle au geste d'Amplitude Faible.

Notre objectif est d'observer le comportement de l'utilisateur à la phase d'interaction. Par objet à manipuler, différents degrés d'information sont présentés à la phase d'interaction par rapport à la phase de création. Nous avons isolé les rôles du couplage et du rendu visuel des rétroactions. Par ailleurs, nous souhaitions connaitre le rôle de la répétabilité du geste sur l'interaction. Voici ce que les résultats montrent.

Les rétroactions visuelles

Dans un premier temps nous avons isolé le rôle du couplage. Celui-ci est déterminant pour inciter l'utilisateur à interagir. Sans couplage, les utilisateurs adoptent des comportements différents. Certains choisissent de reproduire le geste et d'autre non. Nous constatons principalement que le couplage geste-animation est plus incitatif à exécuter une tâche.

Concernant les rétroactions avec couplage du geste, le type de rétroaction influence le comportement de l'utilisateur lors de l'interaction. Nous notons qu'un *feedback* donné a posteriori ne permet pas de guider le geste. En effet, le manque de visibilité sur l'avancement de l'activité ne permet pas d'aider l'utilisateur à agir. Cette observation correspond à la préconisation de (Norman 1989b) sur la visibilité de l'état du système. A contrario, l'utilisateur qui perçoit un retour visuel pendant l'action répète un geste plus proche de l'original. Les rétroactions continues permettent de guider le geste tout au long de son exécution. Ce geste se révèle être complété en moins de temps et en nécessitant un travail moins important.

À propos de l'influence du degré d'informations visuelles, les rétroactions de natures symboliques demandent un travail supplémentaire à l'utilisateur. Par ailleurs, une rétroaction plus explicite donne plus de points de repère à l'exécution du geste. De plus, si l'utilisateur perçoit une certaine latence du système, il vérifie cette perception à la seconde répétition du geste. Par conséquent, c'est l'utilisateur qui s'adapte et « attend le système ». Si le système le permet, l'utilisateur interagit de plus en plus rapidement, sans pour autant fournir moins d'effort. Enfin, un gain de tolérance du système permet une reconnaissance du geste plus rapide. L'utilisateur fournit par conséquent moins d'effort pour effectuer l'action et cela lui permet de compléter la tâche plus rapidement, ce qui semble apprécié à l'usage. Nous constatons enfin que les préférences subjectives des utilisateurs croissent selon l'augmentation des degrés d'informations perçus en continu. Bien qu'il n'y ait pas de différence significative, les performances semblent suivre cette tendance dès la première répétition du geste.

En résumé, il semblerait que l'utilisateur, par essai et apprentissage, s'adapte au système afin de respecter la consigne d'ouvrir un objet. Nous avons vu qu'il effectue un geste plus rapidement quitte à fournir plus de travail pour compléter la tâche d'ouverture au maximum. En revanche, si l'utilisateur perçoit un retard du système (notamment dû à la latence), il adapte son comportement et l'attend. Cette observation rejoint les remarques de (Regan

1995) sur l'effet de la latence qui ralentit les mouvements de l'utilisateur. Enfin, si la rétroaction n'indique pas une complétion totale, l'utilisateur insistera pour ouvrir totalement l'objet en le prolongeant notamment. Cela montre la limite principale d'un couplage en continu et de sa synchronisation. Il serait envisageable d'assister l'utilisateur dans la fin du geste afin de pallier la qualité de synchronisation et de faciliter l'usage.

La répétabilité du geste

Nous avons proposé à l'utilisateur d'effectuer des gestes de faible amplitude pour interagir avec différents objets. Ceux-ci incitent à l'emploi de gestes de grande amplitude. Nous notons que ces gestes sont mémorisés et reproduits conformément au geste original. Ils demandent moins d'effort et de temps d'exécution par rapport aux gestes de grande amplitude, mais les taux de complétion se révèlent être supérieurs. Le travail supplémentaire de l'utilisateur permet de compléter l'action jusqu'à la fin.

De façon générale pour 5 répétitions d'un même geste de manipulation, la quantité de travail fourni ne semble pas diminuer. En revanche, le geste est effectué plus rapidement. Nous constatons que la seconde répétition jouerait un rôle de vérification de la perception de la réactivité du système par rapport à la première exposition à la rétroaction. Ceci étant vérifié par l'utilisateur, le geste se stabilise ensuite dans sa vitesse d'exécution dès la 3e répétition et jusqu'à la 5e répétition.

Autre remarque vis-à-vis de la fatigue : malgré des consignes d'interactions « le bras levé » pouvant entrainer une certaine fatigue, nous n'avons pas observé de syndrome « bras de gorille ». Pour un usage du geste à répétition pendant 25 min (le temps de l'expérimentation en condition Geste Ample), la fatigue sera limitée en suivant ces trois conditions :

- Effectuer un geste court (< 4 secondes) ;
- Avoir les bras partant et se reposant près du corps (comme sur un accoudoir) ;
- La main doit éviter de dépasser le niveau du cœur.

En conclusion, cette expérimentation confirme que notre approche où l'Utilisateur Définit l'Interaction gestuelle fonctionne et est efficace. Nous proposons de discuter des avantages et des limites au chapitre 6.

CHAPITRE 6
CONCLUSION ET PERSPECTIVES

Dans ce chapitre, nous proposons de résumer les principaux résultats de l'évaluation de l'approche UDIg qui permet à l'Utilisateur de Définir l'Interaction gestuelle. Ces résultats et les discussions sont décrits selon nos quatre axes principaux proposés au chapitre 3.

Nous proposerons ensuite des perspectives d'utilisation de la méthode UDIg qui ont fait l'objet de deux expérimentations hors laboratoire. Nous décrivons la présentation d'un prototype implémentant l'UDIg dans le cas d'une manifestation grand public et des tests pratiques avec des patients dans un centre de rééducation fonctionnelle. De ces tests, nous avons pu extraire des préconisations d'usage spécifiques.

Nous proposons enfin pour des travaux futurs des perspectives d'évolution de notre approche.

6.1. CONCLUSION SUR LA MISE EN PLACE DE L'APPROCHE UDIG

Ces travaux ont pour objectif de permettre à l'utilisateur d'effectuer une activité en environnement virtuel à partir d'une gestuelle non imposée. Nous avons proposé l'approche UDIg qui permet à l'Utilisateur de Définir l'Interaction gestuelle. Cette approche soulève un certain nombre de questions liées à son implémentation et son appropriation par les utilisateurs. Nous synthétisons les réponses selon quatre principaux aspects : la mémorisation du geste, sa répétabilité, le rôle de la rétroaction et des considérations sur l'utilisation de systèmes peu performants.

6.1.1. Comment faciliter la proposition et la remémoration des gestes ?

La mise en place d'une approche où l'utilisateur définit lui-même un vocabulaire de gestes pour interagir pose la question de la mémorisation dudit vocabulaire. Nous nous sommes tournés vers l'hypothèse du schème comme aide à la remémoration. En se basant sur le principe du schème, nous avons isolé deux caractéristiques propres à une manipulation d'objets en situation écologique : la colocalisation et les affordances visuelles de l'objet.

Alors que nous pensions que tous deux contribuaient à la remémoration, nos résultats montrent que les affordances visuelles sont le seul critère étudié aidant à la remémoration des gestes. Des gestes proposés au regard de l'affordance d'objets ou de l'environnement virtuel sont remémorés de façon immédiate. En revanche, sans affordance explicite, nous avons observé une confusion à partir de trois gestes à se remémorer. Le rôle de la colocalisation n'a pas d'influence sur la mémorisation mais la position de l'objet virtuel influe sur la position du geste dans l'espace. En effet, dans notre cas nous n'avons pas mis en place de retour haptique, par conséquent l'utilisateur manipule en l'air. Ses gestes sont implicitement dirigés par la vision. Un objet non colocalisé induit un geste dirigé spontanément vers cet objet. La colocalisation semble avoir une influence sur la direction du geste tandis que les affordances ont un rôle d'évocation du geste.

Par conséquent, si l'on se réfère aux situations du quotidien où l'homme utilise des schèmes et se base sur les affordances, la limite de remémoration d'un vocabulaire de gestes issus de schèmes pour interagir serait inexistante. Nous savons que l'emploi de schème n'est pas toujours envisageable en environnement virtuel. Nous avons vu qu'il est possible dans ce cas d'employer des métaphores (confère 2.1.3.b). Il nous semble important que ces métaphores aient un caractère explicite ou présentent des affordances. Celles-ci demeurent nécessaires à l'évocation du geste lors de la création du geste et de l'interaction.

6.1.2. Quel retour sur l'action et quand ?

La rétroaction est un critère essentiel pour que l'utilisateur comprenne le déroulement et l'exécution de l'activité et son avancement. Première constatation, la rétroaction impacte le comportement de l'utilisateur. L'exécution du geste est influencée par le moment d'apparition de cette rétroaction. La nature explicite de cette rétroaction est aussi un critère important. Dans nos évaluations, nous nous sommes principalement focalisés sur le rendu visuel. Nous proposons de détailler les différents moments possibles d'apparition de la rétroaction.

Une rétroaction donnée trop en amont de l'exécution d'un geste et sans contrôle possible revient à laisser la tâche s'exécuter d'elle-même. L'utilisateur n'a plus de contrôle pendant l'exécution, ceci correspond à une rétroaction sans couplage au geste. Nous avons constaté que les utilisateurs perçoivent cette absence de couplage dès la première interaction, certains la percevant immédiatement. Par conséquent, chacun est libre de continuer le geste, ou comme nous l'avons observé, de recopier dans son geste l'action présentée. Bien que cette option de rétroaction soit plus simple à implémenter, elle se révèle être peu incitative ou peu motivante à l'exécution de l'action.

Un autre moyen de signaler la complétion de l'activité est de ne donner une rétroaction qu'à la fin du geste. Nous avons observé que malgré des repères visuels explicites (mais statiques) pendant l'interaction, les utilisateurs reproduisent un geste avec plus de variabilité par rapport au geste créé. Cela provient du manque de visibilité pendant l'exécution du geste. Les utilisateurs fournissent un travail supérieur pour effectuer l'action jusqu'au bout. Ils restent cependant dans l'incertitude d'avoir correctement réutilisé un geste. Nous avons observé que dans notre cadre expérimental, les utilisateurs s'imposent de terminer leur geste. S'il ne fallait donner qu'un retour discret sur la complétion de la tâche, il serait à envisager durant l'action. Nos expérimentations ont souligné le fait de ne pas donner ce retour trop tôt, sans quoi il n'y a plus de sentiment de contrôle. S'il est effectué trop tard, l'exécution du geste laisserait l'utilisateur incertain de sa réussite. Nous estimons que le choix d'apparition d'une rétroaction discrète doit se faire en fonction de l'application et selon l'activité.

La dernière option est de proposer le retour visuel à l'utilisateur pendant l'exécution du geste. Le geste est donc couplé à l'image tout au long de l'activité. L'image en réaction permet d'accompagner l'utilisateur dans l'exécution du mouvement. Précisons néanmoins que la correspondance geste physique-action contrôlée peut être purement sémantique. En effet, il peut potentiellement ne pas y avoir de ressemblance entre les deux (par exemple un geste de rotation peut tout à fait contrôler un mouvement de translation). Dans le cas du coupage geste action, il semblerait que c'est le degré d'informations perçues au cours du geste qui guide l'action. Ce degré comporte à la fois les affordances explicites et la dynamique de l'action. En effet, trop peu d'informations disponibles à l'utilisateur correspondent à un manque de visibilité en retour sur l'action. Comme nous l'avons précisé, cela laisse l'utilisateur incertain. L'affordance dynamique guide par conséquent l'utilisateur pendant l'action. Deux critères sont à retenir :

- La temporalité, qui permet de donner suffisamment d'informations au cours du geste ;
- Le retour visuel, qui guide le geste en donnant des points de repère sur l'activité de l'utilisateur. Celui-ci interprète le rendu visuel grâce aux affordances. Même sans retour haptique, la perception visuelle de la position des objets agit à la manière d'une contrainte « physique » de l'environnement, ce qui guide l'action.

On retrouve ainsi les critères de (Norman 1989b) que nous décrivons au paragraphe 2.1.3.a sur l'usage, qui incluent la visibilité, les affordances, le couplage et les contraintes. Il ne faut pas oublier de considérer le « syndrome de l'immersion » lors d'une interaction avec couplage.

Concernant le sentiment de contrôle devant être perçu par l'utilisateur, il semble intéressant de se référer au *flow* (Csíkszentmihályi 1990) pour comprendre son activité. Dans notre deuxième expérimentation, l'objectif — ou challenge — est d'ouvrir un objet. Une rétroaction peu explicite crée potentiellement de l'anxiété : l'utilisateur est dans l'incertitude

sur l'exécution de l'activité. En revanche, une situation avec rétroaction non couplée au geste peut créer un certain ennui. La question est de savoir si la situation de couplage avec rétroaction couplée et continue induit un sentiment de *flow* chez l'utilisateur. Le couplage et le contrôle sont-ils des prérequis au *flow* ? La question reste ouverte.

6.1.3. Comment permettre la répétabilité du geste ?

Le geste, par son caractère propre et singulier (McNeill 1992), ne peut être répété à l'identique. La condition physique ou émotionnelle de l'utilisateur peut l'influencer. Outre le fait que le geste doit être remémoré par l'utilisateur, sa reproductibilité doit être prise en compte par le système. Par conséquent, il demeure important de conférer au système d'apprentissage artificiel une certaine tolérance à la variabilité du geste.

Nous avons isolé le rôle de la fatigue sur la répétition du geste. L'utilisateur tend naturellement à limiter sa fatigue, et notamment lors des activités à répétition. Dans notre évaluation, nous nous attendions à des gestes d'interaction répétés avec une efficience de plus en plus grande. Le travail s'en trouverait diminué. Mais dans notre cadre expérimental, un travail moindre ne permet pas de compléter totalement la tâche. En effet, nous utilisons des descripteurs du mouvement liés au travail du geste (trajectoires). Nous constatons que les utilisateurs respectent la consigne d'ouvrir l'objet à son maximum, notamment lorsque la rétroaction est couplée au geste. Afin de pallier la fatigue, ils effectuent des gestes plus rapides, mais ne diminuent pas pour autant le travail mécanique fourni. Il semblerait que l'utilisateur explore la tolérance du système, et tente d'en exploiter les limites. Cela se constate dès la première interaction par le geste. Si l'utilisateur perçoit une modification du système (retard ou latence), il vérifie son impression à la 2e répétition et attend le système si nécessaire. Bien que notre approche prenne en compte directement l'adaptation, l'utilisateur tend à optimiser son geste vis-à-vis de la tolérance du système. Cela se constate notamment dans la vitesse d'exécution du geste, notre système étant tolérant aux modifications temporelles du geste.

Concernant la stabilité du geste à l'usage, plus ce degré d'information disponible à l'utilisateur est élevé, plus l'affordance est explicite et plus le geste est guidé. La quantité d'informations perçues est nécessaire pour accompagner l'exécution du geste, mais elle est aussi dépendante de la qualité. Par exemple, nous avons noté les effets de la latence ou d'un couplage geste-image asynchrone. Ceux-ci induisent une perception du retard du système qui modifie le comportement de l'utilisateur. Nous en discutons dans le paragraphe 6.1.4. sur les considérations d'usage de systèmes peu performants.

Nous avons comparé l'usage de gestes d'amplitude réaliste par rapport à des gestes d'amplitude faible pour une même tâche. L'utilisateur avait la possibilité d'effectuer des gestes de faible amplitude pour interagir avec différents objets qui requièrent initialement des gestes amples ou colocalisés. Pour cela, il pouvait interagir de sa position de repos et proposer un geste du poignet ou de la main, plutôt que de lever le bras. Bien que ces gestes basés sur des schèmes aient une contrainte d'amplitude faible, ils sont mémorisés et reproduits sans difficulté. Cela confirme notre observation sur l'absence de rôle de la colocalisation sur la remémoration du geste. Ce type de geste de « téléguidage » de l'action requiert moins d'effort et de temps pour être exécuté par rapport aux gestes de grande amplitude pour une même action. Ils permettent par ailleurs une meilleure complétion.

À propos de l'effet « bras de gorille », malgré des consignes d'interactions avec le bras levé, les utilisateurs n'ont pas signalé de fatigue particulière. Selon nos observations, nous pouvons affirmer que la fatigue pour un usage répété pendant 25 minutes du geste ne sera pas trop prononcée si elle respecte trois critères :

- Le geste demeure court dans sa durée (< 4 secondes) ;
- La main ne se situe pas plus haut que le cœur ;
- L'avant-bras se repose au niveau du buste entre deux actions. Cela évite des mouvements de trop grande amplitude pour atteindre les objets virtuels.

6.1.4. Considérations sur les technologies peu performantes

Dans la mise en place d'un système de réalité virtuelle, nous avons vu que le premier critère à considérer est la transparence de l'interfaçage (Fuchs, Burkhardt, et al. 2006; Norman 1989a). Le choix du capteur est déterminant selon ce que l'on souhaite capter du geste. Le geste peut être décrit de façon qualitative ou quantitative. Toute partie du corps peut être envisagée pour l'interaction. Nous nous sommes focalisés dans ces travaux sur les gestes du bras et de la main, essentiellement pour la manipulation d'objets virtuels. Par conséquent, nous avons expérimenté avec différents capteurs à bas coût : une caméra de profondeur (*Microsoft Kinect*) et des centrales inertielles sans fil (*Movea*).

L'idée générale est d'essayer d'acquérir un maximum d'informations signifiantes et exploitables décrivant le geste. Il est primordial d'éviter de gêner l'utilisateur pour qu'il puisse se focaliser sur son activité. Les capteurs basés sur la vision ou les caméras de profondeur ont l'avantage de permettre une interaction sans équiper l'utilisateur de capteurs. Elles pâtissent cependant de quelques limitations de précision et d'acquisition, notamment liée aux occultations. Les centrales inertielles, à condition d'être peu intrusives, se révèlent être une bonne option. Nous avons vu que le mouvement humain peut, entre autres, être représenté par des orientations. En fonction de la quantité de données et de leur qualité, il est possible de combiner ces capteurs selon le niveau de détails de la description du geste souhaité (par exemple les degrés de liberté). Des techniques de filtrage et de seuillage permettent de pallier le bruit, mais peuvent induire une certaine latence. Là aussi, selon l'activité, un équilibre est à trouver entre précision et réactivité.

La latence et la synchronisation geste-image sont deux limitations dans l'interaction. À propos de la latence, celle-ci est perçue par l'utilisateur et notamment pendant l'interaction avec une rétroaction continue. Dans le cas d'une activité avec répétition d'une même tâche, l'utilisateur vérifie le retard ressenti à la seconde répétition du geste d'interaction. Dans ce cas l'utilisateur attend le système. Une fois le problème et son effet identifiés, si cela ne nuit pas particulièrement à la complétion de l'activité, l'utilisateur ne semble plus s'en soucier pour les répétitions suivantes. Dans le cas d'une rétroaction asynchrone avec un retard, celle-ci est perçue comme une latence du système. Là aussi, l'utilisateur semble attendre le système, puis une fois le problème identifié, continue d'agir malgré le retard. Dans le cas inverse d'une avance du système, nous avons observé pendant les tests que l'utilisateur tend à rattraper le retard en accélérant son mouvement. Et, comme noté précédemment, l'utilisateur tend à accélérer son geste pour limiter la fatigue. Il semblerait qu'une légère avance ne nuise pas à l'activité. Au contraire, elle permet même d'assister l'utilisateur dans la fin de l'exécution de la tâche. Une légère avance à la rétroaction permettrait éventuellement de pallier à la fois la latence et les défauts de synchronisation geste-image. Bien entendu, cela doit se faire dans une

certaine mesure afin de permettre à l'utilisateur de conserver un certain contrôle, sans quoi l'action sera considérée comme décorrélée du geste. Enfin, cette dernière hypothèse demande une évaluation.

Comme discuté précédemment, bien que nous proposions une solution sur mesure à l'utilisateur, celui-ci s'adapte au système. Nos expérimentations ont permis d'apprécier la tolérance du système et de prendre en compte la variabilité de l'activité de l'utilisateur. Nous avons ainsi défini des critères de tolérance du système :

- L'invariance spatiale : l'utilisateur peut effectuer un mouvement dans le référentiel du corps ou dans celui de l'environnement. Il est à noter, que le référentiel choisit doit s'accorder à celui utilisé pour la description des gestes.
- L'invariance temporelle : le mouvement peut aussi bien être effectué lentement ou rapidement ;
- Selon le descripteur gestuel, le système doit avoir une certaine tolérance à la variabilité du geste (que ce soit le travail mécanique, force, accélération, des descriptions plus qualitatives, etc.).

Enfin, nous avons vu que le critère essentiel pour la remémoration et la répétabilité du geste sont les affordances et la visibilité. Par conséquent, un système à bas coût de suivi de point de vue pourrait être envisageable, permettant une perception active de l'environnement. Il faut cependant prendre en compte qu'un système de suivi de point de vue peu performant peut éventuellement provoquer des incohérences sensorimotrices. Une autre possibilité serait de présenter à l'utilisateur un environnement et des objets « actifs », ou dynamiques, afin de l'aider à comprendre l'activité. Il est possible d'exposer l'utilisateur préalablement à l'animation des objets. Le caractère dynamique de l'objet animé fournit des informations supplémentaires qui explicitent davantage l'affordance de l'objet.

6.2. PERSPECTIVES D'USAGE HORS LABORATOIRE ET LIMITES

Ces travaux de recherche ont pour vocation de permettre une interaction gestuelle par gestes non imposés au regard de deux cas d'application : la rééducation à domicile et les activités de création artistique. Les résultats obtenus en laboratoire ont donné un certain nombre de pistes sur l'usage possible et l'appropriation du concept par les utilisateurs. La question est de savoir si un déploiement est possible hors d'un cadre expérimental. Le concept nous paraissant suffisamment mûr et la technologie accessible, nous avons entrepris de mettre l'approche UDIg entre les mains d'utilisateurs finaux potentiels. Cela permet de vérifier en pratique nos préconisations, mais également d'avoir des retours sur l'immédiateté, l'accessibilité, et l'appropriation de l'approche ainsi que la transparence et la robustesse de la technologie. Pour cela, nous avons mis en place deux prototypes :

- Un démonstrateur à destination du grand public a été présenté lors de la manifestation annuelle Futur en Seine 2013 (au Cent-Quatre, à Paris). Celui-ci a permis de tester l'approche et son acceptation par des utilisateurs. Une activité écologique et une activité plus abstraite ont été présentées aux visiteurs. Ces derniers ont eu la possibilité de proposer des gestes et de les réutiliser ensuite pour interagir. Nous avons fait plus de deux cents démonstrations et une vingtaine de visiteurs ont utilisé eux-mêmes le système permettant une cinquantaine de mesures ;
- Un second prototype, mis en place au sein du département ergothérapie-neurologie du centre de rééducation fonctionnelle de Kerpape (Ploemeur, Morbihan). Huit patients ont testé le prototype. Quatre d'entre eux ont été en mesure de comprendre l'environnement virtuel et d'y agir.

Nous détaillons les conditions des expérimentations et nos observations sur l'acception et l'usage de notre approche. Nous détaillerons enfin quelques limites et préconisations d'usage supplémentaires issues de ces tests pratiques.

6.2.1. Prototype « Haut la main ! » à destination du grand public

Nous avons exposé et fait tester au grand public notre approche lors d'une manifestation Futur en Seine 2013. Nos objectifs principaux étaient :

- L'appréhension par le grand public et l'enregistrement de gestes ;
- L'observation de la nature des gestes proposés pour des situations explicites de manipulation faisant appel à des schèmes, et deux situations incitant plutôt à des gestes plus métaphoriques :
    - un déplacement en virtuel ;
    - le contrôle d'un personnage abstrait représenté par un nuage de vapeur.
- La vérification de notre intuition sur une légère avance de la rétroaction qui faciliterait l'usage.

Table 6-1 : Protocole avec deux activités et deux types de retours visuels à l'interaction

	Tâche	Phase de création du geste	Phase d'Interaction	Phase d'Interaction avec gain
1 action	Ouvrir une boite	Animation	Geste et animation synchronisés	Geste et animation synchronisés avance de 10%
1 séquence de 3 actions en série	Ouvrir une porte	Animation	Geste et animation synchronisés	Geste et animation synchronisés avance de 10%
	Déplacement		Geste complété à 85% déclenche animation	Geste complété à 85% déclenche animation
	Ouvrir un store		Geste et animation synchronisés	Geste et animation synchronisés avance de 10%
1 action artistique	Moduler le mouvement d'un personnage	Animation de saut périlleux avec rendu abstrait (particules)	Geste et animation synchronisés	Geste et animation synchronisés avance de 10%

6.2.1.a *L'appréhension de l'approche par les usagers*

Globalement, l'approche et le démonstrateur ont été rapidement pris en main par les utilisateurs. Par rapport à un contexte expérimental, nous avons souhaité rendre l'utilisation plus immédiate. Dans la phase de création du geste, la tâche est simplement présentée à l'utilisateur (lecture de l'animation de chaque tâche). L'utilisateur propose ensuite un geste pendant la lecture de l'animation. Nous enregistrons la seconde répétition du geste — nous avons constaté que cela suffisait à obtenir un geste suffisamment stable. Au besoin, nous demandons aux utilisateurs des répétitions supplémentaires afin d'enregistrer un geste plus stable. À noter qu'il arrive parfois que le système ne reconnaisse pas le geste (moins d'une fois sur dix), dans ce cas la tolérance du système est ajustée. Si cela ne suffit pas, un réapprentissage est effectué. Cela nous semble essentiellement dépendant de la qualité du capteur ou de l'environnement (dans le cas de systèmes magnétiques) et du choix des descripteurs gestuels et de la segmentation du geste. La découpe du *stroke* se faisant à l'œil nu

non pas par l'utilisateur, mais par l'opérateur. Bien que le découpage ne soit pas parfait, il permettait une interaction fonctionnelle.

Figure 6-1 : Utilisateurs du démonstrateur grand public (Crédit photo : Futur en Seine 2013)

6.2.1.b *Nature des gestes proposés*

Premier constat, les gestes ont tous été correctement remémorés. Les objets concrets à manipuler ont induit des gestes issus de schèmes. Il y a clairement des consensus sur les gestes de manipulation des objets, l'affordance explicite jouant probablement son rôle. Certains utilisateurs ont cependant proposé des gestes plus originaux dans le but de tester l'apprentissage par le système.

Concernant le déplacement, nous n'avons pas constaté de problème de remémoration malgré la nature souvent métaphorique des gestes proposés. En effet, nous avons observé différentes propositions de geste et d'interprétations de la rétroaction par les utilisateurs. À la phase de création du geste, les utilisateurs sont exposés à un déplacement en virtuel de la porte au store d'une durée de 8 secondes. La consigne, qui leur est donnée, est de proposer un geste pour provoquer ce mouvement. Contrairement aux tâches de manipulation, ils ne sont pas informés que le déplacement sera initié par une commande (discrète). Nous avons laissé le choix du geste aux utilisateurs. Certaines personnes proposent de marcher sur place, ce qui évoque des gestes issus d'un schème. Nous avons aussi observé des gestes plus métaphoriques comme un pointage discret vers le store (déictique) ou indiquer, voire mimer, la direction avec la main en continu (métaphore d'un survol du sol, de la porte au store, à la manière d'un planeur). Pour ce cas-ci, nous avons choisi d'isoler seulement le début du geste pointant le store. Nous avons observé des gestes en boucle, comme un pédalage dans les airs, dans ce cas nous isolions une boucle. Ainsi à la phase d'interaction, nous avons paramétré une reconnaissance discrète du geste (à partir de 85% de complétion). Les utilisateurs font leur propre interprétation de la rétroaction. Certains répètent le geste en boucle le temps du déplacement. D'autres effectuent le geste de pointage et gardent le bras levé le temps de l'animation. Certains participants, une fois l'animation démarrée, ont compris le caractère discret de la commande de déplacement et reposent leur bras. Enfin, certaines personnes miment le survol dans la pièce. Il n'y a clairement pas de consensus.

Concernant l'application créative, l'utilisateur est exposé à un avatar au rendu abstrait (particules) effectuant un saut périlleux. L'animation est exposée au préalable à la phase de création du geste. Les utilisateurs proposent des gestes de nature plutôt mimétique même si l'avatar n'est pas toujours explicite. On constate cependant une influence de la dynamique et de la direction du mouvement de l'effet visuel sur le geste. À l'interaction, les utilisateurs ont la possibilité de moduler l'effet en ralentissant leur mouvement. Enfin, certaines personnes proposent des gestes différents à l'interaction afin de faire réagir le système et d'influencer l'ordre d'exécution de l'animation et de conférer à l'entité virtuelle un comportement moins déterministe.

6.2.1.c *Rétroaction en avance*

Dans la phase d'interaction, nous faisons répéter deux fois la séquence de tâches. Dans la première, les rétroactions sont directement couplées au geste, ce qui — nous l'avons vu — peut poser des problèmes de synchronisation geste-animation et rendre difficile la fin de l'exécution de la tâche. Nous avons proposé une seconde rétroaction en majorant le taux de complétion du geste de 10%. Cela confère à l'animation une légère avance lorsque l'utilisateur débute son geste dans la boite englobante. Concrètement, cela correspond à une discontinuité au début de l'animation par rapport au geste, la boite s'ouvrant par exemple directement à un angle faible. Les gestes d'interaction étant assez courts (inférieurs à trois secondes), cette légère avance au début de geste n'a pas toujours été perçue au début de l'interaction. Dans les cas perçus, il semblerait que certains utilisateurs doutent du caractère interactif. Pour d'autres, elle se révèle incitative à interagir, l'objet « signalant » qu'il est prêt à être manipulé. Une possibilité serait de donner progressivement une légère avance pendant l'action ou seulement à la fin, en exploitant des données de vitesse ou d'accélération du mouvement comme impulsion physique pour terminer l'action (cela renvoie au concept de physique naïve dans l'interaction (Jacob et al. 2008)). Il est demandé au participant, à la fin de l'usage du prototype, s'ils ont perçu une différence entre les deux sessions d'interaction. La majorité des utilisateurs percevaient une plus grande facilité d'utilisation. D'autres trouvaient le système presque trop réactif.

6.2.2. Prototype proposé aux patients du centre de rééducation fonctionnelle de Kerpape

Le centre de rééducation fonctionnelle de Kerpape nous a offert possibilité de faire des premiers tests avec notre prototype auprès de patients du centre qui le souhaitaient. Cette rencontre a permis de déployer le système in situ et de répondre à plusieurs objectifs :

- L'appréhension du système par les patients ;
- L'observation de la nature des gestes proposés pour des situations explicites de manipulation faisant appel à des schèmes et un déplacement en environnement virtuel. Observer si une séquence d'action peut être effectuée dans sa totalité ;
- Le début d'une réflexion sur les considérations techniques sur le déploiement in situ avec du matériel à bas coût et en détectant des gestes de faible amplitude.

6.2.2.a *L'appréhension de l'approche par les usagers*

L'application proposée permet d'interagir au moyen de gestes définis par l'utilisateur. Il est possible de manipuler des objets d'une cuisine virtuelle (une boite, une porte, un store) et de s'y déplacer de la porte au store. La Table 6-2 décrit le protocole. Tout d'abord, il s'agit de familiariser le patient avec l'approche UDIg avec une seule tâche : ouvrir une boite. À l'interaction, deux rétroactions différentes sont proposées. La première est de nature discrète qui s'active lorsque le geste d'interaction est complété aux deux tiers par rapport au geste original. La seconde est un couplage littéral geste-image sans gain ni avance. Si le patient souhaite continuer, il lui est proposé de définir et de réutiliser ses gestes dans une séquence d'actions avec deux ouvertures d'objet et un déplacement.

Table 6-2 : Protocole avec deux activités et deux types de retours visuels à l'interaction

	Tâche	Phase de création du geste	Phase d'Interaction par déclenchement	Phase d'interaction avec couplage
1 action	Ouvrir une boite	Animation de l'objet	Le geste complété à 66% déclenche l'animation	Geste et animation synchronisés
1 séquence (3 actions en série)	Ouvrir une porte Déplacement Ouvrir un store	Animation de l'objet	Le geste complété à 66% déclenche l'animation	Geste et animation synchronisés (sauf déplacement)

Dans la phase de création, le patient observe une animation : ouverture d'un objet (boite, porte, store) ou déplacement dans la cuisine de la porte au store. Il propose ensuite un geste qu'il répétera au moins trois fois. Seul le dernier geste est enregistré. Dans une deuxième phase, le patient peut réutiliser son geste pour interagir avec l'objet et se déplacer.

Pour ce prototype, comme nous souhaitons simplifier le déploiement, nous avons mis en place un dispositif minimaliste avec écran de taille moyenne (22 pouces), sans relief ni suivi de point de vue. En revanche, les objets sont préanimés et présentés aux patients avant qu'ils ne proposent leurs gestes. Les affordances dynamiques permettent d'apprécier le fonctionnement de l'objet, ses contraintes ou sa temporalité. Dans le cas où la seule animation présentée des objets, notamment la boite, ne permet pas la compréhension de l'activité en virtuel, nous avons apporté une boite réelle qui correspond à celle modélisée en 3D. Le patient peut au besoin s'y référer, l'observer ou la manipuler en situation réelle.

Les participants

Plusieurs rencontres avec des patients ont été planifiées durant une journée. Au total, sept personnes ont accepté de tester le système. Cinq personnes ont pu comprendre et interagir avec le système. En revanche, une personne parmi les cinq ne pouvait bouger que très peu les doigts de la main droite. Si elle a pu tester le système, ses gestes, d'amplitude trop faible, n'ont pu être détectés et captés. Deux personnes n'ont pas pu l'utiliser : la première souhaitait impérativement toucher l'écran et ne semblait pas comprendre le fait de réaliser un geste à distance. La deuxième ne concevait pas d'exécuter un geste sans objet en main. Un

ergothérapeute a tenté de lui montrer la boite en bois réelle et de lui faire pratiquer le geste d'ouverture sur cette boite. Malgré cet entrainement, cela ne lui a pas permis de mieux comprendre l'utilisation de geste de contrôle à distance. L'ergothérapeute souligne que le patient présente les signes d'une apraxie ne lui permettant pas de comprendre et d'effectuer un mouvement dans le vide sans objet tangible. Une huitième personne atteinte de myoclonies a souhaité voir le système, mais son état de fatigue provoquant des crises n'a pas pu lui permettre de l'essayer.

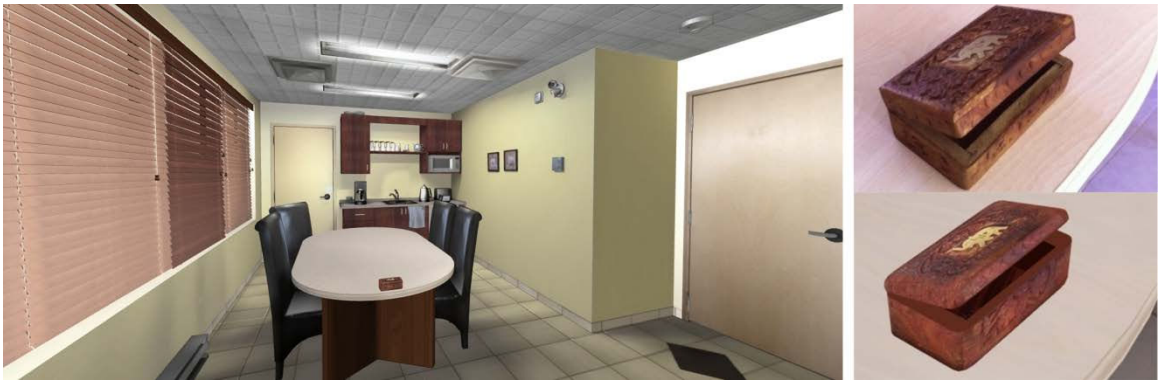


Figure 6-2 : A gauche, cuisine virtuelle ; en haut à droite, boite de thé réelle ; en bas à droite, boite modélisée en 3D

6.2.2.b *Utilisation par les patients*

Activité inspirée d'une situation écologique

Les patients sont invités à faire le geste qui correspond à l'objet, les affordances présentées sur l'objet animé incitent à effectuer un geste basé sur un schème. Cependant, même si le système est capable d'enregistrer n'importe quels gestes, certains patients tiennent à effectuer un geste de type schème, en dépit de leurs difficultés motrices. Ceci demande un effort supplémentaire à l'enregistrement, et complexifie l'interaction. Il aurait été possible de leur proposer d'effectuer des gestes plus simples, voire symboliques, mais cela rentre a priori en contradiction avec les exercices de rééducation fonctionnelle qu'ils réalisent au cours de leur période en hôpital.

Concernant la reconnaissance, les quatre patients qui ont pu interagir rapidement avec la boite ont également pu faire la séquence d'action. Celle-ci a été correctement comprise et les gestes répétés sont corrects. À noter qu'ils ont essayé le système avec le bras droit puis le bras gauche, les deux bras n'ayant parfois pas la même motricité. Pour ces deux latéralités, l'approche a été comprise et le système a globalement reconnu les gestes.

L'environnement affiché est modélisé de façon réaliste. Un environnement simplifié incitant à proposer des gestes même symboliques aurait peut-être permis aux patients de mieux appréhender le concept « un geste = une action ». Cette dernière hypothèse requerrait une étude à part entière.

Rétroaction et couplage geste-action

Première remarque sur la perception de l'environnement virtuel : dans ce déploiement nous n'avons pas mis en place de suivi de point de vue. Les patients qui ont pu percevoir et comprendre les tâches ont été en mesure d'interagir avec les objets virtuels. Cela montre encore l'importance d'affordances explicites afin d'inspirer et de guider et le geste.

À propos du couplage vision et action, les systèmes de reconnaissance de geste ont un problème inhérent de latence, le geste ne pouvant commencer à être détecté qu'après un minimum de mouvement. S'il y a trop de décalage, le patient recommence le geste de façon erronée ou exagérée par rapport au geste original. Les objets animés choisis permettent des mouvements réversibles : la boite, la porte et le store avaient seulement une animation d'ouverture. Or, les patients souhaitent pouvoir les fermer. Le choix de n'enregistrer qu'un geste pour l'ouverture de l'objet dans le protocole pose ainsi un problème de compréhension pour l'utilisateur. Enregistrer un autre geste spécifique à la fermeture complexifie la reconnaissance intergeste par le système. La problématique de la réversibilité de l'action reste encore à étudier.

Concernant le couplage geste-animation, les patients ont la possibilité de contrôler l'ouverture des objets. Or, en se basant sur une situation écologique, le seul objet permettant logiquement une ouverture partielle est le store (il est plus rare de pouvoir ouvrir partiellement une boite ou une porte). Ce couplage a quand même été testé pour les différents objets, mais cela peut poser des problèmes de compréhension. De plus, les tremblements même filtrés par le système peuvent se traduire en animation saccadée, troublant ainsi l'utilisateur. Ce type de rétroaction est donc à utiliser pour des cas précis et en fonction des besoins en réadaptation.

Nous voyons ici les limites à donner trop d'informations à la rétroaction. Cela est aussi lié au « syndrome de l'immersion ». En effet, trop de visibilité et un couplage trop littéral peuvent perturber l'utilisateur. Le compromis entre degré d'informations explicites dynamiques et le moment d'apparition de ces informations restent un problème ouvert.

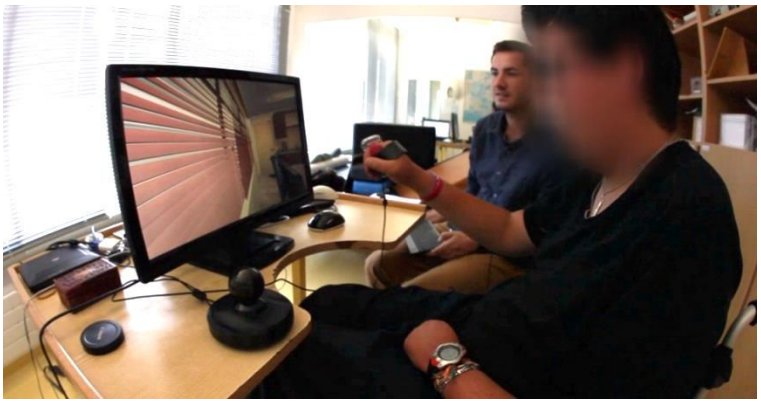


Figure 6-3 : Un patient interagissant dans la cuisine virtuelle

Répétabilité et colocalisation

Il a été constaté que la distance à l'écran avait une incidence sur les gestes proposés. Les premiers patients ont été installés relativement proches de l'écran, le mettant ainsi à portée de

bras. Les patients ont par conséquent effectué des gestes au plus proche de l'image (précisons qu'il n'y avait pas de relief), un patient touchant littéralement l'écran pour interagir malgré les consignes. Pour les patients suivants, l'écran a été délibérément éloigné du fauteuil, les incitants à manipuler à distance, ce qui se révéla fonctionnel. Hormis le patient cité précédemment qui ne peut assimiler l'interaction gestuelle à distance, les autres patients ont su proposer et reproduire leurs gestes à distance.

Afin de faciliter l'exécution de la tâche ou donner des ponts de repère, la question se pose de l'emploi d'une main virtuelle autour de l'objet à manipuler, mais cela demande des évaluations complémentaires.

Mémorisation des gestes

Globalement, les patients capables de proposer des gestes sont en mesure de les reproduire. Parmi les quatre patients ayant réussi les tâches, l'un d'entre eux présente des troubles de la mémoire, y compris à court terme. Paradoxalement, il a pu proposer des gestes et les reproduire aux deux phases d'interaction. En revanche, quelques minutes après avoir utilisé le système, il ne s'est pas souvenu d'avoir ouvert une boite. L'ergothérapeute en l'aidant a pu lui faire remémorer qu'il a visité une cuisine, mais les interactions avec les objets ont été oubliées.

6.2.2.c *Considérations technologiques avec du matériel peu performant*

Les différents capteurs qui équipent les patients ont été présentés. Il s'agit d'un brassard contenant un capteur magnétique *Razer Hydra* et un *MotionPod Movea* positionné sur le bras ; un bracelet contenant un capteur magnétique *Razer Hydra* modifié et un *MotionPod* s'installant sur la main ; et d'une bague (un *MotionPod*) positionnée sur un ou deux doigts (index et majeur ou pouce). La reconnaissance est effectuée seulement avec les *Razer Hydra*. Les capteurs *MotionPod* permettent une capture des données plus précise en vue d'une analyse a posteriori. Dans le but de reconnaitre des mouvements limités en amplitude, nous avons mis en place une amplification des données du mouvement par une normalisation dont les extremums sont définis par l'amplitude du geste que propose l'utilisateur.

À propos de la gestion des gestes non intentionnels, nous avions mis en place dans le cadre de nos expérimentations laboratoire une segmentation par boite englobante prédéfinie. Étant donné que l'utilisateur propose un geste, nous pouvons exploiter l'amplitude du mouvement pour définir une boite englobante sur mesure qui correspond au découpage du *stroke*. Une fois celle-ci définie à la phase de création du geste, quand la main quitte cette boite englobante à l'interaction, le système de reconnaissance se met automatiquement en pause afin de pallier le « syndrome de l'immersion ». Nous conseillons cependant de majorer le volume de cette boite englobante afin de prendre en compte la variabilité spatiale du geste à l'interaction.

À propos de la segmentation des gestes à l'enregistrement, celle-ci a été faire à l'œil nu par l'opérateur, mais cela ne se révèle pas toujours optimal. Cependant, l'objectif premier de ces essais était d'observer comment le système était appréhendé par des utilisateurs. La méthode de reconnaissance utilisée s'avéra satisfaisante et fonctionnelle pour les quatre patients. L'interaction de commande, où le geste reconnu déclenche l'animation, s'avéra plus simple à comprendre et à utiliser.

6.2.2.d *Perspectives d'usage en rééducation*

Ces tests pratiques avec des patients ont été d'une grande utilité pour isoler les avantages et limites d'un système d'apprentissage artificiel, intégrant des possibilités d'interaction sur mesure. L'approche UDIg s'avéra utilisable pour quatre patients, mais comme décrit précédemment, plusieurs limitations subsistent, qu'elles soient sensorimotrices ou cognitives, autant dans la phase d'enregistrement qu'à la phase d'interaction. La reconnaissance du geste reste à améliorer, notamment dans le cas de gestes de très faible amplitude. Toutefois, ces mesures et les résultats permettent dès à présent de mettre en place un premier cahier des charges afin d'améliorer l'approche pour de futures expériences.

Concernant l'adaptation, de nombreuses questions restent ouvertes, notamment comment adapter le système en fonction de la progression du patient à très court terme ? Un patient présentant des myoclonies pouvait avoir une crise à tout moment. La prise en compte de ces conditions sensorimotrices, et côté informatique, une amélioration significative de la caractérisation et de la reconnaissance des mouvements requiert un travail de recherche à part entière.

Le système a été conçu pour s'adapter autant que possible au patient, avec possibilité d'ajuster la tolérance de la reconnaissance. Par ailleurs, il est tout à fait envisageable de l'adapter aux besoins du thérapeute afin d'inciter le patient à effectuer des gestes particuliers, plus précis et à des fins de rééducation. Cela existe notamment en rééducation motrice. (Dukes et al. 2013) proposent un *serious game* afin d'aider à la récupération motrice du bras suite à un accident vasculaire cérébral. L'activité consiste à viser du poing les canards d'un jeu de fête foraine par un geste basé sur le schème du coup de poing. L'extension du bras virtuel peut être réglée par différents gains. Cela permet d'inciter l'utilisateur à dépasser progressivement ses limites et rejoint le concept de *flow* décrit précédemment (2.1.3.a).

Enfin, il est possible d'imaginer que l'apprentissage à cette interaction gestuelle puisse être en soi un exercice thérapeutique primaire avant de faire des activités thérapeutiques en environnement virtuel (comme l'application AGATHE (Klinger et al. 2013) notamment déployée au centre de rééducation de Kerpape).

6.3. Récapitulatif des préconisations d'usage de l'approche où l'Utilisateur Définit l'Interaction gestuelle

La mise en œuvre de l'approche UDIg hors laboratoire a permis de vérifier les concepts proposés. Les retours des utilisateurs permettent de réels apports et de nouvelles réflexions. Nous proposons de récapituler nos observations et nos préconisations sur la Figure 6-4. À noter que, si les flèches donnent un sens de lecture général, les interactions et les processus mis en jeu sont bien plus complexes. Nous proposons une explication de ces critères en commençant par les connaissances sur le comportement humain.

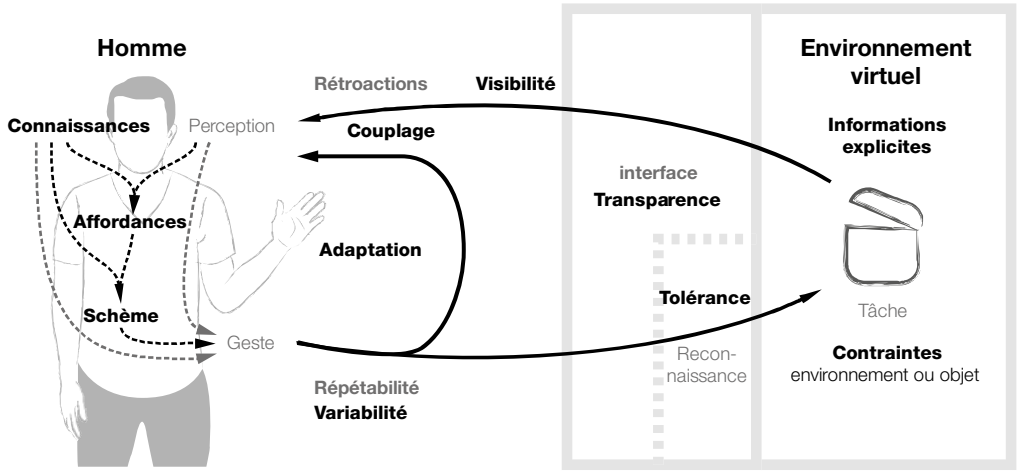


Figure 6-4 : Préconisations pour l'approche où l'Utilisateur Définit l'Interaction gestuelle (les flèches donnent un sens de lecture général)

La mémorisation et le rappel des connaissances

Nous avons vu qu'il est possible de faire appel aux connaissances de l'utilisateur dans la proposition du geste. Afin d'inspirer le geste, nous avons vérifié le rôle capital de l'affordance dans l'évocation du schème. Le schème permet à l'utilisateur d'adopter un comportement similaire dans une nouvelle situation en se basant sur ses connaissances et les affordances de cette situation. Le schème inspire le geste. Nous avons aussi observé empiriquement que le geste peut être issu de connaissances propres (comme mimer une situation imaginée ou improviser des gestes en relation à une boite de couleurs) ou lié simplement à la perception (mimer le mouvement perçu de particules d'une entité virtuelle abstraite). Le geste produit par l'utilisateur porte un sens ou une intention. Sa singularité implique qu'il ne soit pas reproductible à l'identique. Cependant, puisque la répétabilité du geste est un critère essentiel dans le fonctionnement de notre approche, nous allons détailler comment la prendre en compte.

La répétabilité du geste

Le geste, qui dépend de la condition de l'utilisateur, peut varier dans sa répétabilité. L'homme ayant des facultés d'adaptation, l'environnement et sa perception peuvent par

conséquent influer ses gestes. Il demeure important de conférer au système de reconnaissance du geste une certaine tolérance à la variabilité du geste. Cette tolérance contribue d'ailleurs à la souplesse et la transparence de l'interfaçage. Le geste reconnu par le système va modifier l'état du système. Ce nouvel état sera perçu par l'utilisateur grâce aux rétroactions.

Perception de la rétroaction

Le premier prérequis pour interpréter une action est de percevoir correctement la situation virtuelle. Outre un interfaçage de qualité suffisante, l'environnement virtuel doit donner un minimum d'informations explicites dans le but de susciter les affordances. Les contraintes de l'environnement permettent simplement de connaitre les limites de l'utilisation (par exemple les limites physiques de manipulation comme l'amplitude d'ouverture d'un objet). Ces éléments doivent être visibles, mais ce critère est souvent oublié. La transparence et la qualité de l'interfaçage influencent la visibilité. Un système de *headtracking* peut par exemple s'avérer utile pour permettre une perception correcte de l'environnement, mais cela ne semble pas obligatoire. En effet, des patients ont été en mesure d'utiliser des objets virtuels interactifs sans *headtracking*. Nos expériences laissent supposer que le caractère dynamique et donc actif de l'objet présenté pallie les possibilités d'exploration active. Cela amène de nouvelles problématiques qu'il serait intéressant d'explorer.

Nous avons vu que la rétroaction influence la perception de l'utilisateur. En effet, lors de l'interaction, de multiples rétroactions entrent en jeu. Outre le *feedback* visuel de l'environnement, l'utilisateur a comme nous l'avons vu un *feedback* intrinsèque (proprioceptif, kinesthésique, etc.). Nous avons constaté que dans le cas d'un couplage entre le geste et la rétroaction visuelle, si l'utilisateur perçoit une incohérence comme un défaut de synchronisation ou une latence, il tâche de l'identifier et attend le système. Une fois le phénomène interprété, l'utilisateur semble s'adapter en autorégulant son activité tout en exploitant les limites de tolérance du système pour notamment limiter sa fatigue. Ceci rejoint les considérations sur la mise en place d'interfaces énactives décrites en section 2.1.3.a.

En ce qui concerne le couplage, si l'environnement ou l'entité virtuelle ne fournit pas de contraintes, l'utilisateur semble se les imposer. Nous avons constaté dans l'expérimentation du chapitre 4 que des gestes inventés au regard de boites qui demeurent fermées, se limitaient à une certaine amplitude d'ouverture. Bien qu'aucun indice, mis à part la charnière, ne décrive l'amplitude d'ouverture de la boite, l'utilisateur s'impose des contraintes. On peut supposer que celles-ci correspondent à des contraintes de confort physique, ou bien qu'elles reflètent l'expérience propre à l'utilisateur.

Pour conclure, nous proposons un tableau récapitulatif des préconisations d'usage présentées par axe et par chapitre :

Table 6-3 : Récapitulation des préconisations de l'approche où l'Utilisateur Définit l'Interaction gestuelle (les chiffres renvoient aux chapitres correspondants)

UDIg 4, 5 : mise en œuvre en laboratoire 6 : mise en œuvre hors laboratoire 4, 5, 6 : appréhension par les utilisateurs	
La répétabilité 4 : stabilisation du geste par quatre répétitions 4, 5, 6 : affordances 4, 5 : colocalisation 4, 5 : visibilité 5 : geste de faible amplitude 5 : « syndrome bras de gorille » 5 : adaptation de l'utilisateur	La mémorisation 4 : empan mnésique 4, 5 : schème 4, 5, 6 : affordances 4, 5 : visibilité 5 : geste de faible amplitude
La rétroaction 5, 6 : choix de la rétroaction 5, 6 : couplage ou non 5, 6 : qualité du couplage	Les considérations technologiques 4, 5, 6 : « syndrome de l'immersion » 4 : filtrage 4, 5, 6 : normalisation 4, 5, 6 : invariance aux déplacements 4, 5, 6 : situations sans colocalisation 5 : latence 5 : synchronisation geste-animation

6.4. PERSPECTIVES

La proposition de l'approche UDIg, ses validations expérimentales et pratiques ont permis de connaitre ses avantages et ses limites. Nous proposons quelques perspectives d'amélioration et d'évolution du concept, qui peut aussi s'étendre à d'autres domaines d'application.

Une activité complexe peut nécessiter un grand vocabulaire de gestes. Dans notre première expérimentation, nous avons remarqué qu'il demeure difficile de différencier les gestes notamment par l'interprétation des descripteurs ou les coarticulations possibles. Dans notre seconde expérimentation, nous avons obligé l'utilisateur à n'effectuer qu'une tâche à la fois, la segmentation des gestes se faisant par contexte sur un objet imposé. Or, comme le relatent (Bobillier-Chaumon et al. 2005), une activité se constitue de flexibilité, d'opportunisme et est improvisée. Elle émerge de la dynamique des interactions avec l'environnement et dépend de la situation donnée. Dans le cas d'un protocole de rééducation, l'utilisateur peut effectuer une séquence d'actions dans un ordre aléatoire. La pratique artistique peut aussi nécessiter de laisser place à l'improvisation ou de ne pas suivre un scénario préétabli. Il est possible d'établir des zones d'interaction, contextuelles ou par objet (à la manière de boites englobantes) qui nécessiterait un déplacement virtuel ou réel selon l'installation ou la mise en scène. Il est également envisageable de détecter l'intention par des moyens plus avancés, comme le suivi du regard « activant » l'objet observé. Ces solutions pratiques peuvent être implémentées, mais une réelle détection d'intention reste un domaine d'étude à part entière.

Nos travaux de recherche se sont focalisés sur des gestes de manipulation. Nous avons choisi pour l'interaction et l'analyse des descripteurs gestuels qui semblent représentatifs des activités. Bien entendu, il serait intéressant de se pencher sur d'autres types de gestes et d'activités. Par exemple (Billinghurst et al. 2011) proposent de considérer d'autres usages que la manipulation directe ou la prise en compte de gestes simplement déictiques ou symboliques pour l'interaction. Le continuum de Kendon (Kendon 1988) montre que l'usage du geste peut s'étendre de l'interface gestuelle à l'interface multimodale jusqu'à une interface conversationnelle. Il semble nécessaire de pouvoir croiser les descriptions à la fois quantitatives et qualitatives des gestes pour exploiter toute la richesse et l'expressivité du geste dans l'interaction.

Bien que l'idée initiale de l'approche UDIg soit d'éviter à l'utilisateur une adaptation quelconque, nous avons constaté que l'utilisateur ajuste son comportement et s'adapte naturellement à l'approche. L'adaptation du geste de l'utilisateur à l'interaction pourrait provenir de la liberté dans son choix de geste. En effet, il est possible que le geste choisi se révèle contraignant, fatigant ou complexe à l'usage, même s'il est issu d'un schème. Cela provient notamment du fait d'absence de retour haptique qui ne permet pas de reposer ou de guider le geste. Dans ce cas, comment prendre en compte cette adaptation et tâcher de faciliter davantage cette activité ? Afin de répondre à cette question, nous avons notamment invité l'utilisateur à proposer des gestes de « télémanipulation » ou simplifiés pour interagir et nous avons observé que ces gestes sont moins sujets à variation. Ils ont été correctement réutilisés par les utilisateurs. Une tout autre approche serait de faire en sorte que le système se réadapte selon les choix de l'utilisateur ou selon certains paramètres à l'usage.

Nous savons que la condition de l'utilisateur peut évoluer au cours de l'activité, ses préférences d'usage également (comme la facilité, la rapidité ou la précision voulues). Il serait par ailleurs intéressant de faire en sorte que le système puisse détecter si l'utilisateur souhaite

interagir, par une gestion automatique de l'embrayage et du débrayage. Un tel mécanisme permettrait par exemple de laisser place à l'improvisation. Cela renvoie à l'idée d'une interaction adaptée ou adaptative, qui situe respectivement le contrôle de l'adaptation côté utilisateur ou côté système (Bobillier-Chaumon et al. 2005). Comme le système apprend de l'utilisateur, il est possible d'imaginer que l'UDIg permette un réapprentissage « à la volée ». Le cas d'une modification du système par lui-même renvoie à la notion d'autonomie du système, et son couplage à l'homme au principe d'énaction. Ceci ouvre des pistes de recherches intéressantes.

La mise en place de l'approche UDIg hors laboratoire montre que le geste semble une modalité appropriée à nos cas d'application. Certains usagers peuvent éventuellement souhaiter d'autres modalités d'interaction : il semble tout à fait envisageable d'étendre l'UDIg à d'autres modalités, voire à la multimodalité.

Enfin, nous nous sommes appuyés sur deux cas d'application (la thérapie à domicile et les performances scéniques numériques), mais cette approche peut tout à fait concerner d'autres domaines d'application plus ludiques ou à des fins éducatives. Il reste à tirer le meilleur parti de ces technologies prometteuses pour permettre une interaction plus proche de l'homme.

BIBLIOGRAPHIE

Adams, Jack A. 1971. "A Closed-Loop Theory of Motor Learning." *Journal of motor behavior*.

Alani, Tarik, & Hadj Guellif. 1994. "Modeles de Markov Caches: Théorie et Techniques de Base: Partie I."

Baddeley, Alan. 2003. "Working Memory: Looking Back and Looking Forward." *Nature Reviews Neuroscience* 4(10): 829–39.

Bardiot, Clarisse. 2013. *Arts de La Scène et Technologies Numériques: Les Digital Performances*. Les Basiques, Leonardo/Olats.

Baudel, Thomas, & Michel Beaudouin-Lafon. 1993. "Charade: Remote Control of Objects Using Free-Hand Gestures." *Communications of the ACM* 36(7): 28–35.

Benesh, Rudolf, & Joan Benesh. 1956. *An Introduction to Benesh Movement Notation*. Londres: Adam and Charles Black.

Berthoz, Alain. 1997a. "L'adaptation." In *Le sens du mouvement*, Paris, France: Odile Jacob, 345.

———. 1997b. "Référentiels." In *Le sens du mouvement*, Paris, France: Odile Jacob, 345.

Bevilacqua, Frédéric, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, Fabrice Guédy, & Nicolas Rasamimanana. 2010. "Continuous Realtime Gesture Following and Recognition." *Gesture in Embodied Communication and Human-Computer Interaction* 5934: 73–84.

Billinghurst, Marc, & Bill Buxton. 2011. "Gesture Based Interaction." In *Haptic Input*, Cambridge University Press, 35.

Bird, Jon. 2011. "The Phenomenal Challenge of Designing Transparent Technologies." *ACM Interactions* 18(6): 20–23.

Blender Foundation. 1995. "Blender." http://www.blender.org/ (October 25, 2013).

Bobillier-Chaumon, Marc Eric, Sarah Carvallo, Fraieeenck Tarpin-Bernard, & Jacqueline Vacherand-Revel. 2005. "Adapter Ou Uniformiser Les Interactions Personnes-Systèmes?" *Revue d'Interaction Homme-Machine* 6(2): 91–129.

Bolt, RA. 1980. "'Put-That-There': Voice and Gesture at the Graphics Interface." *ACM SIGGRAPH* 14: 262–70.

Bouchard, Stéphane, Stéphanie Dumoulin, Jeanne Talbot, André-anne Ledoux, Jennifer Phillips, Johana Monthuy-blanc, Geneviève Labonté-chartrand, Geneviève Robillard, Matteo Cantamesse, & Patrice Renaud. 2012. "Manipulating Subjective Realism and Its Impact on Presence: Preliminary Results on Feasibility and Neuroanatomical Correlates." *Interacting with Computers* 24: 227–36.

Bouisset, Simon, Eric Berton, Guillaume Rao, & David Amarantini. 2006. "Etude et Modelisation Du Mouvement Humain." In *Le traité de la Réalité Virtuelle Volume 1*, Paris, France: Les presses des Mines, 101–16.

Boutet, Dominique. 2008. “Une Morphologie de La Gestualité: Structuration Articulaire.” *Cahiers de linguistique analogique* 5: 81–115.

Bowman, Douglas A, Jian Chen, Chadwick A Wingrave, John Lucas, Andrew Ray, Nicholas F Polys, Qing Li, Yonca Haciahmetoglu, Ji-sun Kim, Seonho Kim, Robert Boehringer, & Tao Ni. 2006. “New Directions in 3D User Interfaces.” *International Journal of Virtual Reality IJVR* 5(2): 3–14.

Bowman, Douglas A., Ernst Kruijff, Joseph LaViola, Mark Mine, & Ivan Poupyrev. 2004. *3D User Interface Design: Fundamental Techniques, Theory and Practice*. Addison-Wesley.

Braffort, Annelies, & Sylvie Gibet. 2006. “La Gestuelle.” In *Le traité de la Réalité Virtuelle Volume 5*, Paris, France: Les presses des Mines, 83–112.

Bret, Michel, & Marie-Hélène Tramus. 2002. *La Funambule*. Issy les Moulineaux, France: Le Cube.

Burkhardt, Jean-Marie, Yves Duthen, Domitile Lourdeaux, Thierry Morineau, Marc Parentho, & Jacques Tisseau. 2004. “Thème 5: Comportements Humains Individuels et Collectifs.” In *Action Spécifique Humain virtuel: vers un humain synthétique temps-réel aussi vrai que nature*, CNRS.

Burkhardt, Jean-Marie, Patricia Plénacoste, & Laurence Perron. 2006. “Concevoir et Évaluer L’interaction Utilisateur-Environnement Virtuel.” In *Le traité de la réalité virtuelle Volume 2*, Paris, France: Presses des Mines, 552.

Cabral, Marcio C., Carlos H. Morimoto, & Marcelo K. Zuffo. 2005. “On the Usability of Gesture Interfaces in Virtual Reality Environments.” *ACM CLICH Latin American conference on Human-computer interaction*: 100–108.

Cadoz, Claude. 1994. “Le Geste Canal de Communication homme/Machine. La Communication ‘Instrumentale’.” *Technique et science informatiques* 13(1): 31–61.

———. 2006. “Retour Au Réel: Le Sens Du Feedback Rétroaction – Interaction – Énaction.” *Le feedback dans la création musicale, Rencontres musicales pluridisciplinaires*.

Cage, John, Merce Cunningham, David Tudor, Gordon Mumma, Nam June Paik, Stan VanDerBeeck, Robert Moog, & Billy Klüver. 1965. *Variations V*.

Cao, Caroline G. L., Christine L. MacKenzie, & Shahram Payandeh. 1996. “Task and Motion Analyses in Endoscopic Surgery.” *ASME Dynamic Systems and Control Division*: 583–90.

Cao, Xiang, & Ravin Balakrishnan. 2005. “Evaluation of an on-Line Adaptive Gesture Interface with Command Prediction.” In *Graphics Interface*, Canadian Human-Computer Communications Society, 187–94.

Caramiaux, Baptiste, Patrick Susini, Olivier Houix, & Frédéric Bevilacqua. 2011. “Study of the Impact of Sound Causality on Gesture Responses.” In *Etudes de la relation geste-son en performance musicale*, Thèse de l’Université Pierre et Marie Curie Paris 6, IRCAM.

Card, Stuart K, Thomas P Moran, & Allen Newell. 1983. *The Psychology of Human Computer Interaction*. Routledge.

Charbonneau, Emiko, Andrew Miller, Chadwick Wingrave, & Joseph J. LaViola Jr. 2009. “Understanding Visual Interfaces for the Next Generation of Dance-Based Rhythm Video Games.” In *ACM SIGGRAPH Symposium on Video Games*, , 119–26.

Choi, E., S. Kwon, D. Lee, H. Lee, & M. K. Chung. 2012. "Can User-Derived Gesture Be Considered as the Best Gesture for a Command?: Focusing on the Commands for Smart Home System." *Human Factors and Ergonomics Society Annual Meeting* 56(1): 1253–57.

Cicerone, Keith D., Cynthia Dahlberg, Kathleen Kalmar, Donna M. Langenbahn, James F. Malec, Thomas F. Bergquist, Thomas Felicetti, Joseph T. Giacino, J. Preston Harley, Douglas E. Harrington, Jean Herzog, Sally Kneipp, Linda Laatsch, & Philip A. Morse. 2000. "Evidence-Based Cognitive Rehabilitation: Recommendations for Clinical Practice." *Archives of physical medicine and rehabilitation* 81(12): 1596–1615.

Cifuentes, Jenny, Minh Tu Pham, Richard Moreau, Flavio Prieto, & Pierre Boulanger. 2013. "An Arc-Length Warping Algorithm for Gesture Recognition Using Quaternion Representation." *IEEE Engineering in Medicine and Biology Society*: 6248–51.

Ciger, Jan. 2012. "Razer Hydra HMD Hack." http://janoc.rd-h.com/archives/212.

Cole, Jonathon, Shaun Gallagher, & D McNeill. 2002. "Gesture Following Deafferentation: A Phenomenologically Informed Experimental Study." *Phenomenology and the Cognitive Sciences* 1(1): 49–67.

Coster, Xavier De, Charles De Groote, Arnaud Destin, Pierre Deville, Laurent Lamouline, Thibault Leruitte, & Vincent Nuttin. 2010. 1 *Mahalanobis Distance, Jaro-Winkler Distance and nDollar in UsiGesture*. Louvain, Belgium.

Csíkszentmihályi, Mihály. 1990. *Flow: The Psychology of Optimal Performance*. New York, NY, USA: Harper Perennial Modern Classics.

Van Dam, Andries. 1997. "Post-WIMP User Interfaces." *Communications of the ACM* 40(2): 63–67.

Dieterich, H., U. Malinowski, T. Kuhme, & M. Schneider-Hufschmidt. 1993. "State of the Art in Adaptive User Interfaces." In *Adaptive User Interfaces: Principles and Practice*, Amsterdam: Elsevier, 13–48.

Dixon, Steve, & Barry Smith. 2007. *Digital Performance: A History of New Media in Theater, Dance, Performance Art, and Installation*. Cambridge, USA: MIT Press.

Dreyfus, Hubert L. 1984. *Intelligence Artificielle: Mythes et Limites*. Paris, France: Flammarion.

Dukes, Patrick S., Austen Hayes, Larry F. Hodges, & Michelle Woodbury. 2013. "Punching Ducks for Post-Stroke Neurorehabilitation: System Design and Initial Clinical Evaluation." In *IEEE 3DUI 3D User Interfaces*, Orlando, USA.

Engelbart, Douglas C., & William K English. 1968. "A Research Center for Augmenting Human Intellect." *AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference* 33: 395–410.

Eshkol, Noa, & Avraham Wachman. 1958. *Movement Notation*. London, UK: Weidenfeld and Nicolson.

Farrington, Jeanne. 2011. "From the Research: Myths Worth Dispelling. Seven Plus or Minus Two." *Performance Improvement Quarterly* 23(4): 113–16.

Fdili Alaoui, Sarah, Caramiaux Baptiste, Marcos Serrano, & Frédéric Bevilacqua. 2012. "Movement Qualities as Interaction Modality." In *Designing Interactive Systems*, ACM, 761–69.

Fitts, P.M. 1954. "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement." *Journal of experimental psychology* 47(6): 381.

Françoise, Jules. 2011. "Realtime Segmentation and Recognition of Gestures Using Hierarchical Markov Models." UPMC, IRCAM, Telecom ParisTech.

Françoise, Jules, Ianis Lallemand, Thierry Artières, Frédéric Bevilacqua, Norbert Schnell, & Diemo Schwarz. 2013. "Perspectives Pour L'apprentissage Interactif Du Couplage Geste-Son." *JIM Journées d'Informatique Musicale*: 77–84.

Freeman, William T., & Craig D. Weissman. 1995. "Television Control by Hand Gestures." In *IEEE Workshop on Automatic Face and Gesture Recognition*, Zurich.

Fuchs, Philippe, Bruno Arnaldi, & Jacques Tisseau. 2003. "La Réalité Virtuelle et Ses Applications." In *Le traité de la Réalité Virtuelle Volume 1*, Paris, France: Les Presses de l'Ecole des Mines de Paris, 3–52.

Fuchs, Philippe, Alain Berthoz, & Jean-Louis Vercher. 2006. "Introduction à La Réalité Virtuelle." In *Le traité de la Réalité Virtuelle Volume 1*, Paris, France: Les presses des Mines, 380.

Fuchs, Philippe, Jean-Marie Burkhardt, & Domitile Lourdeaux. 2006. "Approche Théorique et Pragmatique de La Réalité Virtuelle." In *Le traité de la Réalité Virtuelle Volume 2*, Paris, France: Les presses des Mines, 19.

Gardner, Brett S. 2011. "Responsive Web Design: Enriching the User Experience." *Connectivity and the User Experience* 11(1): 13.

Gibson, James J. 1966. *The Senses Considered as Perceptual Systems*. Boston, MA, USA: Houghton Mifflin.

———. 1979. Boston: Houghton Mifflin *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin.

Gold, Brian J, Marc Pomplun, Nichola J Rice, & Robert Sekuler. 2008. "A New Way to Quantify the Fidelity of Imitation: Preliminary Results with Gesture Sequences." *Experimental brain research* 187(1): 139–52.

Goldin-Meadow, Susan. 1999. "The Role of Gesture in Communication and Thinking." *Trends in cognitive sciences* 3(11): 419–29.

Goldin-Meadow, Susan, & Martha Wagner Alibali. 2013. "Gesture's Role in Speaking, Learning, and Creating Language." *Annual review of psychology* 64: 257–83.

Good, Michael D., John A. Whiteside, Dennis R. Wixon, & Sandra J. Jones. 1984. "Building a User-Derived Interface." *Communications of the ACM* 27(10): 1032–43.

Gourlay, Douglas, Kwok Chan Lun, Y. N. Lee, & John Tay. 2000. "Virtual Reality for Relearning Daily Living Skills." *International journal of medical informatics* 60(3): 255–61.

Guyon, Isabelle, Vassilis Athitsos, Pat Jangyodsuk, Hugo Jair Escalante, & Ben Hamner. 2012. "Results and Analysis of the ChaLearn Gesture Challenge 2012." *ChaLearn Gesture Challenges Workshop*: 1–20.

Hand, Chris. 1997. "A Survey of 3D Interaction Techniques." *Computer graphics forum* 16(5): 269–81.

Hartson, Rex. 2003. "Cognitive, Physical, Sensory, and Functional Affordances in Interaction Design." *Behaviour & Information Technology* 22(5): 315–38.

Herbelin, Bruno, Jan Ciger, & Anthony L Brooks. 2011. "Customising Games for Non-Formal Rehabilitation." *International Journal on Disability and Human Development* 10(1): 5–9.

Hiltz, Starr Roxanne, & Murray Turoff. 1981. "The Evolution of User Behavior in a Computerized Conferencing System." *Communications of the ACM* 24(11): 739–51.

Hurtienne, Jörn, & Patrick Langdon. 2009. "Prior Knowledge in Inclusive Design: The Older, the More Intuitive." In *Proceedings of the 23rd British Computer Society Human Computer Interaction Workshop and Conference (HCI'09)*,.

Hurtienne, Jörn, K Weber, & L Blessing. 2008. "Prior Experience and Intuitive Use: Image Schemas in User Centred Design." In *Designing inclusive futures*, Springer, 107–16.

Hutchins, Edwin L, James D Hollan, & Donald A Norman. 1986. "Direct Manipulation Interfaces." In *User Centered System Design: new perspectives on human-computer interaction*, Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 87–124.

Jacob, Robert J. K., Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, & Jamie Zigelbaum. 2008. "Reality-Based Interaction: a Framework for Post-WIMP Interfaces." *ACM SIGCHI Human Factors in Computing Systems*: 201–10.

Jensenius, Alexander Refsum. 2007. "Action-Sound: Developing Methods and Tools to Study Music-Related Body Movement." University of Oslo.

Kallmann, Marcelo, & Daniel Thalmann. 1999. "Direct 3D Interaction with Smart Objects." *ACM VRST Virtual Reality Software and Technology*: 124–30.

Keates, Simeon, & Peter Robinson. 1998. "The Use of Gestures in Multimodal Input." In *ACM conference on Assistive technologies ASSETS*, New York, NY, USA, 35–42.

Kendon, Adam. 1980. "Gesticulation and Speech: Two Aspects of the Process of Utterance." In *The relationship of verbal and nonverbal communication*, The Hague, The Netherlands: Mouton publishers, 207–27.

———. 1988. "How Gestures Can Become Like Words." *Cross-cultural perspectives in nonverbal communication*: 131–41.

———. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.

Kita, Sotaro, & Ingeborg Van Gijn. 1998. "Movement Phases in Signs and Co-Speech Gestures, and Their Transcription by Human Coders." *Gesture and sign language in human*.

Klinger, Evelyne. 2006. "Apports de La Réalité Virtuelle à La Prise En Charge de Troubles Cognitifs et Comportementaux." Telecom Paris.

Klinger, Evelyne, Abdelmajid Kadri, Eric Sorita, Jena-Luc Le Guiet, Pauline Coignard, Philippe Fuchs, Laure Leroy, Nicolas du Lac, Fabrice Servant, & Pierre-Alain Joseph. 2013. "AGATHE: A Tool for Personalized Rehabilitation of Cognitive Functions Based on Simulated Activities of Daily Living." *IRBM* 34(2): 113–18.

Koffka, Kurt. 1935. *Principles of Gestalt Psychology*. London, UK: Lund Humphries.

Kristensson, Per-Ola, & Shumin Zhai. 2004. "SHARK^2: a Large Vocabulary Shorthand Writing System for Pen-Based Computers." *ACM UIST User interface software and technology* 6(2).

Krueger, Myron W, Thomas Gionfriddo, & Katrin Hinrichsen. 1985. "VIDEOPLACE an Artificial Reality." *ACM SIGCHI Human Factors in Computing Systems*: 35–40.

Kuntz, Sébastien, & Ján Cíger. 2012. "Low-Cost and Home-Made Immersive Systems." *The International Journal of Virtual Reality*: 1–9.

Kurtenbach, Gord, & Eric Hulteen. 1990. "Gestures in Human-Computer Communication." In *The Art and Science of Interface Design*, Addison-Wesley Publishing Co., 309–17.

Laban, Rudolf, & Lisa Ullmann. 1971. *The Mastery of Movement*.

Lampton, Donald R., Bruce W. Knerr, Stephen L. Goldberg, James P. Bliss, & Michael J. Moshell. 1994. "The Virtual Environment Performance Assessment Battery (VEPAB): Development and Evaluation." *Presence: Teleoperators and Virtual Environments* 3(2): 145– 157.

Lange, Belinda, Chien-Yen Chang, Evan Suma, Bradley Newman, Albert Skip Rizzo, & Mark Bolas. 2011. "Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool Using the Microsoft Kinect Sensor." *IEEE Engineering in Medicine and Biology Society* 2011: 1831–34.

Langley, Pat. 1999. "User Modeling in Adaptive Interfaces." *International Conference on User Modeling*: 357–70.

Laurel, Brenda. 1986. "Interface as Mimesis." In *User Centered System Design: new perspectives on human-computer interaction*, Hillsdale, New-Jersey: Lawrence Erlbaum Associates, 67–85.

LaViola, Joseph J., & Daniel F. Keefe. 2011. "3D Spatial Interaction: Applications for Art, Design, and Science." In *ACM SIGGRAPH Courses*, Vancouver, British Columbia, Canada: ACM, 72.

Lécuyer, Anatole, Fabien Lotte, Richard B Reilly, & Trinity College. 2008. "Brain-Computer Interfaces, Virtual Reality, and Videogames." *IEEE Computer* 41(10): 66–72.

Lee, Johnny Chung. 2008. "Hacking the Nintendo Wii Remote." *IEEE Pervasive Computing* 7(3): 39–45.

———. 2010. "In Search of a Natural Gesture." *ACM Crossroads* 16(4): 9–12.

———. 2011. "Myth of the Dying Mouse: Why There Is No Such Thing as Convergence in Consumer Electronics." In *Web 2.0 Expo*, New York, NY, USA: O'Reilly Media, Inc., UBM TechWeb.

Levin, Golan, Zachary Lieberman, Jaap Blonk, & Joan La Barbara. 2003. *Messa Di Voce*. London, UK: ICA.

Liddell, Scott K, & Robert E Johnson. 1989. "American Sign Language: The Phonological Base." *Sign language studies* 64: 195–278.

De Loor, Pierre, Kristen Manac, Alexandra Fronville, & Jacques Tisseau. 2008. "Requirement for an Enactive Machine: Ontogenesis, Interaction and Human in the Loop." *International Conference on Enactive Interfaces*.

Lyons, Damian M., & Michael A. Arbib. 1989. "A Formal Model of Computation for Sensory-Based Robotics." *Robotics and Automation, IEEE* 5(3).

Manitsaris, Sotiris. 2010. “Vision Par Ordinateur Pour La Reconnaissance Des Gestes: Analyse et Modélisation Stochastique Du Geste Dans L’interaction Musicale.” Université de Macédoine.

Mäntyjärvi, Jani, Juha Kela, Panu Korpipää, & Sanna Kallio. 2004. “Enabling Fast and Effortless Customisation in Accelerometer Based Gesture Interaction.” *ACM MUM Mobile and ubiquitous multimedia*: 25–31.

Martin, Jérôme. 2000. “Reconnaissance de Gestes En Vision Par Ordinateur.” IPNG.

Marvit, David L., Albert H. M. Reinhardt, B. Thomas Adler, & Hitoshi Matsumoto. 2007. “User Definable Gestures for Motion Controlled Handheld Devices.”

McGrenere, Joanna, & Wayne Ho. 2000. “Affordances : Clarifying and Evolving a Concept.” *Graphics Interface* (May): 179–86.

McMahan, Ryan P., Alexander Joel D. Alon, Shaimaa Lazem, Robert J. Beaton, David Machaj, Michael Schaefer, Mara G. Silva, Anamary Leal, Robert Hagan, & Douglas A. Bowman. 2010. “Evaluating Natural Interaction Techniques in Video Games.” In *IEEE 3DUI 3D User Interfaces*, Waltham, MA, 11–14.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press.

“Meant to Be Seen.” 2009. http://www.mtbs3d.com/phpBB/ (October 25, 2013).

Microsoft. 2013. *Human Interface Guide Kinect for Windows V1.8*.

Milgram, Paul, Haruo Takemura, Akira Utsumi, & Fumio Kishino. 1994. “Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum.” In *Photonics for Industrial Applications*, , 282–92.

Miller, Georges A. 1956. “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” *Psychological review* 63(2): 81–97.

Mine, Mark R. 1995. UNC Chapel Hill computer science technical report *Virtual Environment Interaction Techniques*. Chapel Hill.

Mine, Mark R., P. Brooks Jr. Frederick, & Carlo H. Sequin. 1997. “Moving Objects In Space: Exploiting Proprioception In Virtual-Environment Interaction.” In *ACM SIGGRAPH*, New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 19–26.

Moreau, Richard, Victor M. Ochoa Mayorga, Minh Tu Pham, Pierre Boulanger, Tanneguy Redarce, & Olivier Dupuis. 2008. “A Method to Evaluate Skill Transfer and Acquisition of Obstetric Gestures Based on the Curvatures Analysis of the Position and the Orientation.” *Journal of Biomedical Informatics* 41(6): 991–1000.

Morrison, K, & S McKenna. 2002. “Contact-Free Recognition of User-Defined Gestures as a Means of Computer Access for the Physically Disabled.” In *Workshop on Universal Access and Assistive Technology*, Cambridge, UK, 99–103.

Mullaly, John. 1998. “IBM RealThings.” In *ACM SIGCHI Human Factors in Computing Systems*, New York, NY, USA: ACM, 13–14.

Myers, Brad A. 1998. “A Brief History of Human Computer Interaction Technology.” *ACM Interactions* 5(2): 44–54.

Neisser, Ulric. 1976. *Cognition and Reality: Principles and Implications of Cognitive Psychology.* WH Freeman/Times Books/Henry Holt & Co.

Nielsen, Jakob, & Rolf Molich. 1990. “Heuristic Evaluation of User Interfaces.” In *ACM SIGCHI Human Factors in Computing Systems*, New York, NY, USA: ACM, 249–56.

Nielsen, Michael, Moritz Störring, Thomas B. Moeslund, & Erik Granum. 2004. “A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI.” *Gesture-Based Communication in Human-Computer Interaction*: 409–20.

Norman, Donald A. 1989a. “Human-Centered Design.” In *The Design of Everyday Things*, New-York: Doubleday, 188.

———. 1989b. “Seven Principles for Transforming Difficult Tasks into Simple Ones.” In *The design of everyday things*, New-York: Doubleday, 188.

———. 2010. “Natural User Interfaces Are Not Natural.” *Interactions* 17(3): 6–10.

Oppermann, Reinhard, & Rossen Rashev. 1997. “Adaptability and Adaptivity in Learning Systems.” *Knowledge transfer* 2(2): 173–79.

Otaduy, Miguel A., Takeo Igarashi, & Joseph J. LaViola Jr. 2009. “Interaction: Interfaces, Algorithms, and Applications.” In *ACM SIGGRAPH Courses*, New York, NY, USA: ACM, 14:1–14:66.

Pagliei, Lorenzo. 2013. *Voir-Toucher*. Paris, France: Centre-Pompidou.

Paillard, J. 1974. “Le Traitement Des Informations Spatiales.” *De l’espace corporel à l'espace écologique* (7): 54.

Paljic, Alexis, Sabine Coquillart, Jean-Marie Burkhardt, & Paul Richard. 2002. “A Study of Distance of Manipulation on the Responsive Workbench (tm).” In *Immersive Projection Technology Workshop IPT*, Orlando.

Pausch, Randy. 1991. “Virtual Reality on Five Dollars a Day.” In *ACM SIGCHI Human Factors in Computing Systems*, , 265–70.

Piaget, Jean. 1971. *Biology and Knowledge: An Essay on the Relations Between Organic Regulations and Cognitive Processes.* University of Chicago Press.

Pirhonen, Antti. 2010. “Gestures in Human-Computer Interaction--Just Another Modality?” In *Gesture in Embodied Communication and Human-Computer Interaction*, Springer, 281–88.

Piumsomboon, Thammathip, Adrian Clark, Mark Billinghurst, & Andrew Cockburn. 2013. “User-Defined Gestures for Augmented Reality.” *ACM SIGCHI Human Factors in Computing Systems*: 955.

“Project Holodeck.” 2012. http://www.projectholodeck.com/ (October 25, 2013).

Quek, Francis, David Mcneill, Robert Bryll, & Karl E Mccullough. 2002. “Multimodal Human Discourse : Gesture and Speech.” *ACM TOCHI Transactions on Computer-Human Interaction* 9(3): 171–93.

Rabiner, Lawrence R. 1989. “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.” *Proceedings of the IEEE* 77(2): 257–86.

Rautaray, Siddharth S., & Anupam Agrawal. 2012. “Vision Based Hand Gesture Recognition for Human Computer Interaction: a Survey.” *Artificial Intelligence Review*: –-54.

Regan, E.C. 1995. “Some Evidence of Adaptation to Immersion in Virtual Reality.” *Displays* 16(3): 135–39.

Rich, Elaine. 1989. “Stereotypes and User Modeling.” In *In Users models in dialog systems*, ed. W. Wahlster Kobsa. Berlin: Springer-Verlag, 35–51.

Rigal, Robert. 2003. *Motricité Humaine. Fondements et Applications Pédagogiques - Tome 2 Développement Moteur*. 3è édition. Presses de l’Université du Québec.

Rizzo, Albert A., & J. Galen Buckwalter. 1997. “Virtual Reality and Cognitive Assessment and Rehabilitation: The State of the Art.” *Studies in health technology and informatics* 44: 123–45.

Rizzolatti, Giacomo, Luciano Fadiga, Leonardo Fogassi, & Vittorio Gallese. 1997. “The Space Around Us.” *Science* 277(5323): 190–91.

Rizzolatti, Giacomo, & Corrado Sinigaglia. 2011. “Le Cerveau Qui Agit.” In *Les neurones miroirs*, Paris, France: Odile Jacob, 236.

Rubine, Dean. 1991. “Specifying Gestures by Example.” In *ACM SIGGRAPH*, , 329–37.

Ruiz, Jaime, Y Li, & E Lank. 2011. “User-Defined Motion Gestures for Mobile Interaction.” In *ACM SIGCHI Human Factors in Computing Systems*, , 197–206.

Russel, M. Taylor II, Thomas C. Hudson, Adam Seeger, Hans Weber, Jeffrey Juliano, & Aron T Helser. 2001. “VRPN : A Device-Independent , Network-Transparent VR Peripheral System.” In *ACM VRST, Virtual reality software and technology*, , 55–61.

Sanchez, Rosa, & Alain Baumann. 2006. *NOU I_D: Electronic Interactive Opera for Dance and Voice*. Barcelona, Spain: KONIC THTR.

Schlömer, Thomas, Benjamin Poppinga, Niels Henze, & Susanne Boll. 2008. “Gesture Recognition with a Wii Controller.” In *ACM TEI Tangible and Embedded Interaction*, Bonn, Germany: ACM, 11–14.

Schultz, Maureen, Janet Gill, Sabiha Zubairi, Ruth Huber, & Fred Gordin. 2003. “Bacterial Contamination of Computer Keyboards in a Teaching Hospital.” *Infection control and hospital epidemiology* 24(4): 302–3.

Serre, Jean-Claude. 1984. “La Danse Parmi Les Autres Formes de Motricité.” *La recherche en danse* 26: 135–56.

Shneiderman, Ben. 1998. *Designing The User Interface: Strategies for Effective Human-Computer Interaction*. Addison Wesley Longman.

Shotton, Jamie, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, & Andrew Blake. 2013. “Real-Time Human Pose Recognition in Parts from a Single Depth Image.” *Communications of the ACM* 56(1): 116–24.

Simoneau, Serge, Marie St-Vincent, & Denise Chicoine. 1996. *Les Lésions Musculosquelettiques Associées Au Travail Répétitif, Mieux Les Comprendre Pour Mieux Les Prévenir*. Bibliothèque nationale du Québec Association Paritaire pour la Sante et la Securite du Travail Secteur Fabrication de Produits en Metal et Electriques.

Slater, Mel, J Howell, & A Steed. 2000. “Acting in Virtual Reality.” *Proceedings of the third international conference on Collaborative virtual environments*: 103–10.

Smets, Gerda J. F., Pieter Jan Stappers, & Kees Overbeeke. 1994. “Designing in Virtual Reality: Implementing Perception-Action Coupling with Affordances.” In *ACM VRST Virtual reality software and technology*, , 97–110.

Sowa, Timo, & Ipke Wachsmuth. 2002. "Interpretation of Shape-Related Iconic Gestures." In *Gesture and sign language in human-computer interaction*, Springer Berlin Heidelberg, 21–33.

Starner, Thad, Joshua Weaver, & Alex Pentland. 1998. "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video." *IEEE Pattern Analysis and Machine Intelligence* 20(12): 1371–75.

Steil, Jochen J, & Heiko Wersing. 2006. "Recent Trends in Online Learning for Cognitive Robotics." *ESANN*: 77–87.

Stewart, John, Armen Khatchatourov, & Charles Lenay. 2004. "Enaction and Engineering."

Stokoe, William C. 1972. *Semiotics and Human Sign Language*. Walter de Gruyter.

Stößel, Christian, Hartmut Wandke, & Lucienne Blessing. 2010. "Gestural Interfaces for Elderly Users: Help or Hindrance?" In *Gesture in Embodied Communication and Human-Computer Interaction*, Springer, 269–80.

Suma, Evan A., Belinda Lange, Skip Rizzo, David Michael Krum, & Mark Bolas. 2011. "FAAST Flexible Action and Articulated Skeleton Toolkit." In *IEEE Virtual Reality*, Singapore: MxR Lab, University of Southern California, 247 – 248.

"The VR Geeks Association." 2010. Http://www.vrgeeks.org/ (October 25, 2013).

Thouvenin, Indira. 2009. "Interaction et Connaissance: Construction D'une Expérience Dans Le Monde Virtuel." Université de Technologie Compiègne.

Tognazzini, Bruce. 1999. "A Quiz Designed to Give You Fitts." *Ask Tog*: 1–7. http://www.asktog.com/columns/022DesignedToGiveFitts.html (October 25, 2013).

Tough, Christopher. 2013. *Can Gesture Recognition Repair the Disembodiment of Live Electronic Music?*

Tyndiuk, F, C Schlick, B Claverie, & G Thomas. 2003. "Modèles et Facteurs Humains En IHM - Application à La Réalité Virtuelle." *Modeles Formels de l'Interaction (MFI)*: 301–6.

Valli, Alessandro. 2008. "The Design of Natural Interaction." *Multimedia Tools and Applications* 38(3): 295–305.

Varela, Francisco J, Evan Thompson, & Eleanor Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Vercher, Jean-Louis. 2006. "La Latence Temporelle Dans La Boucle de Réalité Virtuelle." In *Le traité de la Réalité Virtuelle Volume 1*, Paris, France: Les presses des Mines, 339–55.

Verna, Didier. 2000. *Télé-Opération et Réalité Virtuelle: Assistance à L'opérateur Par Modélisation Cognitive de Ses Intentions*. Paris: Thèse Telecom Paris.

Vik, Chris. 2011. "Kinectar Performance Platform." http://ethnotekh.com/project/kinectar/.

Viviani, Paolo, & Carlo Terzuolo. 1980. "32 Space-Time Invariance in Learned Motor Skills." *Advances in psychology*.

Voida, Stephen, Mark Podlaseck, Rick Kjeldsen, & Claudio Pinhanez. 2005. "A Study on the Manipulation of 2D Objects in a Projector/camera-Based Augmented Reality Environment." In *ACM SIGCHI Human Factors in Computing Systems*, New York, NY, USA: ACM, 611–20.

"VR Juggler." 2007. http://vrjuggler.org/ (October 25, 2013).

"VRoot.org." 2006. http://vroot.org/ (October 25, 2013).

Wagner, Susan M., Howard Nusbaum, & Susan Goldin-Meadow. 2004. "Probing the Mental Representation of Gesture: Is Handwaving Spatial?" *Journal of Memory and Language* 50(4): 395–407.

Wassner, Hubert. 2008. "La Wii Au Secours Des Handicapés." *Article de Ducassou, Emmanuel*. http://www.handimarseille.fr/le-magazine/sciences-et-techno/article/la-wii-au-secours-des-handicapes (October 23, 2013).

Winograd, Terry, & Fernando Flores. 1989. *L'intelligence Artificielle En Question*. Paris, France: PUF.

Wobbrock, Jacob O., Meredith Ringel Morris, & Andrew D. Wilson. 2009. "User-Defined Gestures for Surface Computing." In *ACM SIGCHI Human Factors in Computing Systems*, Boston, MA, USA: ACM Press, 1083–92.

Wobbrock, Jacob O., Andrew D. Wilson, & Yang Li. 2007. "Gestures Without Libraries, Toolkits or Training: a $1 Recognizer for User Interface Prototypes." In New York, NY, USA: ACM, 159–68.

Zamborlin, Bruno. 2012. "Mogee." http://www.brunozamborlin.com/mogees/ (October 30, 2013).

Zhang, Shiqi, Chun Yuan, & Yan Zhang. 2008. "Self-Defined Gesture Recognition on Keyless Handheld Devices Using MEMS 3D Accelerometer." *IEEE Conference on Natural Computation*: 237–41.

Zhao, Liwei. 2001. "Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures." University of Pennsylvania.

TABLES DES ILLUSTRATIONS

TABLE DES FIGURES

Figure 1-1 : L'interaction en trois générations, : (a) lignes de commande (CLI), (b) manipulation directe (GUI), (c) styles d'interactions émergents et basés sur le monde réel (NUI), adapté de (Jacob et al. 2008).....3
Figure 1-2 : Les quatre thèmes de l'interaction basée monde réel selon (Jacob et al. 2008).....4
Figure 1-3 : Deux utilisateurs en immersion et interaction avec des dispositifs à bas coûts : à gauche casque Vuzix VR920, Gametrak et Wiimote / à droite casque Sony HMD, SpacePoint Fusion et Razer Hydra. Photos de l'association VRGeeks extraites de (Kuntz et al. 2012).....6
Figure 1-4 : Un patient accompagné d'une thérapeute s'entraine avec une application de réalité virtuelle (Klinger et al. 2013).....8
Figure 1-5 : Captures d'écran de l'installation interactive *Videoplace* (Krueger et al. 1985).....10
Figure 1-6 : Positionnement des problématiques soulevées par une approche où l'Utilisateur Définit l'Interaction gestuelle.....13
Figure 2-1 : Continuum des sources de connaissances selon (Hurtienne et al. 2009).....18
Figure 2-2 : Différents niveaux d'adaptation et de prescription des systèmes et du travail (Bobillier-Chaumon et al. 2005).....19
Figure 2-3 : Différents types de processus d'adaptation : de l'adaptabilité à l'adaptativité d'après (Dieterich et al. 1993), extrait de (Bobillier-Chaumon et al. 2005).....21
Figure 2-4 : Taux de consensus en ordre décroissant pour quarante tâches (en bleu) et ratio de l'utilisation de gestes bimanuels (ligne rouge) extrait de (Piumsomboon et al. 2013).....25
Figure 2-5 : Sept phases de l'interaction homme système selon (Norman 1989a).....26
Figure 2-6 : Notion de *flow* (Csíkszentmihályi 1990).....29
Figure 2-7 : Le continuum de Kendon en relation à la parole.....33
Figure 2-8 : Découpage d'un geste et phénomène de coarticulation entre deux gestes, adapté de (Kita et al. 1998).....37
Figure 2-9 : Schéma général d'un système de reconnaissance de gestes (Braffort et al. 2006).....38
Figure 2-10 : Capteurs grand public. (a) Caméra de profondeur *Microsoft Kinect*, (b) Bracelet à centrale inertielle *Movea MotionPod*, (c) Manette à capteur magnétique *Razer Hydra* et modification possible...39
Figure 2-11 : Projet Holodeck, combinaison de capteurs *Playstation Move* et *Razer Hydra*. Crédit photo : ("Project Holodeck" 2012).....41
Figure 2-12 : Association geste et son à l'apprentissage et à la performance (Bevilacqua et al. 2010).....43
Figure 2-13 : Relations du façonnage interactif à d'autres champs de recherche (Françoise et al. 2013).....44
Figure 2-14 : Techniques possibles de reconnaissance de gestes.....44
Figure 2-15 : Association de classe de gestes a un geste continu, extrait de (Françoise 2011). (a) apprentissage de trois gestes de référence. (b) reproduction d'un geste en continu en entrée du modèle. (c) extraction des classes de gestes par le système de reconnaissance.....45
Figure 2-16 : Interaction (gauche) et rétroaction (droite), extrait de (Cadoz 2006).....49
Figure 2-17 : Informations disponibles lors d'un mouvement (interaction) et *feedback* (Rigal 2003).....50
Figure 2-18 : À gauche, représentation de trajectoires angulaires de différents gestes experts dans l'espace des quaternions (Moreau et al. 2008) ; à droite, distance de Mahalanobis (Coster et al. 2010). En distance euclidienne, x appartient à la classe C1alors qu'avec la distance de Mahalanobis, x appartient à la classe C2.....52
Figure 3-1 : Notre approche UDIg en deux étapes.....56
Figure 3-2 : Couplage geste et animation d'objet aux phases de création et d'interaction du geste (a). À noter en (b) que l'invariance aux variations temporelles permet de conserver la synchronisation du mouvement d'un geste à l'animation. Adapté de (Françoise et al. 2013).....63
Figure 4-1 : Affordances présentes à trois niveaux : physique, cognitif et fonctionnel selon (Hartson 2003).....67
Figure 4-2 : Étapes de l'expérimentation, liste des boites, conditions d'affordance, et chronologie de l'expérimentation.....70
Figure 4-3 : Les boites sont présentées à l'utilisateur selon quatre combinaisons de conditions d'expérimentation. A0: pas d'affordance explicite, A1: affordance explicite. C0: manipulation distante, C1 manipulation colocalisée.....71
Figure 4-4 : Un utilisateur pendant l'expérimentation et détail du dispositif.....72
Figure 4-5 : Schéma du système de reconnaissance de gestes utilisé pour l'UDIg.....73
Figure 4-6: Notre approche du découpage d'une geste basé sur (Kendon 1980). À droite, un utilisateur effectuant un geste de manipulation dans une boite englobante.....74
Figure 4-7 : Temps de réflexion en millisecondes. Les valeurs basses représentent une meilleure performance..77

Figure 4-8: Distances euclidiennes moyennées des trajectoires entre le geste original et sa reproduction. Les valeurs basses représentent une meilleure performance (en mètres) 77
Figure 4-9 : Taux de reconnaissance du geste (1 correspond à 100%). Les valeurs hautes représentent une meilleure performance 77
Figure 4-10 : Comparaison entre les trajectoires du geste créé et du geste reproduit à l'interaction (référentiel absolu). La courbe noire représente le geste original. La courbe colorée représente la distance à la courbe originale. 78
Figure 5-1 : Continuum représentant les affordances et les informations qui définissent ces affordances (McGrenere et al. 2000) 85
Figure 5-2 : Schéma technocentré des modalités d'interaction selon les dispositifs (Lee 2011) 87
Figure 5-3 : Cuisine réelle (gauche) et reproduction 3D temps-réel (droite) 89
Figure 5-4 : Un utilisateur interagissant avec la porte virtuelle ; à gauche, condition geste d'Amplitude Réaliste (AR) ; à droite, condition geste d'Amplitude Faible (AF) 94
Figure 5-5: Notre approche du découpage d'une geste basé sur (Kendon 1980). À droite, un utilisateur effectuant un geste de manipulation dans une boite englobant 94
Figure 5-6 : Résultats globaux selon les métriques ; en bleu : condition geste d'Amplitude Réaliste (AR), en vert condition geste d'Amplitude Faible 96
Figure 5-7 : Schéma de placement des hypothèses A à E dans l'analyse des conditions de rétroaction et de répétitions 97
Figure 5-8 : Résultats des performances pour le Travail, le Temps et le Taux de complétion selon la condition de couplage des rétroactions 97
Figure 5-9 : Résultats des performances par rétroactions pour le Travail, pour le Temps et pour le Taux de complétion 98
Figure 5-10 : Résultats des performances par répétitions en fonction des rétroactions pour le Travail, pour le Temps et pour le Taux de complétion 98
Figure 5-11 : Résultats des questionnaires subjectifs sur la qualité de l'interaction par rétroaction 100
Figure 5-12 : Moyennes des résultats des questionnaires subjectifs sur la qualité de l'interaction en fonction de la rétroaction (détail des résultats sur la Figure 5-11) 101
Figure 5-13 : Résultats des performances pour les conditions d'Amplitude pour le Travail, pour le Temps et pour le Taux de complétion 102
Figure 5-14 : Résultats des performances pour les conditions d'Amplitude en fonction des Répétitions pour le Travail, pour le Temps et pour le Taux de complétion 103
Figure 5-15 : Résultats du questionnaire subjectif d'utilisabilité entre conditions de gestes d'Amplitude Réalise (AR en bleu) et d'Amplitude Faible (AF en vert) 104
Figure 5-16 : Résultats des performances en fonction des Répétitions pour le Travail, le Temps et le Taux de complétion 104
Figure 6-1 : Utilisateurs du démonstrateur grand public (Crédit photo : Futur en Seine 2013) 117
Figure 6-2 : A gauche, cuisine virtuelle ; en haut à droite, boite de thé réelle ; en bas à droite, boite modélisée en 3D 120
Figure 6-3 : Un patient interagissant dans la cuisine virtuelle 121
Figure 6-4 : Préconisations pour l'approche où l'Utilisateur Définit l'Interaction gestuelle (les flèches donnent un sens de lecture général) 124

LISTE DES TABLES

Table 2-1: Intérêts et inconvénients des approches d'uniformisation et d'adaptation, extrait de (Bobillier-Chaumon et al. 2005)....22
Table 2-2 : Tendances et limites en conception inclusive....25
Table 2-3: Avantages et Limites des capteurs à bas coût....41
Table 4-1: Comparaison des copropriétés des affordances selon Gibson et selon Norman (McGrenere et al. 2000) 67
Table 5-1 : Choix des rétroactions présentées à l'utilisateur classées par dynamique et chronologie de l'expérimentation....92
Table 6-1 : Protocole avec deux activités et deux types de retours visuels à l'interaction 116
Table 6-2 : Protocole avec deux activités et deux types de retours visuels à l'interaction 119
Table 6-3 : Récapitulation des préconisations de l'approche où l'Utilisateur Définit l'Interaction gestuelle (les chiffres renvoient aux chapitres correspondants)....126

Appendices

ANNEXE A

TABLEAU COLLABORATIF DESCRIPTIF DES CAPTEURS GRAND PUBLIC SELON L'USAGE ET LES SPÉCIFICATIONS TECHNIQUES

	Brand	Apple	Essential Reality	Hillcrest			In2game	LEAP
	Model	iPod Touch	P5 Glove	FSRK-2.4G-1	FSRK-BT-1	FSRK-USB-1	Gametrack	LeapMotion
USAGE	Body parts	hand, any	hand + 5 fingers	any	any	any	2 hands, any	2 hand + 10 fingers
	Position		Absolute for hand				3 DOF	3DOF
	Orientation	3 DOF, Absolute	Absolute for hand	3 DOF, Absolute	3 DOF, Absolute	3 DOF, Absolute	2 DOF, Absolute	3 DOF, Absolute
	Range	any	1m				2m	<1m
	Drift	low		?	?	?	low	distance-dependent
	Wireless	wifi bluetooth	plugged on base	RF	Bluetooth	USB	cables	USB
	Batteries	Integrated	no	2 x AAA batteries	2 x AAA batteries	no	no	no
	VRPN		yes				yes	
	SDK	100$	OpenSource	OpenSource : libfreespace	OpenSource : libfreespace	OpenSource : libfreespace	OpenSource	OpenSource
INPUT	Accelerometer (Axes)	3		3	3	3		
	Gyroscope (Axes)	3		3	3	3		
	Magnetic compass	yes		?	?	?		
	Thermometer			?	?	?		
	GPS	yes						
	Microphones	1with headset						
	RGB Camera	2						
	IR Camera		yes					?
	IR Emitter		8 LED					yes
	Magnetic tracking							
	Buttons	4 +screen	4					
	AnalogStick							
	Analog Buton		Analogic fingers bend					
	Tactile	Multipoints						
OUTPUT	Visual feedback	2,7" screen						
	Rumble	analogic						
	Speaker	stereo						
	Price $US	200	70	149	119	99	40	70

	Brand	Movea	Microsoft		Nintendo			
	Model	MotionPod	Kinect	Kinect 2 (TBA)	Wiimote	Nunchuck	Wiimote + WiiMotion+	Wiimote + SensorBar
USAGE	Body parts	any	whole body, no fingers	whole body, face tracking, heartbeat	hand, any	hand, any	hand, any	pointing hand, any
	Position		Absolute w/ constraint	Absolute w/ constraint				3 DOF, Absolute
	Orientation	3 DOF, Absolute	Partial	Partial	3 DOF, Relatives	3 DOF, Relatives	3 DOF, Relatives	2 DOF, Absolute
	Range	any	4m		any	wiimote dependent	any	4m
	Drift	low	Microsoft	distance-dependent	high		low	high
	Wireless	RF	vision-based, USB	vision-based, USB	Bluetooth	plugged on wiimote	Bluetooth	Bluetooth
	Batteries	Integrated	no	no	2 x AAA batteries	no	2 x AAA batteries	2 x AAA batteries
	VRPN		yes		yes	yes	yes	yes
	SDK	2000 $	OpenNI, Kinect SDK, Freenect	N/A	OpenSource	OpenSource	OpenSource	OpenSource
INPUT	Accelerometer (Axes)	3	3	3	3	3	3	3
	Gyroscope (Axes)	3					2+1	
	Magnetic compass	yes						
	Thermometer	?						
	GPS							
	Microphones		4 with beamforming	?				
	RGB Camera		yes	yes				
	IR Camera		yes	yes	yes		yes	yes
	IR Emitter		yes	yes				yes
	Magnetic tracking							
	Buttons				11	2	11	11
	AnalogStick					1		
	Analog Buton							
	Tactile							
OUTPUT	Visual feedback	1LED	3 colors LED	3 colors LED	4 LEDs		4 LEDs	4 LEDs
	Rumble				bool		bool	bool
	Speaker				mono		mono	mono
	Price $US	1000$ each, to 1$ in quatity	150	N/A	30	35	40	50

	Brand	NaturalPoint	Playstation				PNI	Razer
	Model	TrackIR	Move	Move + PSEye	Navigation controller	Sixaxis	Spacepoint Fusion	Hydra
USAGE	Body parts	any	hand, any	hand, any	hand, any	hand, any	any	2 hands, any
	Position	3 DOF, Absolute		3 DOF, Absolute				3 DOF, Absolute
	Orientation	3 DOF, Absolute	3 DOF, Absolute	3 DOF, Absolute	3 DOF, Relatives	3 DOF, Relatives	3 DOF, Absolute	3 DOF, Absolute
	Range		any	4m	any	any	any	1m from base
	Drift		distance-dependent	distance-dependent		high	medium	distance-dependent
	Wireless		Bluetooth	Bluetooth	Bluetooth	Bluetooth	USB	plugged on base
	Batteries		Integrated	Integrated	Integrated	Integrated	no	no
	VRPN						yes	yes
	SDK		Opensource, Move.Me + PS3 400$	Opensource, Move.Me + PS3 400$	Opensource, Move.Me + PS3 400$	Open Source	OpenSource	yes
INPUT	Accelerometer (Axes)		3	3		3	3	
	Gyroscope (Axes)		3	3		3	3	
	Magnetic compass		yes	yes			yes	
	Thermometer		yes	yes				
	GPS							
	Microphones		2	2 + array				
	RGB Camera			yes				
	IR Camera	1						
	IR Emitter	1						
	Magnetic tracking							6 axis
	Buttons		9	9	10	17		7x2
	AnalogStick		1	1	1	2		1x2
	Analog Buton							1x2
	Tactile							
OUTPUT	Visual feedback		RGB LEDs analogic	RGB LEDs analogic	1 LED	5 LEDs analogic		
	Rumble							
	Speaker							
	Price $US	150	40	50	30	40	100	50

Tableau collaboratif disponible sur le site de l'association VRGeeks :
http://www.vrgeeks.org/wiki/device-classification

ANNEXE B
MATÉRIEL ET RÉSULTATS DÉTAILLÉS EXPÉRIMENTATION 1

Détail des moyennes et écarts types des résultats selon nos trois métriques
(Distance Euclidienne, Temps, Taux de complétion)

Distance Eucl. M	A1C1	A1C0	A0C1	A0C0
1G	0,0651	0,1196	0,0651	0,1196
2G	0,0814	0,0983	0,1164	0,1368
4G	0,0805	0,0969	0,0831	0,1182
6G	0,0982	0,1200	0,1341	0,1405
Distance Eucl. SD	A1C1	A1C0	A0C1	A0C0
1G	0,0436	0,1332	0,0436	0,1332
2G	0,0629	0,0520	0,1332	0,1042
4G	0,0560	0,0563	0,0730	0,1204
6G	0,0760	0,0767	0,1063	0,0999

Temps M	A1C1	A1C0	A0C1	A0C0
1G	2287,3095	4243,3810	2287,3095	4243,3810
2G	1819,7619	2934,3333	2350,6429	3427,2857
4G	1584,0595	1947,5952	3369,8214	2721,0119
6G	1923,1825	1986,0952	4617,1905	4359,0079
M	*1775,6680*	*2289,3413*	*3445,8849*	*3502,4352*
Temps SD	A1C1	A1C0	A0C1	A0C0
1G	2388,1672	4011,4742	2388,1672	4011,4742
2G	1090,8426	1813,5417	1738,9915	3569,9075
4G	877,2735	931,9816	4652,8440	2225,6386
6G	1833,8552	1512,0650	6111,2260	5351,6756

Complétion M	A1C1	A1C0	A0C1	A0C0
1G	1,0000	0,9825	1,0000	0,9825
2G	0,6281	0,6381	0,6630	0,5938
4G	0,3871	0,3725	0,5440	0,5178
6G	0,4351	0,3862	0,2521	0,2552
Complétion SD	A1C1	A1C0	A0C1	A0C0
1G	0,0000	0,1132	0,0000	0,1132
2G	0,3451	0,3701	0,3416	0,3495
4G	0,3298	0,3159	0,3556	0,3604
6G	0,3428	0,3137	0,3051	0,3388

Questionnaire sur l'expérimentation et l'utilisabilité

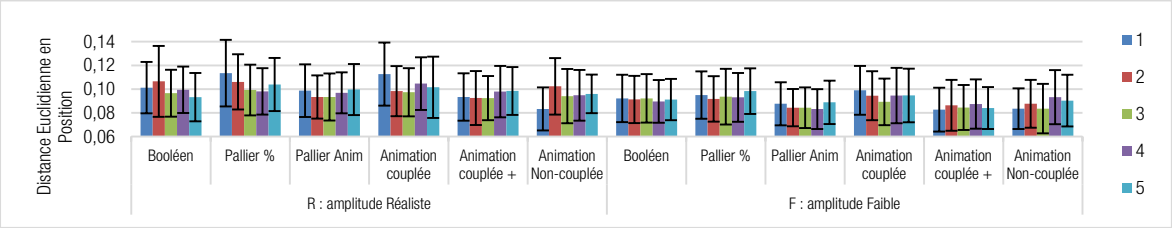
PRÉNOM : _____ AGE : ___ ans GAUCHER / DROITER DALTONISME OUI / NON
DATE : __ /__/__ DIO : ___ mm WIRT : ___%

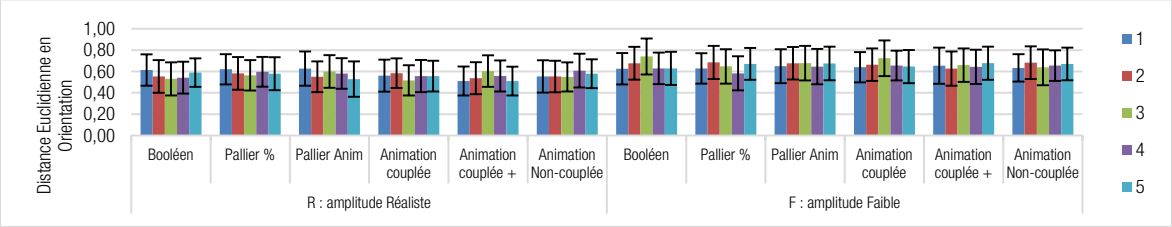
Critère	Questions à noter de 1 à 5	1 Mauvais ou pas d'accord / 5 Bon ou d'accord
Expertise	Mon niveau d'expertise informatique ?	1 2 3 4 5
Consigne	Les consignes étaient-elles claires ?	1 2 3 4 5
Utilisabilité	Cela m'a-t-il demandé beaucoup de concentration ?	1 2 3 4 5
	Cela m'a-t-il demandé beaucoup de précision ?	1 2 3 4 5
	Mes mouvements me paraissaient naturels ?	1 2 3 4 5
	Mes mouvements me paraissaient spontanés ?	1 2 3 4 5
	Utilisiez-vous des mnémoniques pour associer un geste à une boite ou une couleur ? Si oui donner un exemple : _____ _____ _____	oui / non
	J'étais confortablement installé(e) ?	1 2 3 4 5
	A l'enregistrement, 4 gestes me paraissent suffisants pour m'en rappeler à posteriori ?	1 2 3 4 5
	Mes mouvements étaient aisés à effectuer ?	1 2 3 4 5
	Les capteurs gênaient mes mouvements ?	5 4 3 2 1
	Était-ce agréable à utiliser ?	1 2 3 4 5
	J'ai ressenti une fatigue du bras ?	5 4 3 2 1
	Était-ce facile d'utilisation ?	1 2 3 4 5
Immersion	L'environnement virtuel me paraissait réaliste ?	1 2 3 4 5
	Les boites en bois me paraissaient réalistes ?	1 2 3 4 5
	Je me sentais présent dans la pièce virtuelle ?	1 2 3 4 5

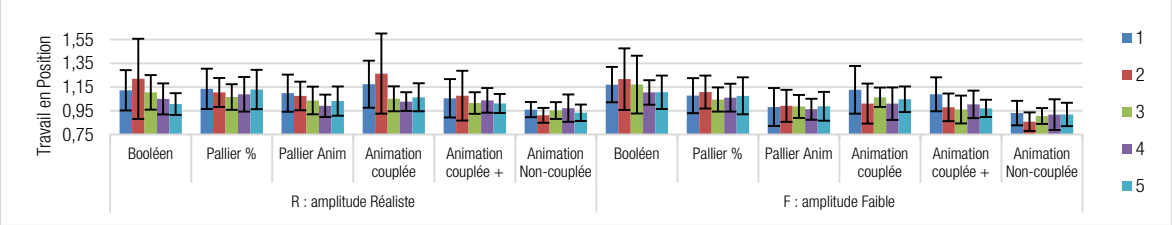
ANNEXE C

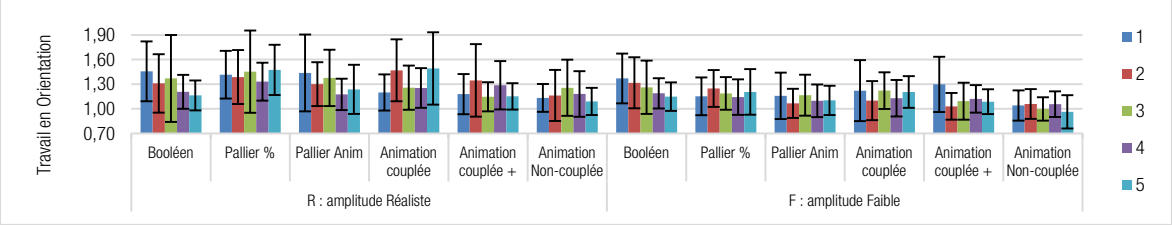
Matériel et résultats détaillés expérimentation 2

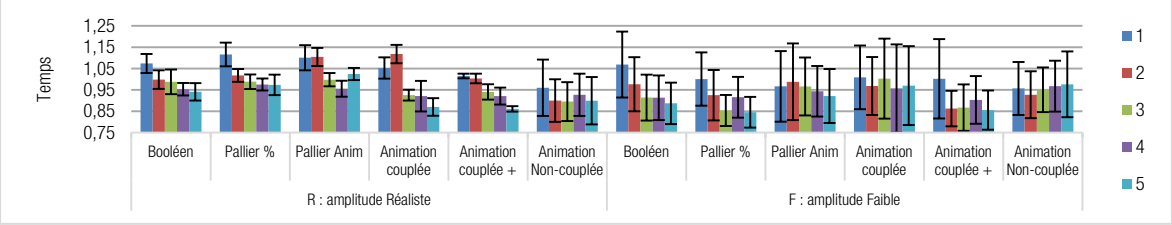
Détails des résultats de l'expérimentation 2

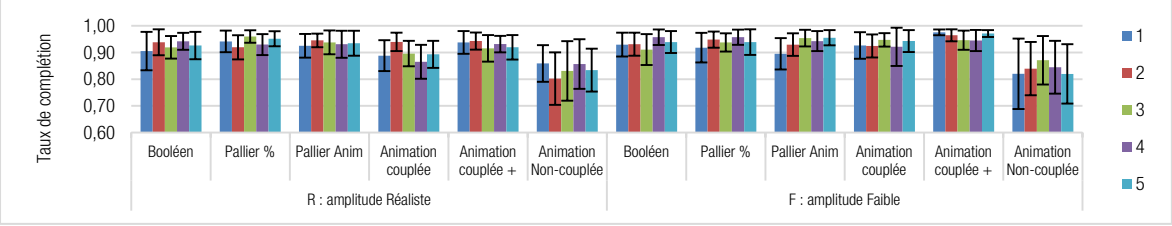












Résultats moyennés de l'expérimentation 2.
Chaque valeur correspond à la moyenne des performances de chaque métrique pour 20 utilisateurs et 3 objets

			AR Amplitude Réaliste						AF : Amplitude Faible					
		Répétition	Bool	Pa%	PaAn	AnCo	AnC+	TANC	Bool	Pa%	PaAn	AnCo	AnC+	TANC
Distance Euclidienne (orientation)	M	1	0,61	0,62	0,63	0,56	0,51	0,55	0,63	0,63	0,65	0,64	0,65	0,63
		2	0,55	0,58	0,55	0,58	0,54	0,55	0,68	0,68	0,68	0,66	0,63	0,68
		3	0,53	0,56	0,60	0,52	0,60	0,55	0,74	0,65	0,68	0,72	0,66	0,64
		4	0,54	0,60	0,58	0,56	0,56	0,61	0,63	0,58	0,65	0,66	0,64	0,65
		5	0,59	0,58	0,53	0,56	0,51	0,58	0,63	0,67	0,67	0,65	0,68	0,67
	SD	1	0,15	0,14	0,16	0,15	0,14	0,15	0,15	0,14	0,16	0,14	0,17	0,13
		2	0,15	0,15	0,14	0,14	0,15	0,15	0,15	0,15	0,15	0,15	0,16	0,15
		3	0,15	0,14	0,15	0,14	0,15	0,14	0,17	0,16	0,16	0,17	0,16	0,17
		4	0,15	0,14	0,14	0,15	0,15	0,16	0,15	0,16	0,17	0,14	0,16	0,14
		5	0,13	0,15	0,17	0,14	0,14	0,13	0,16	0,15	0,16	0,15	0,16	0,15
Distance Euclidienne (position)	M	1	0,10	0,11	0,10	0,11	0,09	0,08	0,09	0,09	0,09	0,10	0,08	0,08
		2	0,11	0,11	0,09	0,10	0,09	0,10	0,09	0,09	0,08	0,09	0,09	0,09
		3	0,10	0,10	0,09	0,10	0,09	0,09	0,09	0,09	0,08	0,09	0,08	0,08
		4	0,10	0,10	0,10	0,10	0,10	0,09	0,09	0,09	0,08	0,09	0,09	0,09
		5	0,09	0,10	0,10	0,10	0,10	0,10	0,09	0,10	0,09	0,09	0,08	0,09
	SD	1	0,02	0,03	0,02	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
		2	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
		3	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
		4	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
		5	0,02	0,02	0,02	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
Travail (orientation)	M	1	1,46	1,42	1,44	1,20	1,18	1,13	1,37	1,15	1,16	1,22	1,30	1,04
		2	1,31	1,39	1,30	1,47	1,35	1,16	1,32	1,25	1,07	1,10	1,03	1,06
		3	1,37	1,45	1,38	1,26	1,15	1,26	1,26	1,19	1,17	1,22	1,09	1,00
		4	1,21	1,33	1,18	1,25	1,29	1,18	1,19	1,14	1,10	1,13	1,12	1,06
		5	1,16	1,47	1,24	1,49	1,15	1,09	1,15	1,21	1,10	1,20	1,09	0,96
	SD	1	0,37	0,29	0,47	0,22	0,25	0,17	0,30	0,23	0,28	0,37	0,34	0,18
		2	0,36	0,33	0,27	0,38	0,44	0,31	0,31	0,22	0,18	0,24	0,16	0,18
		3	0,53	0,50	0,34	0,27	0,18	0,34	0,33	0,20	0,25	0,22	0,23	0,14
		4	0,21	0,23	0,19	0,24	0,29	0,28	0,18	0,22	0,20	0,22	0,17	0,16
		5	0,18	0,31	0,30	0,44	0,16	0,17	0,17	0,28	0,18	0,19	0,15	0,20
Travail (position)	M	1	1,12	1,13	1,10	1,17	1,05	0,96	1,17	1,08	0,98	1,13	1,09	0,93
		2	1,22	1,10	1,08	1,26	1,08	0,91	1,22	1,11	0,99	1,01	0,98	0,86
		3	1,10	1,07	1,04	1,05	1,02	0,95	1,17	1,04	0,99	1,06	0,96	0,91
		4	1,05	1,09	0,99	1,03	1,04	0,97	1,10	1,06	0,96	1,01	1,00	0,92
		5	1,01	1,13	1,03	1,06	1,01	0,93	1,11	1,08	0,99	1,05	0,97	0,92
	SD	1	0,17	0,17	0,16	0,20	0,16	0,06	0,15	0,15	0,16	0,20	0,14	0,10
		2	0,34	0,12	0,12	0,34	0,21	0,06	0,26	0,14	0,14	0,17	0,12	0,08
		3	0,15	0,11	0,12	0,11	0,09	0,07	0,24	0,10	0,10	0,08	0,12	0,07
		4	0,13	0,15	0,09	0,08	0,10	0,12	0,10	0,12	0,09	0,14	0,12	0,13
		5	0,09	0,17	0,12	0,12	0,08	0,07	0,14	0,16	0,12	0,11	0,07	0,10
Taux de complétion	M	1	0,91	0,94	0,93	0,89	0,94	0,86	0,93	0,92	0,90	0,93	0,98	0,82
		2	0,94	0,92	0,95	0,94	0,94	0,80	0,93	0,95	0,93	0,92	0,96	0,84
		3	0,92	0,96	0,94	0,90	0,92	0,83	0,91	0,94	0,95	0,95	0,95	0,87
		4	0,94	0,93	0,93	0,87	0,93	0,86	0,96	0,96	0,94	0,92	0,95	0,84
		5	0,93	0,95	0,93	0,89	0,92	0,83	0,94	0,94	0,96	0,94	0,97	0,82
	SD	1	0,07	0,04	0,04	0,06	0,04	0,07	0,04	0,06	0,06	0,05	0,01	0,13
		2	0,05	0,05	0,03	0,03	0,03	0,10	0,04	0,03	0,04	0,04	0,02	0,10
		3	0,04	0,02	0,04	0,05	0,05	0,11	0,06	0,03	0,03	0,03	0,04	0,09
		4	0,03	0,04	0,05	0,06	0,03	0,09	0,03	0,03	0,04	0,07	0,04	0,10
		5	0,05	0,03	0,05	0,05	0,05	0,08	0,04	0,05	0,03	0,04	0,01	0,11
Temps	M	1	1,07	1,12	1,10	1,05	1,02	0,96	1,07	1,00	0,97	1,01	1,00	0,96
		2	1,00	1,02	1,10	1,12	1,00	0,90	0,98	0,93	0,99	0,97	0,86	0,93
		3	0,99	0,99	1,00	0,93	0,94	0,90	0,91	0,85	0,97	1,00	0,87	0,95
		4	0,95	0,98	0,96	0,92	0,92	0,93	0,91	0,92	0,94	0,96	0,90	0,97
		5	0,94	0,97	1,02	0,87	0,86	0,90	0,89	0,85	0,92	0,97	0,86	0,98
	SD	1	0,04	0,06	0,06	0,05	0,01	0,13	0,15	0,13	0,17	0,15	0,19	0,12
		2	0,04	0,03	0,04	0,04	0,02	0,10	0,13	0,12	0,18	0,14	0,08	0,11
		3	0,06	0,03	0,03	0,03	0,04	0,09	0,11	0,07	0,14	0,19	0,11	0,10
		4	0,03	0,03	0,04	0,07	0,04	0,10	0,10	0,10	0,12	0,21	0,11	0,12
		5	0,04	0,05	0,03	0,04	0,01	0,11	0,10	0,07	0,13	0,18	0,09	0,15

Tables des résultats des ANOVA selon nos trois métriques et les conditions de rétroaction, répétabilité et d'amplitude

Tests of Between-Subjects Effects

Dependent Variable: **Travail**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7,535[a]	49	,154	,900	,669
Intercept	1334,671	1	1334,671	7811,133	0,000
Amplitude	1,839	1	1,839	10,763	,001
Rétroaction	1,841	4	,460	2,694	,030
Répétition	,895	4	,224	1,310	,264
Amplitude * Rétroaction	,628	4	,157	,918	,452
Amplitude * Répétition	,259	4	,065	,378	,824
Rétroaction * Répétition	1,068	16	,067	,391	,985
Amplitude * Rétroaction * Répétition	1,005	16	,063	,368	,989
Error	162,324	950	,171		
Total	1504,530	1000			
Corrected Total	169,859	999			

Tests of Between-Subjects Effects

Dependent Variable: **Temps**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4,841[a]	49	,099	1,012	,452
Intercept	933,710	1	933,710	9563,514	0,000
Amplitude	,734	1	,734	7,518	,006
Rétroaction	,604	4	,151	1,546	,187
Repétition	2,094	4	,523	5,361	,000
Amplitude * Rétroaction	,348	4	,087	,891	,469
Amplitude * Repétition	,207	4	,052	,530	,714
Rétroaction * Repétition	,331	16	,021	,212	1,000
Amplitude * Rétroaction * Repétition	,523	16	,033	,335	,993
Error	92,751	950	,098		
Total	1031,302	1000			
Corrected Total	97,592	999			

Tests of Between-Subjects Effects

Dependent Variable: **Complétion**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,456[a]	49	,009	1,193	,175
Intercept	870,486	1	870,486	111639,598	0,000
Amplitude	,056	1	,056	7,150	,008
Rétroaction	,111	4	,028	3,559	,007
Répétition	,025	4	,006	,796	,528
Amplitude * Rétroaction	,061	4	,015	1,947	,101
Amplitude * Répétition	,019	4	,005	,610	,655
Rétroaction * Répétition	,104	16	,007	,836	,644
Amplitude * Rétroaction * Répétition	,080	16	,005	,641	,852
Error	7,407	950	,008		
Total	878,350	1000			
Corrected Total	7,863	999			

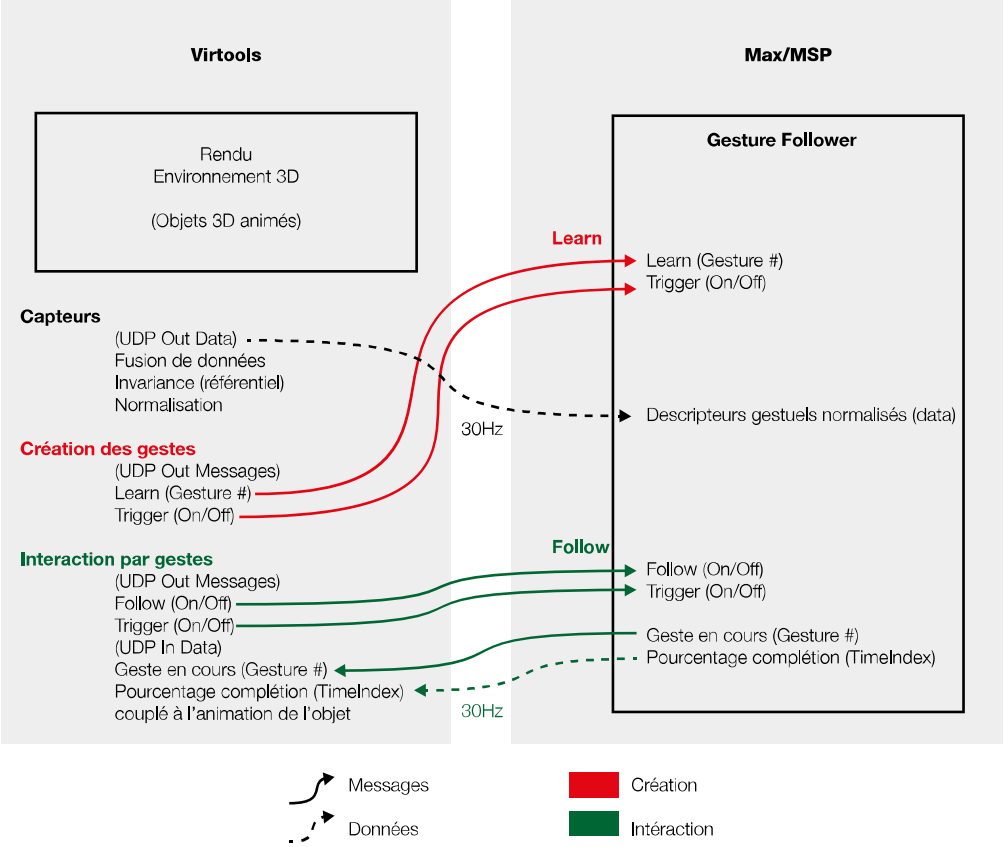
Nombre d'échantillons pour les analyse ANOVA et tests *post hoc* (Tukey)

		échantillons pour les 6 rétroactions	DDL	échantillons pour les 5 rétroactions couplées	DDL
Condition d'amplitude	AR	600	1	500	1
	AF	600		500	
Conditions de Rétroactions	Bool	200	5	200	4
	Pa%	200		200	
	PaAn	200		200	
	AnCo	200		200	
	AnC+	200		200	
	TANC	200		-	
Répétitions	1	240	4	200	4
	2	240		200	
	3	240		200	
	4	240		200	
	5	240		200	

Questionnaire de préférences subjectives des rétroactions (questions posées après chaque série de 15 gestes par rétroaction, notes sur échelle Likert de 1 à 5)

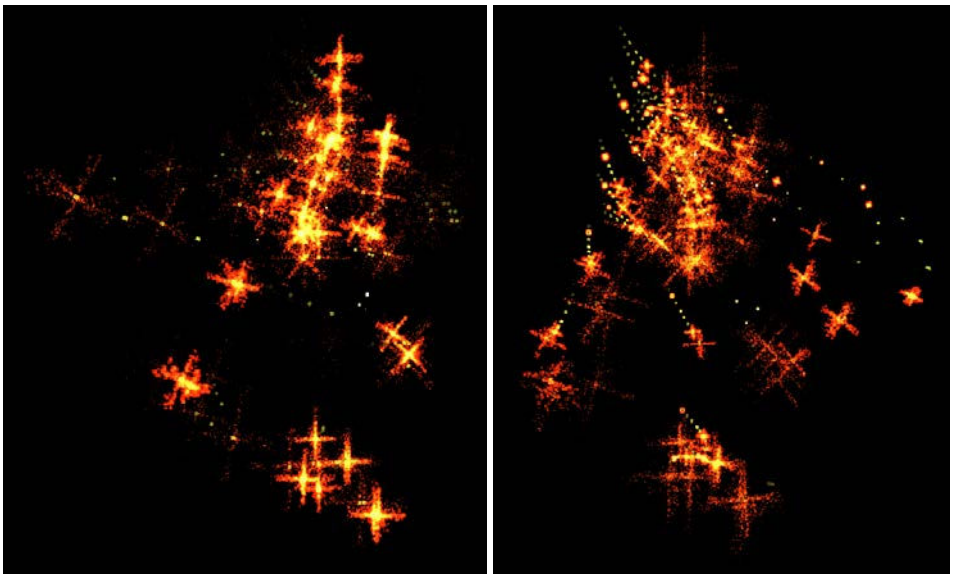
Rétroactions	Bool		Pa%		PaAn		AnCo		AnC+		TANC	
Conditions d'amplitude du geste	Réaliste	Faible	Réaliste	Faible	Réaliste	Faible	Réaliste	Faible	Réaliste	Faible	Réaliste	Faible
Réalisme du mouvement par rapport à une situation réelle (1 pas réaliste, 5 réaliste)												
Réactivité de la rétroaction (1 pas réactif, 5 réactif)												
Motivation à répéter le geste (1 pas motivant, 5 motivant)												
Agréable à utiliser (1 pas agréable, 5 agréable)												
Facile à utiliser (1 difficile, 5 facile)												

Communication réseau UDP (messages et données) entre logiciels 3DVia Virtools 5.0
et Cycling'74 Max/MSP 6 (extension GestureFollower)



ANNEXE D
TESTS UTILISATEURS GRAND PUBLIC FUTUR EN SEINE 2013

Environnement virtuel proposé pour l'activité dans une cuisine virtuelle



Environnement virtuel proposé pour l'activité créative

Interaction basée sur des gestes définis par l'utilisateur :
Application à la réalité virtuelle

RÉSUMÉ : Ces travaux de recherche proposent une nouvelle méthode d'interaction gestuelle. Nous nous intéressons en particulier à deux domaines d'application : la thérapie à domicile par la réalité virtuelle et les arts scéniques numériques. Nous partons du constat que les interfaces standardisées ne sont pas adaptées à ces cas d'usage, car elles font appel à des gestes prédéfinis et imposés. Notre approche consiste à laisser la possibilité à l'utilisateur de faire apprendre ses gestes d'interaction au système. Cela permet de prendre en compte ses besoins et aptitudes sensorimotrices. L'utilisateur réutilise ensuite son langage gestuel appris par le système pour interagir dans l'environnement virtuel. Cette approche pose des questions de recherche centrées sur la mémorisation, la rétroaction visuelle et la fatigue. Pour aborder ces trois aspects, nous étudions d'abord le rôle de l'affordance visuelle des objets et celui de la colocalisation dans la remémoration d'un vocabulaire de gestes. Ensuite, nous évaluons l'influence de différents types de rétroactions visuelles sur l'évolution des gestes répétés par l'utilisateur dans une série de tâches de manipulation. Nous comparons également les performances entre des gestes d'amplitude réaliste et des gestes d'amplitude faible pour effectuer la même action. Aussi, nous attachons une importance à rendre l'interface accessible en utilisant des dispositifs bas coûts et peu intrusifs. Nous explorons les moyens de pallier les contraintes techniques liées aux systèmes peu performants. Pour cela, nous avons conduit des expériences où plus de six mille gestes proposés par une quarantaine d'utilisateurs ont été étudiés.

Mots clés : réalité virtuelle, facteur humain, interaction gestuelle, 3DUI, apprentissage artificiel

User-Defined Gestural Interaction for Virtual Reality

ABSTRACT: In this thesis, we propose and evaluate new gestural interfaces for 3DUI. This work is motivated by two application cases: the first one is dedicated to people with limited sensory-motor abilities for whom generic interaction methods may not be adapted. The second one is artistic digital performances, for which gesture freedom is part of the creative process. For those cases, a standardized approach is not possible and thus user-specific or dedicated interfaces are needed. We propose a user-defined gestural interaction that allows the user to make the system learn the gestures that he has created, in a specific phase, prior to using the system. Then, the user reuses his created gestures to interact in the virtual environment. This approach raises research questions about memorization of gestures, effects of fatigue and effects of visual feedbacks. To answer those questions, we study the memorization of user created gestures regarding the role of affordances and colocalization on gesture recall. Then, we study the role of different visual feedbacks on gesture repetitions for a set of manipulation tasks. We also compare full-collocated gestures to loose gestures with lower amplitude. Also, the approach has been designed to be affordable using low-cost devices. We explore solutions to deal with the lower data quality of such devices. The results of the user-studies are based on the analysis of six thousand gestures performed by forty subjects.

Keywords : virtual reality, human factors, gestural interaction, 3DUI, machine learning

MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY