Interaction naturelle avec une scène virtuelle de micromanipulation

Laura Cohen

▶ **To cite this version:**

Laura Cohen. Interaction naturelle avec une scène virtuelle de micromanipulation. Sciences de l'ingénieur [physics]. Université Pierre et Marie Curie, 2015. Français. NNT : . tel-01142158

HAL Id: tel-01142158

https://hal.sorbonne-universite.fr/tel-01142158

Submitted on 14 Apr 2015

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

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

Thèse

présentée à

L'Université Pierre et Marie Curie

par

Laura COHEN

pour obtenir le grade de

Docteur de l'Université Pierre et Marie Curie

Spécialité : Robotique

Interaction naturelle avec une scène virtuelle de micromanipulation

---

Soutenance prévue le 14 avril 2015

JURY

M.	A. FERREIRA	Professeur à l'École Nationale Supérieure d'Ingénieurs de Bourges	Rapporteur
M.	P. HÉNAFF	Professeur à l'École des Mines de Nancy	Rapporteur
M.	Y. AMIRAT	Professeur à l'Université Paris-Est Créteil	Examinateur
M.	M. CHETOUANI	Professeur à l'Université Pierre et Marie Curie	Examinateur
M.	S. HALIYO	Maître de conférences à l'Université Pierre et Marie Curie	Examinateur
M.	S. RÉGNIER	Directeur de thèse Professeur à l'Université Pierre et Marie Curie	Examinateur

Résumé

Le développement des micro et nanotechnologies a ouvert un champ nouveau pour visualiser et organiser la matière aux toutes petites échelles. L'adoption actuelle des systèmes de micromanipulation et de leurs interfaces reste cependant limitée par la complexité de leur utilisation. Ce travail se concentre sur la synthèse d'une interface naturelle pour interagir avec le micromonde. L'étude de la littérature montre un manque d'outils adaptés à la détection des décisions. L'opérateur doit ainsi apprendre un langage symbolique complexe pour communiquer avec l'interface. Pour dépasser ces limites, trois aspects sont abordés :

- Un simulateur d'une tâche typique de micromanipulation sous AFM est proposé. Ce dernier masque la complexité de la tâche réelle pour la rendre accessible à des non spécialistes.
- Pour détecter la décision de saisie ou de dépose d'un objet au sein du simulateur, un modèle non symbolique de prédiction de l'intention inspiré des sciences cognitives est proposé. Ce dernier exploite les invariants en vitesse du geste ciblé lorsque l'objet cible est connu.
- Dans des contextes plus réalistes, la cible ne peut pas être connue a priori. Une méthode de sélection est proposée, basée sur un champ de neurones dynamiques activé selon la direction de déplacement du geste et du regard. La bulle d'activité principale observée correspond au focus d'attention de l'utilisateur.

Les résultats obtenus montrent une amélioration qualitative et quantitative significative par rapport aux approches classiques de reconnaissance de gestes symboliques. Ce travail montre l'apport de nouvelles méthodes d'interaction non symboliques pour des interfaces centrées sur l'utilisateur.

Mots clés : Interaction, Intention, Focus d'attention, Micromanipulation, Modèles cognitifs, Réalité virtuelle

Abstract

Micromanipulation is an important tool for recent developments in microscale production, characterization and analysis. However, micromanipulation systems and interfaces are not widely used in the research field or in the industry. They remain unhandy and complex to manage from the user's point of view. Hence, the operator has to learn a complex symbolic language to communicate with the interface, like clicking on specific buttons or performing gestural or vocal commands. The objective of this work is to propose more natural interfaces dedicated to micromanipulation. To overcome these limitations, three aspects are addressed :

- An intuitive simulator for a typical micromanipulation task with an AFM cantilever is implemented. The complexity of the real system is masked. The objective is to introduce a simulator suitable for non-expert users.
- To detect the decision to grab or drop an object, a non symbolic model of intention prediction inspired by cognitive science is proposed. It exploits targeted gesture invariants on velocity assuming that the target is known.
- A selection method is proposed to adapt the achieved results to realistic contexts where the target position is not known a priori. This model exploits the direction of motion of the gesture and the gaze to activate a dynamic neural field. The main bubble of activity within the field corresponds to the user's focus of attention

The results show a significative quantitative and qualitative improvement compared with classical approaches to detecting decisions. This work shows the contribution of non symbolic interaction approaches for new user-centered interfaces.

Keywords : Interaction, Intention, Focus of attention, Micromanipulation, Cognitive models, Virtual reality

Remerciements

Les travaux présentés dans ce manuscrit ont été effectués à l'Institut des Systèmes Intelligents et de Robotique. Ils n'auraient pu être réalisés sans le soutien, les conseils et les remarques avisées de nombreuses personnes. Je souhaite exprimer ici ma gratitude à tous ceux qui ont participé, directement ou non, à cette thèse.

Je tiens tout d'abord à exprimer ma reconnaissance aux membres du jury qui ont accepté d'évaluer mes travaux de recherche. Je remercie mes deux rapporteurs, Antoine Ferreira et Patrick Hénaff pour leurs remarques, ainsi que pour le temps qu'ils ont consacré à cette thèse. Je remercie aussi Yacine Amirat pour avoir accepté de participer à l'évaluation de ce travail.

Je souhaite adresser tous mes remerciements à mon directeur de thèse Stéphane Régnier. Sa rigueur, son implication, ses qualités scientifiques ont été un moteur tout au long de cette recherche. Ce manuscrit de thèse doit beaucoup à ses nombreuses remarques et son investissement constant.
Je remercie mon encadrant Mohamed Chetouani pour la pertinence de ses remarques et la confiance qu'il m'a accordée au cours de ces trois années.
Je tiens également à remercier Sinan Haliyo pour ses conseils toujours avisés et ses nombreuses relectures et corrections de ce manuscrit, ainsi que son optimisme bienvenu dans les moments difficiles.

Un très grand merci à Sofiane Boucenna pour son aide et ses conseils. Nos nombreuses discussions scientifiques m'ont apporté beaucoup.

Merci à tous les doctorants du groupe MICROB pour leur soutien, et la solidarité dont ils ont fait preuve. Ils m'ont permis de passer ces trois ans dans une ambiance chaleureuse et sympathique. Je tiens également à remercier le groupe IMI2S pour les bons moments passés ensemble.

Je remercie particulièrement Soukeyna Bouchebout et Wilfried Dron pour leur soutien quotidien, leur présence m'a été plus que précieuse. Je tiens aussi à associer Nils Melchior à ces remerciements, à qui notre groupe amical doit beaucoup.

Le soutien de ma famille a été essentiel, merci à mes parents et grands-parents, ainsi qu'à Isabelle, Quentin et Louise. Je remercie également Sihem, Lucie et Natacha.

Enfin, je remercie Guillaume pour sa gentillesse, son soutien et sa présence.

Table des matières

Table des matières **i**

Table des figures **v**

Introduction générale **1**

1 Interfaces naturelles pour l'interaction en microrobotique **5**
1 Interagir avec le micromonde 6
1.1 La micromanipulation 6
1.2 Interagir avec une scène de micromanipulation 7
1.3 Interfaces de restitution en micromanipulation 7
1.4 Interfaces d'acquisition en micromanipulation 8
1.5 Vers une interface naturelle pour la micromanipulation 9
2 Les interfaces naturelles 10
2.1 Propriétés 10
2.2 Les symboles dans les interfaces 10
2.3 Sens et actions non symboliques 11
3 État de l'art des interfaces pour la micromanipulation 11
3.1 Interfaces de restitution 12
3.1.1 Interfaces de réalité virtuelle 12
3.1.2 Le retour haptique 12
3.2 Les interfaces d'acquisition 14
3.2.1 Les bras de téléopération 14
3.2.2 Les interfaces tactiles 14
3.2.3 Les gants numériques 15
3.2.4 Les interfaces par reconnaissance de gestes 15
3.3 Synthèse 17
4 Approche proposée 19

4.1 Les interfaces gestuelles à l'échelle macroscopique 19
4.1.1 L'interaction par langage gestuel 19
4.1.2 L'interaction par manipulation directe 21
4.1.3 L'interaction basée comportement 22
4.2 Système proposé 23
4.2.1 Analyser le comportement naturel de l'opérateur 23
4.2.2 Les modèles haut niveau du focus d'attention et de l'intention 24
4.2.3 Synthèse 24

2 Un système bas niveau d'évaluation de l'interface 27
1 La micromanipulation téléopérée par contact adhésif 28
1.1 Le principe de la micromanipulation par adhésion 28
1.2 Un simulateur physique pour l'évaluation 30
1.2.1 Le logiciel Blender 31
1.2.2 La main virtuelle 32
1.2.2.1 Modélisation de la main 32
1.2.2.2 Animations de la main 33
1.3 Couplage main virtuelle-simulateur 34
1.3.1 Système maître-esclave pour la téléopération 34
1.3.2 Méthode de saisie/dépose 34
1.3.2.1 Phase de déplacement 35
1.3.2.2 Phase de saisie 35
1.3.2.3 Phase de dépose 35
1.3.3 Retour visuel sur l'échec/le succès de la tâche 35
1.4 Couplage utilisateur-main virtuelle 37
1.4.1 La méthode d'acquisition avec le capteur Kinect 37
1.4.1.1 Choix de la librairie 38
1.4.2 La détection des déplacements avec la Kinect 38
1.4.3 La détection des décisions de l'utilisateur 39
2 Détection des décisions par reconnaissance de gestes 39
2.1 Taxinomie du geste humain 39
2.1.1 Définition du geste 39
2.1.2 Le geste dans l'IHM 40
2.2 La méthode par reconnaissance de gestes 41
3 Expériences utilisateur 42
3.1 Méthode d'évaluation 43
3.2 Protocole du test utilisateur 43
3.3 Résultats quantitatifs 44
3.4 Résultats qualitatifs 45
3.5 Analyse des résultats 45
4 Conclusion 46

3 Une interface basée sur la prédiction de l'intention 49
1 Sélection des signaux de bas niveau pour modéliser l'intention 50

1.1 Définition fonctionnelle de l'intention 50
1.2 État de l'art des signaux caractéristiques de l'intention 51
1.2.1 Étude de l'acteur 52
1.2.2 Étude de l'observateur 53
1.3 Les signaux invariants du geste ciblé humain 55
1.3.1 Profil gaussien de la vitesse 55
1.3.2 Loi d'isochronie du mouvement 56
2 Étude de l'influence de l'intention sur les invariants du geste 57
2.1 Influence de l'intention sur la cinématique du geste 58
2.2 Influence de l'utilisateur sur la cinématique du geste 60
2.3 Reconnaissance de l'intention 60
2.4 Conclusion 61
3 Modèle haut niveau de prédiction de l'intention 62
3.1 Modèle cognitif de prédiction de l'intention par un observateur humain 62
3.2 Modèle de prédiction de l'intention de saisie/dépose pour la micromanipulation 64
3.2.1 Influence du contexte de la micromanipulation 64
3.2.2 Estimation des prédicteurs de la vitesse 64
3.2.3 Application à la saisie et la dépose d'une microsphère sur un substrat 65
3.3 Conclusion 67
4 Évaluation de l'interface pour la micromanipulation 67
4.1 Protocole du test utilisateur 67
4.2 Résultats comparatifs de la reconnaissance de gestes et de la prédiction de l'intention 68
4.2.1 Résultats quantitatifs 68
4.2.2 Résultats qualitatifs 69
5 Conclusion 70

4 Estimation du focus d'attention pour des scènes multicibles 73
1 Les mécanismes attentionnels 74
1.1 Définitions 74
1.1.1 L'attention sélective et le focus d'attention 74
1.1.2 La notion de saillance 75
1.1.2.1 La saillance orientée "bottom-up" 75
1.1.2.2 La saillance orientée "top-down" 76
1.2 Approche proposée 76
2 Les indices bas niveau du focus d'attention 77
2.1 Le regard et la pose du visage 77
2.2 Le geste 78
2.3 Choix des signaux pour notre interface 78
3 Modélisation haut niveau du focus d'attention 79
3.1 Contraintes et objectifs 80
3.1.1 Contraintes liées à la scène et au contexte 80
3.1.2 Contrainte de prédictivité 81

3.2 État de l'art des modèles de l'attention conjointe 81
3.3 Les champs neuronaux dynamiques 83
3.4 Estimation d'une carte de saillance basée comportement 85
3.4.1 Le stimulus d'activation du champ 85
3.4.2 L'influence des neurones voisins 87
3.4.3 Le mécanisme d'hystérésis 88
3.4.4 Carte de saillance estimée 89
3.4.4.1 Choix de l'écart-type 89
3.4.4.2 Choix de la constante de temps pour l'hystérésis . 92
3.5 Sélection du focus d'attention à partir de la carte de saillance . . . 92
4 Évaluation du modèle d'estimation 93
4.1 Protocole expérimental 94
4.2 Cas discret 95
4.2.1 Évaluation de l'influence du nombre d'objets 95
4.2.2 Évaluation de l'influence de la difficulté de la tâche 97
4.3 Cas continu 98
5 Conclusion 99

Conclusions et perspectives **101**

Annexes **105**

A Discrétisation de l'équation des champs neuronaux dynamiques **105**

Bibliographie **107**

Liste des publications **119**

Table des figures

1.1 Téléopération des systèmes de micromanipulation. 7
1.2 L'utilisateur téléopère une plateforme de micromanipulation par l'intermédiaire d'un modèle virtuel pour améliorer le retour visuel et l'intuitivité [Sauvet 12] 8
1.3 Interface haptique avec retour visuel en 3D pour la manipulation d'une molécule de VIH [Bolopion 10a] 9
1.4 Les différents types de téléopérations avec réalité virtuelle 13
1.5 Trois exemples d'interfaces haptiques [Dimension] [Immersion] [Mohand-Ousaid 12] 13
1.6 Manipulation de pinces optiques avec une interface tactile [Grieve 09] . . . 14
1.7 Manipulation de pinces optiques avec un gant numérique (Cyberglove) [Park 07] 15
1.8 Interface de vision par ordinateur 16
1.9 Exemples de dictionnaires de gestes symboliques pour l'IHM [Gillian 14b] [Van den Bergh 11] 20
1.10 La surface de la main de l'utilisateur détectée avec la Kinect est représentée par des particules physiques qui interagissent avec des sphères virtuelles dans le simulateur. [Hilliges 12] [Van den Bergh 11] 21
1.11 Schéma d'interface naturelle pour la micromanipulation. 25

2.1 Principe de la caractérisation d'un échantillon par un microscope à force atomique et système réel 29
2.2 Saisie (en haut) et dépose (en bas) d'une microsphère avec la poutre d'un AFM par adhésion. F_{i-j} est la force d'adhésion entre i et j. 29
2.3 Téléopération d'un système de micromanipulation 30

2.4 Téléopération d'un simulateur de micromanipulation en réalité virtuelle. (1) Les translations et rotations de la main de l'utilisateur sont appliquées à la main virtuelle ainsi que ses décisions (saisie, dépose) (2) Un système maître-esclave est utilisé pour téléopérer la poutre de l'AFM dans le simulateur réaliste. (3) La réussite ou l'échec de la tâche de saisie/dépose est retourné à l'interface naturelle. (4) Un retour visuel est donné à l'utilisateur. 31
2.5 Anatomie de la main humaine et degrés de liberté des articulations. Les articulations interphalangienne sont notées IPP (inter-phalangienne proximale) et IPD (inter-phalangienne distale) 32
2.6 Animation de la main virtuelle pour la saisie et la dépose 33
2.7 Phase de déplacement 35
2.8 Phase de saisie 36
2.9 Phase de dépose 36
2.10 Articulations suivies par le SDK Micosoft Kinect 37
2.11 Exemple de dictionnaire de gestes proposé par [Zeller 97] 40
2.12 Taxinomie des gestes 41
2.13 Séquence de téléopération avec l'approche par reconnaissance de gestes main ouverte et main fermée 42
2.14 Pourcentage de succès des deux tâches avec la méthode par reconnaissance de gestes (à gauche) et durée moyenne d'une tâche (au milieu). Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. La courbe de droite montre l'évolution de la durée moyenne de l'ensemble des utilisateurs pour réaliser une tâche en fonction du temps 44
2.15 Résultats du test utilisateur SUS. Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. 45

3.1 L'intention a priori "Je veux manger cette pomme" déclenche une intention dans l'action de saisie qui provoque une action motrice du bras en direction de la pomme. 51
3.2 Tâche d'atteinte et de saisie pour évaluer si la cinématique du geste est dépendante de l'intention sociale ou individuelle (à gauche) et résultats expérimentaux (à droite) [Becchio 10] 52
3.3 Images extraites des vidéos montrées à l'observateur. A droite, seuls des points lumineux sont visibles. Ils correspondent aux articulations du pouce, de l'index et du poignet [Manera 10]. 54
3.4 Le profil gaussien invariant de la vitesse de la main pour les gestes ciblés. La moyenne est notée μ, l'écart type σ et le maximum v_{max}. L'unité de longueur par défaut du logiciel Blender est appelée "unité Blender". Elle n'a pas d'équivalent dans le monde réel. 56
3.5 Illustration du principe d'isochronie du mouvement pour les gestes ciblés. La courbe est tracée à partir de 60 gestes d'atteinte d'une cible dans le simulateur. 57

3.6 Le squelette est suivi avec le SDK du capteur Kinect. Lorsque la main est suffisamment proche de la bille, la saisie est déclenchée automatiquement. Une fois la bille saisie, lorsque la main est suffisamment proche de la cible, elle est déposée automatiquement. 58
3.7 Influence de la tâche sur les paramètres du profil gaussien des vitesses lors du geste ciblé. 59
3.8 Exemple de l'influence de deux utilisateurs sur les paramètres des gaussiennes pour les actions de saisie et de dépose de micro-objets virtuels . . . 60
3.9 Modèle cognitif computationnel de prédiction de l'intention 63
3.10 Modèle de prédiction de l'intention basé sur les invariants de la vitesse de la main lors du geste ciblé. $v_{pred}(t)$ est la vitesse prédite et $v(t)$ la vitesse réelle de la main acquise avec la Kinect. 65
3.11 Exemple d'une séquence de prédiction de l'intention 66
3.12 Expérience utilisateur d'évaluation de l'interface basée sur la prédiction de l'intention. 68
3.13 Pourcentage de succès des deux tâches avec la méthode de prédiction de l'intention et la reconnaissance de gestes (à gauche) et durée moyenne d'une tâche (au milieu). Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. La courbe de droite montre l'évolution de la durée moyenne de l'ensemble des utilisateurs pour réaliser une tâche en fonction du temps. La méthode de prédiction de l'intention est représentée en noir et la reconnaissance de gestes en bleu . . 69
3.14 Résultats du questionnaire utilisateur SUS. Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. 70

4.1 Exemples de cartes de saillance bottom-up établies à partir de deux types de caractéristiques : l'orientation (en haut) et la couleur (en bas) [Gao 07] 75
4.2 Les saccades visuelles sur une même scène dépendent de la tâche. Des saccades différentes sont observées selon la question posée à l'observateur [Yarbus 67]. 76
4.3 Modèle du focus d'attention basé sur le comportement de l'opérateur proposé. 77
4.4 La pose de la tête 80
4.5 Cas discret et cas continu du focus d'attention selon la tâche. 81
4.6 Trois méthodes d'activation du champ de neurones. L'intensité du bleu représente la valeur du stimulus en entrée du champ. 87
4.7 Une convolution spatiale est réalisée entre la différence de gaussiennes et l'activité neuronale pour modéliser l'influence des neurones voisins. 88
4.8 Exemple de l'influence de l'hystérésis sur l'activité d'un neurone. 89
4.9 Cartes de saillance obtenues pour les trois méthodes d'activation du champ neuronal : locale, prédictive et hybride 90

4.10 Cartes de saillance obtenues pour un même stimulus en entrée à partir de différents paramètres d'hystérésis et d'excitation-inhibition. L'écart-type de la gaussienne excitatrice est représenté en pourcentage de la distance entre deux neurones voisins. L'écart-type de la gaussienne inhibitrice est de 1.5 fois celui de la gaussienne excitatrice pour obtenir une fonction d'excitation proximale et d'inhibition distale comme proposé par Amari [Amari 77]. La constante de temps τ est un indicateur de l'effet mémoire dû à l'influence des instants précédents. 91
4.11 Méthode d'estimation du focus d'attention dans le cas discret et le cas continu à partir de la scène virtuelle et de la carte de saillance. 93
4.12 Test utilisateur réalisé pour plusieurs configurations : la disposition, le nombre d'objets et leur taille varient. 94
4.13 Définition du critère de prédictivité proposé dans ce travail. La courbe bleue représente l'estimation du focus d'attention pour un objet cible donné en fonction du temps lors d'une tâche. 95
4.14 Taux de réussite en fonction du nombre d'objets présents évalué sur l'ensemble de la durée de la tâche. La prédictivité est évaluée en pourcentage de la durée totale de la tâche. 96
4.15 Taux de réussite en fonction de la difficulté de la tâche évalué sur l'ensemble de la durée du geste. La prédictivité est évaluée en pourcentage de la durée totale de la tâche. 97
4.16 Évolution de la distance entre le focus d'attention estimé et la cible lors du geste d'atteinte. 98

Introduction générale

Le développement récent des micro et nanotechnologies a ouvert un champ nouveau pour visualiser et organiser la matière aux toutes petites échelles. Les applications sont nombreuses et touchent des domaines variés comme la synthèse de matériaux, l'électronique ou les technologies pour la santé. Cet essor implique un besoin important d'interagir avec les micro-objets. En particulier, un problème essentiel dans le micromonde est la manipulation d'objets virtuels ou physiques. Cette manipulation passe par un ensemble de tâches élémentaires, par exemple la saisie, le déplacement et la dépose de micro-objets. À ces échelles, la taille des objets, les champs de force complexes et non intuitifs et la sensibilité à l'environnement rendent la manipulation et son automatisation complexes. Une solution émerge avec la conception d'interfaces dédiées, pour qu'un opérateur humain puisse commander la tâche de manipulation dans un cadre manuel ou semi-automatisé. Les interfaces actuellement proposées sont peu naturelles et intuitives. Cette contrainte limite l'adoption des systèmes de micromanipulation et leurs interfaces associées.

Parmi les interfaces de micromanipulation, sont distinguées les interfaces d'acquisition, capables de transmettre des information de l'opérateur vers le micromonde et des interfaces de restitution, dont l'objectif est la communication des données du micromonde vers les modalités de perception humaines. Au niveau de la restitution, des interfaces de réalité virtuelle sont développées pour donner accès à l'être humain à la perception visuelle à ces échelles, et des interfaces haptiques donnent accès à la perception du toucher. Malgré ces avancées, il n'existe pas d'interface de micromanipulation couramment utilisée, ni dans l'industrie, ni dans la recherche. En particulier, l'étude de la littérature montre un manque d'outils intuitifs d'acquisition des décisions de l'opérateur. Cette propriété est le cœur de mon travail de thèse. Celui-ci cherche à définir une nouvelle classe d'interfaces naturelles dédiée à cette problématique.

Les méthodes actuelles de détection des décisions reposent pour la plupart sur l'utilisation de symboles comme des clics sur des bouton prédéfinis. Elles nécessitent d'apprendre et de retenir un langage symbolique pour communiquer les décisions à l'interface. Ces approches sont donc peu adaptées à des utilisateurs naïfs. Dans le domaine connexe des interfaces homme-machine à l'échelle macroscopique, de nouvelles solutions dédiées à l'interaction naturelle émergent. Les caméras de profondeur comme la Kinect disposent d'outils de suivi des gestes de l'opérateur. Ces dispositifs rendent possible une interaction basée directement sur une modalité de communication intuitive et peu contraignante pour l'utilisateur. Malgré cette avancée, plusieurs travaux remettent en question le naturel de ces nouvelles interfaces. Ces dernières exploitent généralement un langage gestuel pour communiquer. Elles semblent donc insuffisantes pour s'abstraire de la nécessité d'apprendre ce langage symbolique.

Pour dépasser ces limites, cette thèse s'attache à proposer de nouvelles **méthodes et outils non symboliques dédiés à la détection des décisions de l'opérateur d'une interface de micromanipulation**.

Le premier chapitre de ce travail spécifie les propriétés d'une interface naturelle dédiée à la micromanipulation. Un état de l'art propose une synthèse des interfaces existantes dans ce domaine pour les évaluer et les comparer selon leur caractère naturel. Il en ressort un manque d'outils intuitifs de détection des décisions. Les tâches typiques de la micromanipulation sont présentées et les différents aspects d'une interface sont caractérisés. Les propriétés d'une interface naturelle adaptée à ces tâches sont spécifiées. Des travaux illustratifs des interfaces existantes en micromanipulation sont analysés d'après les propriétés données. Cette étude montre les limites des interfaces actuelles dans l'analyse du comportement de l'opérateur lors de la prise de décisions.

Dans le chapitre 2, un système bas niveau d'évaluation de l'interface est proposé. Ce système simule une tâche de manipulation de microsphères avec la poutre d'un microscope à force atomique. Afin de fournir une interface destinée à des non-spécialistes, la complexité de l'interaction réelle est masquée. L'entrée de ce système correspond à la tâche que l'utilisateur veut réaliser. Dans ce but, une interface de restitution visuelle en réalité virtuelle est proposée. Cette interface inclut une main virtuelle afin de faciliter l'identification de l'effecteur avec la main réelle de l'utilisateur. Le système robotique réel ou simulé est téléopéré par l'intermédiaire de cette interface. L'enjeu est d'inclure les décisions de l'utilisateur dans ce système en automatisant les sous-tâches qui n'impliquent pas de décision. La difficulté principale de ce travail consiste à détecter la décision de l'utilisateur sans contraindre l'interaction. Une solution par reconnaissance de gestes "naturels" est proposée. Avec cette méthode, la saisie est déclenchée lorsqu'un geste statique "main fermée" est détecté et la dépose lorsqu'un geste "main ouverte" est reconnu. Cette approche est exploitée comme base de comparaison dans cette thèse.

Le chapitre 3 propose une nouvelle approche non symbolique et prédictive pour détecter les décisions de l'utilisateur et dépasser les limites de la reconnaissance de gestes. Cette approche est basée sur un modèle cognitif computationnel de reconnaissance de l'in-

tention [Oztop 05]. Il s'agit de s'inspirer du fonctionnement du cerveau humain, capable d'affecter du sens aux actions non symboliques d'autrui. De plus, l'être humain reconnaît le but d'une action de manière prédictive, avant que celle-ci ne soit terminée. Cet aspect prédictif est une propriété intéressante pour créer une interface avec des propriétés d'anticipation des actions de l'utilisateur. Une telle interface serait capable d'annuler le délai observé entre l'action réelle de l'utilisateur et l'animation de la main virtuelle. L'intention est une notion de haut niveau, elle ne peut pas être extraite directement à partir des capteurs. Elle est donc modélisée à partir de signaux comportementaux extractibles de bas niveau. La première difficulté consiste à sélectionner des signaux de bas niveau à la fois suffisamment invariants et caractéristiques de l'intention. Dans ce but, une étude des invariants du geste ciblé est réalisée. À partir de ces invariants, un modèle computationnel de prédiction de l'intention est proposé. Ce dernier repose sur un modèle récurrent issu des sciences cognitives. Ce système est finalement évalué dans le simulateur dédié d'un point de vue quantitafif et qualitatif.

Le chapitre 4 s'intéresse à généraliser le système proposé à des contextes d'interactions plus réalistes. En particulier, l'approche proposée précédemment considère des objets cibles connus dans leur environnement. Cependant, certaines tâches de manipulations impliquent des scènes plus complexes dans lesquelles il est impossible de déterminer a priori la cible de l'opérateur. Une stratégie de sélection sans a priori apparaît comme une solution prometteuse pour se rapprocher d'une scène réelle de micromanipulation. L'objectif de ce chapitre est la proposition d'une méthode pour réaliser cette sélection dans le cadre d'une interface naturelle. L'enjeu consiste à déterminer la cible visée sans qu'elle soit formulée de manière explicite. Une solution consiste à analyser le comportement naturel de l'opérateur lors de tâches de sélection. D'un point de vue cognitif, la sélection est effectuée par les mécanismes attentionnels. L'être humain est capable d'inférer le focus d'attention de l'autre pour réaliser des tâches collaboratives qui impliquent une attention conjointe. La modélisation de cette capacité constitue une piste pour déterminer le focus d'attention à partir d'indices comportementaux. Le modèle proposé doit être capable d'appréhender la sélection d'objets discrets lors de la saisie et de zones continues du substrat lors de la dépose. Les champs neuronaux dynamiques sont une solution prometteuse pour modéliser l'attention sélective de manière continue sur la surface du substrat. La compétition neuronale proposée par ce modèle possède des propriétés dynamiques adaptées pour réaliser le caractère sélectif du processus d'attention. Une contrainte de ce système consiste à anticiper les actions de l'opérateur. Dans ce but, différentes méthodes d'activation du champ sont proposées à partir d'indices comportementaux basés sur la main et le regard. Une expérience utilisateur est enfin mise en place pour explorer l'influence de ces différents stimuli sur l'estimation du focus d'attention de l'opérateur lors de la sélection d'objets.

Enfin, ce document se conclut par une analyse de ce travail et des perspectives associées.

Chapitre 1

Interfaces naturelles pour l'interaction en microrobotique

Le développement récent des micro et nanotechnologies a ouvert un champ nouveau pour visualiser et organiser la matière aux toutes petites échelles. Les applications sont nombreuses et touchent des domaines variés comme la synthèse de matériaux, l'électronique ou les technologies pour la santé. Cet essor implique un besoin important d'interagir avec les micro-objets. En particulier, un problème essentiel dans le micromonde est la manipulation de micro-objets virtuels ou physiques. Cette manipulation passe par un ensemble de tâches élémentaires, par exemple la saisie, le déplacement et la dépose de micro objets. À ces échelles, la taille des objets, les champs de force complexes et non intuitifs et la sensibilité à l'environnement rendent la manipulation et son automatisation complexes. Une solution émerge avec la conception d'interfaces dédiées. Les interfaces actuellement proposées sont peu naturelles et intuitives. Cette contrainte limite l'adoption des systèmes de micromanipulation et leur interface associée.

L'objet de ce chapitre est de **spécifier les propriétés d'une interface naturelle dédiée à la micromanipulation**. Un état de l'art propose une synthèse des interfaces existantes dans ce domaine pour les évaluer et les comparer selon leur caractère naturel. Il en ressort un manque d'outils intuitifs de détection des décisions de l'opérateur. Cette propriété est le coeur de ce travail de thèse. Celui-ci cherche à définir une nouvelle classe d'interface naturelle dédiée à cette problématique.

Dans une première partie, les tâches typiques de la micromanipulation sont présentées et les différents aspects d'une interface sont caractérisés. La deuxième partie spécifie les propriétés d'une interface naturelle adaptée à ces tâches. Des travaux illustratifs des interfaces existantes en micromanipulation sont analysés dans une troisième partie à partir des

propriétés données. Cette étude montre les limites des interfaces actuelles dans l'analyse du comportement de l'opérateur lors de la prise de décisions. Enfin, l'approche proposée dans ce travail est décrite pour appréhender cette problématique.

1 Interagir avec le micromonde

1.1 La micromanipulation

Dans le domaine de la micromanipulation, certaines applications sont automatisables. Ces applications concernent des tâches simples, dans des conditions contrôlées lorsque l'environnement est connu a priori. Ainsi, l'automatisation est limitée à des tâches répétitives de caractérisation ou d'assemblage, lorsqu'un grand nombre d'échantillons similaires doivent être traité [Fatikow 07] [Agnus 13]. Le caractère interactif de la manipulation ouvre une classe nouvelle d'application. Par exemple, la micromanipulation avec un micro préhenseur et une interface haptique [Mohand-Ousaid 14] ou par pince optique [Pacoret 13] sont des techniques très souvent utilisées dans notre laboratoire. Le premier exemple s'appuie sur un système physique de manipulation sans contact. Le second exploite des champs de forces pour déplacer un objet piégé dans un piège optique. Pour ce dernier, un retour d'effort est transmis à l'utilisateur, qui peut ainsi toucher le micromonde. Une autre application est la manipulation de micro-objets virtuels comme dans les simulateurs moléculaires pour la conception de nouveaux principes actifs. De même, certaines tâches impliquent une prise de décision complexe par un expert humain. Par exemple, une expertise humaine peut être nécessaire pour réaliser le séquençage de sous-tâches automatisables.

Parmi les tâches qui nécessitent l'expertise d'un opérateur pour manipuler des micro-objets physiques ou virtuels, trois champs d'applications principaux sont distingués : la manipulation, l'exploration et la caractérisation.

La manipulation Il s'agit de tâches qui impliquent le déplacement d'un objet virtuel dans son environnement. Les actions unitaires impliquées dans cette tâches sont la sélection, la saisie, le déplacement et la dépose sur le site cible.

L'exploration Contrairement à la manipulation, l'exploration n'implique pas de déplacement L'exploration haptique exploite des bras à retour d'effort pour retransmettre des forces micrométriques à l'opérateur. L'objectif est de comprendre la physique mise en jeu à cette échelle.

La caractérisation Elle consiste à obtenir des informations sur la topologie d'un substrat ou d'un objet. L'opérateur déplace l'outil de mesure sur l'objet à caractériser pour ressentir par exemple sa forme.

Le tâches unitaires impliquées dans l'ensemble de ces champs d'applications sont la

saisie, le **déplacement** et la **dépose**. Lorsque plusieurs objets peuvent être saisis, une tâche unitaire de **sélection** d'objets ou d'outils est aussi nécessaire. Dans ce travail, l'ensemble de ces tâches est qualifié au sens large de "micromanipulation" et nous nous concentrerons sur la manipulation de micro-objets.

1.2 Interagir avec une scène de micromanipulation

Pour réaliser ces tâches, il est essentiel de fournir à l'opérateur une interface avec le micromonde. Créer une interface consiste à mettre en œuvre des systèmes pour donner à l'opérateur la possibilité d'agir dans le micromonde et de percevoir ses propriétés sous différentes formes. L'information circule dans les deux sens : de l'homme vers le micromonde et réciproquement. Ces interfaces peuvent être séparées selon deux aspects : les interfaces d'acquisition (la plus classique est le couple clavier-souris) et les interfaces de restitution comme les écrans. La figure 1.1 montre le sens de transmission des informations au sein de l'interface. La restitution fournit à l'utilisateur un retour sur l'état de la sortie du système microrobotique. Celui-ci peut ajuster son entrée avec le système d'acquisition. Ainsi, l'opérateur ferme la boucle du système.

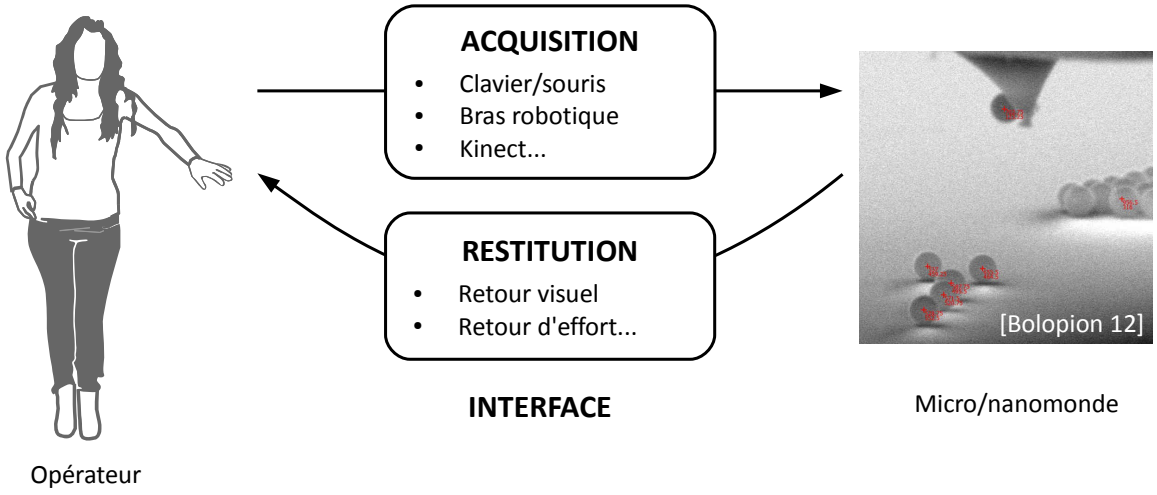


Figure 1.1 - *Téléopération des systèmes de micromanipulation.*

1.3 Interfaces de restitution en micromanipulation

Les interfaces de restitution donnent à l'utilisateur la capacité de percevoir le micromonde à son échelle. Ces interfaces doivent être capables de transmettre des informations interprétables par les modalités sensorielles humaines. De nombreuses solutions sont proposées dans la littérature qui exploitent plusieurs de ces modalités [Bolopion 13a]

[Haag 14]. Pour des tâches de manipulation ou d'assemblage, les sens principalement impliqués sont la vue et le toucher.

Le sens du toucher est exploité avec des interfaces haptiques qui retranscrivent, à l'échelle humaine, les forces micrométriques pour de la micromanipulation avec retour d'effort. À ces échelles, le retour visuel direct est aussi très limité puisqu'il est nécessaire de passer par un microscope optique ou électronique. Le point de vue est fixé, les informations de profondeur ne sont pas disponibles. L'interprétation des scènes est complexe en 3D. La réalité virtuelle apparaît comme une solution pour compenser ces lacunes : en représentant la scène réelle par une scène virtuelle (fig 1.2), l'opérateur peut changer à volonté de point de vue, mieux comprendre les scènes complexes et téléopérer de manière fine et précise un système de microrobotique [Sauvet 12]. De plus, il devient possible de simuler des scènes virtuelles, pour effectuer des tests ou mieux comprendre certains phénomènes sans avoir à recourir à une plateforme réelle de micromanipulation.

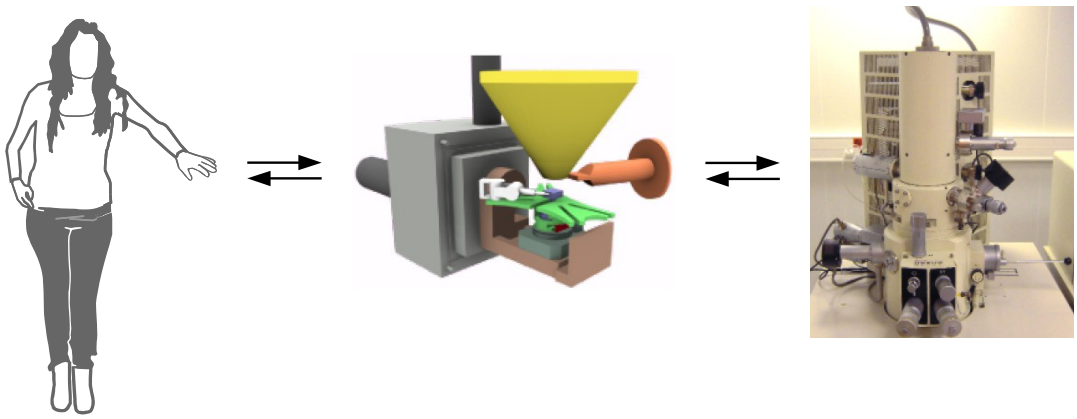


Figure 1.2 - *L'utilisateur téléopère une plateforme de micromanipulation par l'intermédiaire d'un modèle virtuel pour améliorer le retour visuel et l'intuitivité [Sauvet 12]*

1.4 Interfaces d'acquisition en micromanipulation

Le problème de l'acquisition de données issues de l'opérateur est complexe. Les signaux acquis doivent contenir des informations suffisantes sur la tâche que cherche à accomplir l'opérateur. Au niveau des interfaces de restitution, les signaux sont pensés pour être naturellement interprétables par l'homme. Contrairement à celles-ci, les interfaces d'acquisition sont limités à des signaux pauvres devant la complexité et la diversité des signaux communicatifs humains. L'interface la plus classique, basée sur le couple clavier/souris, a rapidement montré ses limites dans le cadre d'une interaction complexe en trois dimensions. Pour répondre à ce besoin, de nouvelles interfaces d'acquisition sont développées. Elles reposent souvent sur des bras manipulateurs [Bolopion 12]. Ces bras permettent de déplacer un curseur virtuel en 3D dans le micromonde et de spécifier une commande en utilisant un symbole, par exemple un clic pour saisir la molécule sur laquelle se situe le curseur.

Malgré ces avancées, il n'existe aujourd'hui pas d'interface de micromanipulation couramment utilisée dans l'industrie. Du point de vue de l'utilisateur, elles restent complexes d'utilisation, peu maniables et difficiles à maîtriser. Par exemple, les utilisateurs non-spécialistes évoquent la difficulté d'identifier la poignée physique et le curseur virtuel. De plus, ces dernières nécessitent un apprentissage puisque l'interaction se fait par le biais d'un langage basé sur des commandes prédéfinies, par exemple les clics sur différents boutons. De nombreux travaux de la littérature s'accordent sur la nécessité de créer des interfaces plus naturelles et intuitives. Ces derniers ne traitent pas cette question de manière approfondie du point de vue de l'utilisateur.

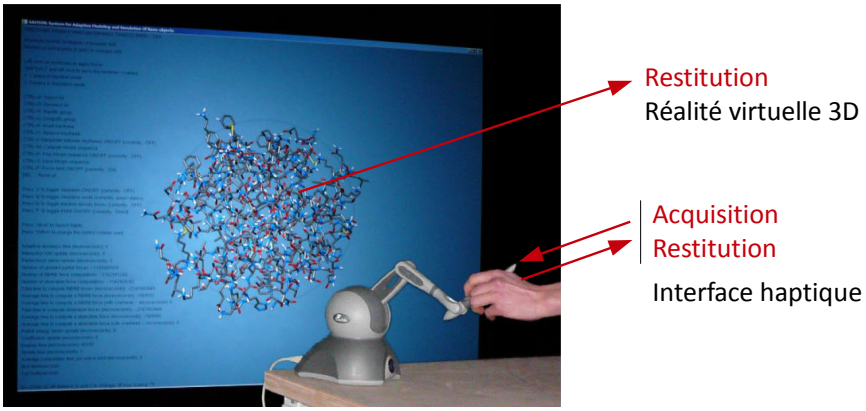


Figure 1.3 - *Interface haptique avec retour visuel en 3D pour la manipulation d'une molécule de VIH [Bolopion 10a]*

1.5 Vers une interface naturelle pour la micromanipulation

La discipline connexe des interfaces homme-machine (IHM) propose des solutions pour répondre à ce besoin. En particulier, de nouvelles interfaces plus proche d'une interaction avec le monde réel sont proposées. L'opérateur se repose sur des méthodes d'interaction qu'il maîtrise déjà et il n'a besoin d'aucune instruction pour utiliser l'interface. Ce travail adopte cette définition d'une interface naturelle, par opposition aux interfaces classiques qui exploitent un dictionnaire de symboles pour communiquer avec l'utilisateur. L'approche symbolique facilite la reconnaissance de la décision de l'utilisateur. En effet, l'être humain doit s'adapter au langage prédéfini par l'interface. L'enjeu principal de l'approche naturelle consiste à adopter le point de vue inverse, adapter l'interface au comportement naturel humain. Les signaux comportementaux humains sont complexes et possèdent une grande variabilité. Ces difficultés doivent être traitées pour reconnaître une décision de l'utilisateur d'après son comportement naturel. Cette approche implique une analyse du comportement humain. Pour mieux comprendre la manière dont l'homme interagit avec ses partenaires sociaux et son environnement, une solution consiste s'intéresser aux modèles issus des sciences cognitives. En modélisant le comportement humain de manière appropriée, il devient possible de créer une interface utilisable sans aucune consigne préalable.

2 Les interfaces naturelles

2.1 Propriétés

Le naturel d'une interface est défini comme le degré objectif avec lequel les actions (caractérisées par les mouvements, les forces, ou les parties du corps humain impliquées) utilisées pour effectuer la tâche dans l'interface correspondent aux actions utilisées pour effectuer cette tâche dans le monde réel [Bowman 12] [Malizia 12].

Dans le cadre de la micromanipulation, deux types de tâches sont distingués. Les tâches de déplacement sont généralement continues, comme les phases d'atteinte d'une cible ou de déplacement d'un objet. Il existe aussi des tâches discrètes qui correspondent à des actions telles que la saisie ou la dépose d'un objet.

Des solutions d'interfaces naturelles sont proposées dans la littérature pour les tâches de déplacement [Poupyrev 96] [Bowman 97]. Cependant, peu de solutions existent pour les actions discrètes (sélection, saisie, dépose). En effet, le déplacement de la main de l'utilisateur peut être suivi par un accéléromètre, une souris ou une Kinect. Le problème se complexifie lorsqu'il s'agit d'affecter du sens à une action discrète. Ces dernières correspondent à des décisions de l'opérateur. Il n'existe pas de caractéristique quantitative mesurable de ces décisions. Par exemple, aucun capteur ne peut extraire directement la décision d'un utilisateur de saisir ou de déposer un objet. Les tâches de sélection et de manipulation sont traitées dans quelques travaux de la littérature, mais il s'agit simplement de sélectionner ou saisir automatiquement l'objet lorsque l'utilisateur le touche [Bowman 12]. La question centrale est alors la réalisation de ces tâches discrètes tout en conservant l'aspect naturel du point de vue de l'utilisateur.

2.2 Les symboles dans les interfaces

D'après la définition donnée, proposer une interface naturelle implique de créer un lien qui n'est pas arbitraire entre l'action réelle et la commande gestuelle associée. D'après Saussure, un signe renvoie à autre chose que lui-même : "Le lien unissant le signifiant au signifié est arbitraire" [Benveniste 39]. Cette définition du signe s'applique au langage. Dans le cadre d'un langage gestuel, ce travail généralise la notion de signe sous le terme de symbole. Les interfaces basées sur des commandes gestuelles prédéfinies sont donc aussi des interfaces symboliques.

Par exemple, [Ren 13] proposent une interface par commande gestuelle. Dans ces travaux, l'opérateur déclenche la saisie d'un objet virtuel en levant la main gauche. Ce type d'interface exploite une modalité "naturelle" (le geste), mais ne correspond pas à l'action qu'il aurait effectuée dans la réalité. Ainsi, l'opérateur doit apprendre une commande gestuelle qui est reliée à la réalité de manière arbitraire. La communication basée sur des signes linguistiques est remplacée par une communication basée sur des symboles gestuels.

Dans la taxonomie du geste, les gestes communicatifs sont distingués des gestes techniques. Les interfaces symboliques sont particulièrement adaptées à la reconnaissance de gestes communicatifs, symboliques par nature. Cependant, la micromanipulation implique de reconnaître des actions techniques, sans but du point de vue de la communication. Dans ce cadre, une interface naturelle est définie dans ce travail comme une **interface non symbolique**.

2.3 Sens et actions non symboliques

Les interfaces dédiées à des tâches de manipulation doivent reconnaître des actions non-symboliques par nature, comme la saisie, le déplacement et la dépose d'objets. Pour ce type de tâches, la méthode classique consiste à remplacer le geste naturel par un geste symbolique de moindre complexité. Ce geste est plus contraint afin de faciliter la reconnaissance par une limitation de la variabilité. Pour créer une interface non symbolique, il faut donc affecter un sens à une action non symbolique sans que ce lien soit arbitraire. Ce problème peut être assimilé au problème de l'ancrage symbolique, illustré initialement par Searle [Searle 82] avec l'expérience de pensée de la chambre chinoise, puis décrit par Harnad [Harnad 90].

Dans cette thèse, deux approches d'interfaces "naturelles" seront proposées et comparées. La première approche consiste à reconnaître des gestes main ouverte/main fermée pour manipuler un objet virtuel. Ces gestes entretiennent un lien non arbitraire avec la tâche réelle de manipulation. Pour saisir un objet, il faut fermer la main autour de ce dernier et ouvrir les doigts pour le déposer. La deuxième approche consiste à s'inspirer du fonctionnement des capacités cognitives humaines pour comprendre les actions d'autrui. Ces approches sont détaillées dans la section 4.

3 État de l'art des interfaces pour la micromanipulation

Depuis les années 90, différents systèmes ont été développés pour donner à l'être humain la possibilité de toucher et de manipuler les échelles microscopiques. Ces systèmes exploitent différentes modalités au niveau de l'acquisition de la restitution. Dans un premier temps, les interfaces existantes pour acquérir et restituer des données sont présentées et comparées selon la définition d'une interface naturelle proposée dans la section précédente. Certaines de ces interfaces sont mixtes, orientées à la fois acquisition et restitution, comme les interfaces haptiques. Ces dernières sont par exemple capables de transmettre une information de force et position de l'utilisateur, mais aussi de lui restituer un retour d'effort. Des exemples illustratifs de systèmes complets (acquisition + restitution) en micromanipulation sont ensuite comparés en mettant l'accent sur les symboles employés.

3.1 Interfaces de restitution

3.1.1 Interfaces de réalité virtuelle

Le retour visuel direct en microscopie est souvent insuffisant pour réaliser un travail de précision. Il est compliqué, voire impossible, d'avoir une bonne compréhension des mouvements de l'outil manipulé, surtout dans des scènes complexes en trois dimensions. Pour pallier ce problème, il existe quelques exemples dans la littérature d'utilisation de la réalité virtuelle [Sulzmann 95, Sitti 98, Probst 07].

La réalité virtuelle comble les lacunes du retour visuel. Elle apporte la possibilité de disposer de plusieurs vues et de mieux appréhender la profondeur de la scène. Elle donne la possibilité de rendre des micro-objets concrets à l'échelle de l'opérateur. Les interfaces de réalité virtuelle sont donc particulièrement adaptées à des tâches complexe de micro-manipulation, et présentent l'avantage de pouvoir simuler des scènes réelles.

Deux types de téléopération différents sont envisagés avec ces outils de la réalité virtuelle (fig.1.4) :

- **la téléopération augmentée** : l'opérateur interagit avec une scène de réalité virtuelle qui représente la scène réelle. La simulation est couplée avec la plate-forme de micro-manipulation pour que les actions de l'utilisateur soient restituées pour manipuler le micro-objet réel. Ce mode de téléopération fournit à l'utilisateur des indications, pas directement disponibles par les capteurs, pour l'assister lors d'une tâche de manipulation.

- **la téléopération virtuelle** : l'opérateur interagit avec une scène de simulation virtuelle du micromonde. Cette scène joue le rôle de plateforme de tests pour des algorithmes ou des stratégies de manipulation. Un exemple est la manipulation par microscope à force atomique (AFM). Les scènes réelles sont tellement complexes que des simulations virtuelles sont utilisées pour tester différentes stratégies de manipulation. Une autre application notable est la simulation moléculaire. Un grand nombre de simulateurs sont aussi développés pour la pédagogie [Millet 13].

3.1.2 Le retour haptique

L'haptique désigne la science du toucher, par analogie avec l'acoustique ou l'optique. Au sens strict, l'haptique s'intéresse au toucher et aux phénomènes kinesthésiques. Les interfaces haptiques ou interfaces à retour d'effort font actuellement l'objet de très nombreux développement. Elles sont largement employées dans la plupart des systèmes de téléopération. L'haptique est l'un des moyens privilégiés d'interaction aux échelles micro et nanoscopiques. Elle restitue un effort que les opérateurs ont naturellement l'habitude

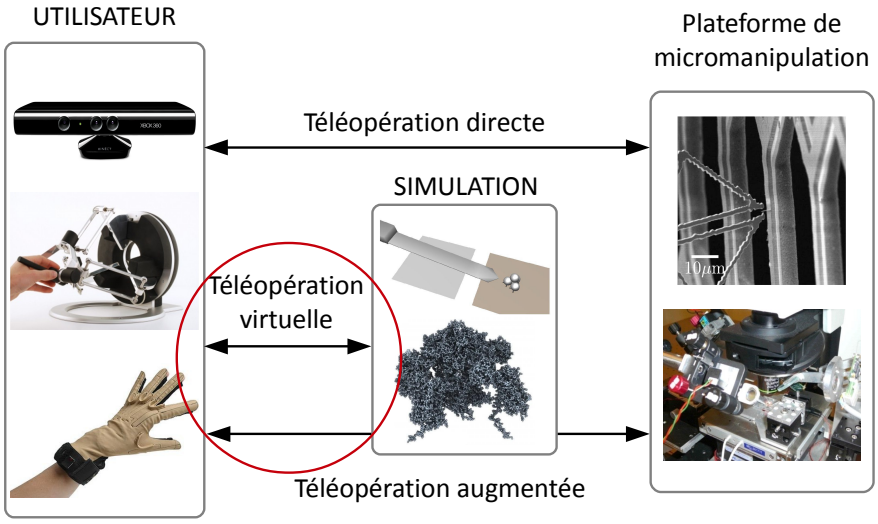


Figure 1.4 - *Les différents types de téléopérations avec réalité virtuelle*

de ressentir lors de manipulations directes à l'échelle macroscopique [Ferreira 06].

Plusieurs bras haptiques commerciaux disposent d'un retour d'effort, par exemple le Phantom [Immersion], l'Omega [Dimension] ou le Virtuose [Haption]. Celles-ci sont dédiées à l'interaction macroscopique. [Bolopion 12] propose une interface de microassemblage d'objets sphériques en trois dimensions avec un bras Omega. Ce système fournit un retour d'effort sur trois axes alors que les mesures s'effectuent uniquement sur l'axe vertical. Il existe aussi des interfaces non commerciales, spécialement dédiées à certaines tâches de micromanipulation [Mohand-Ousaid 12]. La particularité de cette interface vient de ses caractéristiques mécaniques qui visent à recouvrir l'intégralité de la perception haptique humaine. La figure 1.5 illustre ces exemples.

Bras haptique Omega | Bras haptique Phantom | Interface haptique dédiée développée à l'ISIR

Figure 1.5 - *Trois exemples d'interfaces haptiques [Dimension] [Immersion] [Mohand-Ousaid 12]*

Les applications pédagogiques sont aussi un champ d'application des interfaces haptiques. Elles donnent la possibilité aux étudiants d'améliorer leur compréhension du micromonde en incluant à leur raisonnement les notions de forces d'interaction ressenties à

l'aide du retour d'effort [Bivall 11] [Millet 13]. Cependant, ces interfaces ont des limites. Le bras haptique implique une couche supplémentaire entre l'utilisateur et l'objet simulé. Il est nécessaire pour l'utilisateur de passer par une phase d'apprentissage pour interagir avec la scène virtuelle. De plus, ce type d'interface doit faire face à des problèmes complexes comme la transparence et la stabilité. Les utilisateurs non spécialistes évoquent aussi la difficulté d'identifier la poignée physique et le curseur virtuel.

3.2 Les interfaces d'acquisition

3.2.1 Les bras de téléopération

Les bras de téléopération peuvent être de simples joysticks qui extraient une position en 3D, ou permettre une acquisition et un retour d'effort. Une interface haptique est un bras de téléopération avec retour d'effort. L'aspect restitution de ces interfaces est détaillé dans la partie précédente.

3.2.2 Les interfaces tactiles

Le succès des interfaces tactiles multi-touch ces dernières années a entrainé la création d'une nouvelle génération d'interfaces plus intuitives. Parmi ces systèmes, les écrans tactiles combinent une acquisition tactile et une restitution visuelle. Plusieurs travaux exploitent ce type d'interfaces en micromanipulation, en particulier pour la téléopération de pinces optiques. Par exemple, un iPad est utilisé pour la manipulation de pinces optiques à 5 doigts [Bowman 11] , ou encore d'une table multi-touch [Grieve 09]. La position des doigts est directement utilisée pour déplacer l'objet piégé. Ainsi, les déplacements sont réalisés de manière non symbolique : une stricte équivalence apparaît entre le déplacement réel de la main et le déplacement de l'effecteur. La saisie est déclenchée lorsque l'extrémité digitale de l'opérateur entre en contact avec l'interface. La dépose est réalisée lorsque l'opérateur relâche ce contact. Ainsi, ces interfaces réalisent ces actions unitaires sans utiliser de symboles prédéfinis.

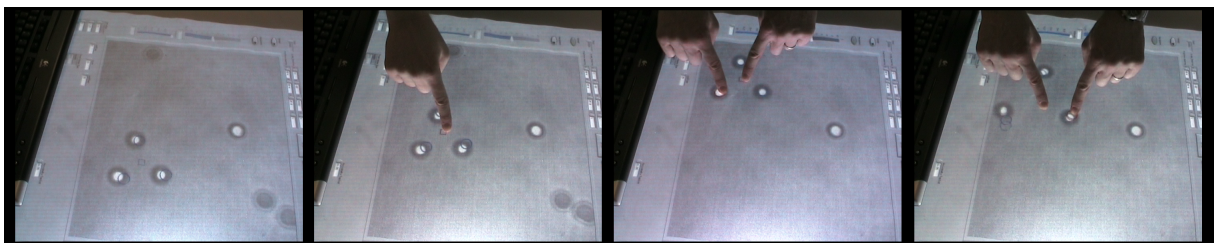


Figure 1.6 - *Manipulation de pinces optiques avec une interface tactile [Grieve 09]*

Les écrans tactiles sont bien adaptés aux problèmes à deux dimensions, et résolvent le problème soulevé par les interfaces haptiques de l'identification du point d'action de la

main avec le système téléopéré. La profondeur peut être gérée par un curseur [Grieve 09] ou encore par une fonction zoom en éloignant ou rapprochant deux doigts [Bowman 11]. Ainsi, l'adaptation de ces solutions à une interaction en 3 dimensions implique d'introduire un langage symbolique (curseur ou fonction zoom). Elles ne sont donc pas satisfaisantes pour répondre à la définition d'une interface naturelle donnée dans la deuxième partie.

3.2.3 Les gants numériques

Pour pouvoir exploiter au maximum les possibilités de manipulation de l'être humain dans un espace en trois dimensions, des interfaces basées sur des gants numériques sont proposées. Les gants numériques embarquent des capteurs qui renvoient les angles des articulations des doigts. Les mouvements de la main sont numérisés, l'utilisateur peut saisir un objet virtuel et le manipuler. Il existe des interfaces de manipulation de pinces optiques qui exploitent les gants numériques [Park 07] ou encore de simulation moléculaire [Ai 98]. Les gants numériques optimisent les capacités de manipulation de la main humaine, puisqu'ils retranscrivent tous ses degrés de liberté et donnent la possibilité d'interagir en 3 dimensions. Cependant, ces solutions sont lourdes à mettre en place. S'équiper du gant est long et le dispositif embarqué augmente la charge cognitive de la tâche. De plus, la mécanique de la main est modifiée par l'ajout d'un système embarqué.

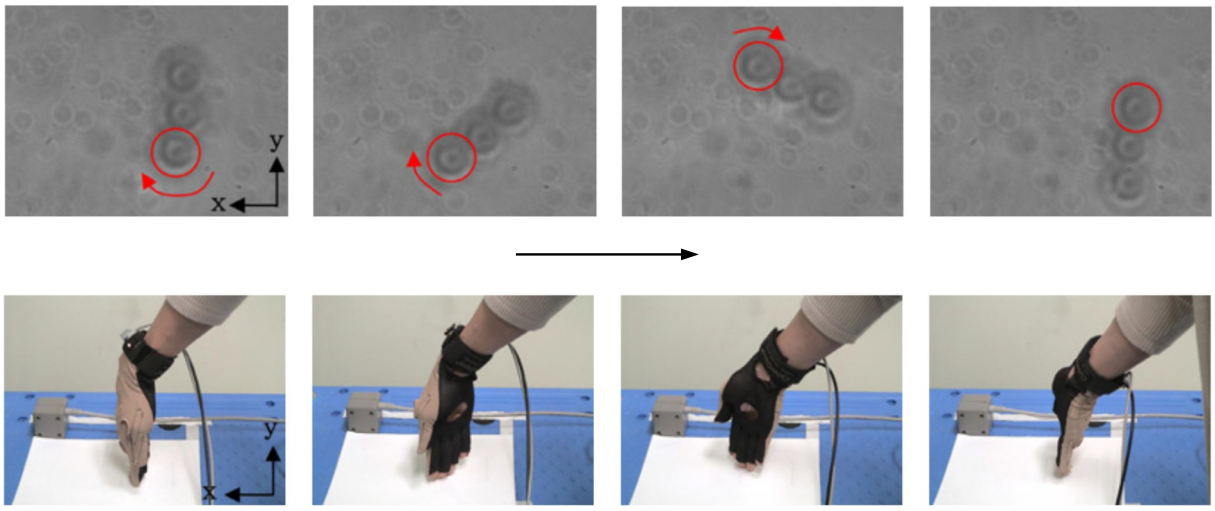


Figure 1.7 - *Manipulation de pinces optiques avec un gant numérique (Cyberglove) [Park 07]*

3.2.4 Les interfaces par reconnaissance de gestes

Dans une interface par reconnaissance de gestes, l'utilisateur est filmé par un capteur de vision, par exemple une caméra RGB. Les images reçues sont ensuite traitées par des méthodes de vision par ordinateur. Ces méthodes réalisent le suivi de la main et la reconnaissance de geste nécessaires à l'interaction avec la réalité virtuelle en temps réel.

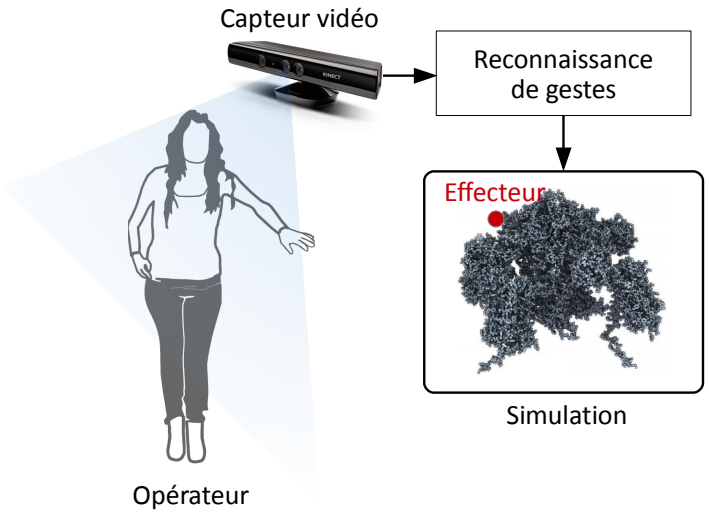


Figure 1.8 - *Interface de vision par ordinateur*

Il existe peu d'interface de vision par ordinateur pour la micromanipulation dans l'état de l'art. Deux articles de la fin des années 90 proposent des interfaces pour la simulation moléculaires basées sur des méthodes de suivi de la main et reconnaissance de geste avec une ou plusieurs caméras RGB [Zeller 97, Pavlovic 96]. Les contraintes techniques de l'époque rendent le temps réel difficile à atteindre, et limitent les performances. Ces travaux sont restés sans suite.

Plus récemment, [Whyte 06] propose un système de manipulation de pinces optiques basé sur de la détection par vision par ordinateur de billes blanches fixées au pouce et à l'index de l'utilisateur. La taille apparente des billes est utilisée pour inférer la position en profondeur des doigts de l'utilisateur. Cette méthode rend la transcription de la profondeur plus naturelle que l'approche basée sur un curseur exploitée avec les interfaces tactiles. Cependant, elle contraint la pose de la main de l'utilisateur afin d'éviter les occlusions des billes blanches. De plus, les actions discrètes (saisie et dépose) ne sont pas gérées.

Enfin, quelques travaux évoquent la perspective d'utiliser les nouvelles technologies d'interaction issues entre autres du jeu vidéo (Microsoft Kinect, Leap Motion) pour parvenir à une interaction plus naturelle et intuitive dans le cadre de la micromanipulation [Lv 13, LaViola 11]. Cette dernière solution a récemment été exploitée pour la manipulation de pièges optiques [McDonald 13]. Cette méthode exploite des gestes symboliques discrets pour déclencher des actions, par exemple faire un signe de la main pour créer un nouveau piège optique puis pour le supprimer. Le déplacement est géré de manière non symbolique : il suit les déplacements de la main de l'utilisateur. L'utilisation de la Kinect permet de gérer la profondeur de manière naturelle.

3.3 Synthèse

Le tableau 1.1 présente les signaux d'acquisition et de restitution de différents systèmes d'interfaces caractéristiques de micromanipulation. Les signaux d'acquisition qui correspondent à un déplacement pour positionner un curseur sont relevés. Ceux qui correspondent à la gestion des évènements, par exemple la préhension, le lâcher ou la suppression d'un objet sont répertoriés. Enfin, l'ensemble de ces signaux sont classés selon leur caractère symbolique/non symbolique.

L'aspect restitution de ces systèmes est exploré en profondeur pour correspondre aux modalités sensorielles humaines. La réalité virtuelle cherche à se rapprocher visuellement du monde réel [Bolopion 10b]. Le retour haptique est exploité dans de nombreux travaux pour ressentir des forces à l'échelle microscopique, peu intuitives à l'échelle humaine. De ce point de vue, l'aspect restitution est traité de manière naturelle.

Par opposition, au niveau de l'acquisition, l'ensemble des ces systèmes exploitent un **langage prédéfini symbolique** pour la reconnaissance d'évènements. Les interfaces haptiques utilisent des clics sur des boutons, par exemple [Bolopion 11] pour réaliser la saisie d'un atome dans une interface de réalité virtuelle en trois dimensions. Les solutions basées sur des écrans tactiles rendent l'interaction plus naturelle, mais sont limitées à une interaction en 2D. La gestion de la 3D passe par des méthodes symboliques.

En vision par ordinateur, [Pavlovic 96] et [McDonald 13] proposent des systèmes basés sur un langage gestuel discret pour interagir avec la scène virtuelle. Ce langage est arbitraire, par exemple un signe de la main est utilisé pour créer un piège optique. Il s'agit donc d'interfaces symboliques, qui ne sont pas naturelles au sens de la définition donnée. Pour utiliser l'interface, une phase d'apprentissage de ce langage gestuel est donc nécessaire. Enfin, il est important de noter qu'aucun de ces travaux n'évalue l'aspect naturel de l'interaction du point de vue de l'utilisateur.

Le geste est une modalité intuitive pour l'interaction en 3D. Il est au centre de la communication entre êtres humains. Les **méthodes d'interaction gestuelles** sont donc une solution prometteuse pour créer une interface plus naturelle destinée à la micromanipulation. La vision par ordinateur est une solution adaptée pour détecter ces gestes. En particulier, les nouveaux capteurs de vision tels que la Kinect sont optimisés pour une interaction en 3D. Cette solution est donc retenue dans ce travail. Cependant, aucune approche actuelle n'exploite ces outils de manière naturelle pour l'utilisateur. La section suivante s'intéresse aux solutions dans le domaine de l'interface homme machine (IHM) à l'échelle macroscopique pour dépasser ces limites.

	Acquisition		**Restitution**
	Déplacements	Événements	
Bras manipulateur			
[**Millet 08**]	Non symbolique : force, position	-	Visuel : RV 3D Haptique
[**Bolopion 11**]	Non symbolique : force, position	Symbolique : clics	Visuel : RV 3D Haptique
[**Sauvet 12**]	Non symbolique : force, position	Symbolique : clics	Visuel : RV 3D Haptique
Tactile			
[**Grieve 09**]	Non symbolique : position 2D des doigts	Symbolique : gestes, clics	Visuel : scène réelle 2D
[**Bowman 11**]	Non symbolique : position 2D des doigts	-	Visuel : scène réelle 2D
Gants numériques			
[**Park 07**]	Non symbolique : position 2D des doigts	-	Visuel : scène réelle 2D
Vision par ordinateur			
[**Pavlovic 96**]	Non symbolique : position 3D de la main	Symbolique : gestes, voix	Visuel : RV 3D
[**Whyte 06**]	Non symbolique : positions 3D pouce et index avec marqueurs	Non symbolique : comportement de l'opérateur	Visuel : RA
[**McDonald 13**]	Non symbolique : position 3D de la main	Symbolique : gestes dynamiques	Visuel : RA et scène réelle

Tableau 1.1 - *Signaux d'acquisition et de restitution d'interfaces de micromanipulation. La restitution visuelle est notée RA pour réalité augmentée et RV pour réalité virtuelle.*

4 Approche proposée

4.1 Les interfaces gestuelles à l'échelle macroscopique

Une solution pour créer une interface de micromanipulation plus naturelle consiste à s'intéresser aux approches issues du domaine connexe de l'IHM.

Ce travail se concentre sur les interfaces qui peuvent inclure les tâches typiques de la micromanipulation : la saisie, le déplacement et la dépose d'objets. D'autres travaux ne traitent pas explicitement ces tâches mais proposent des méthodes adaptées à leur réalisation. Ainsi, des approches qui prennent en compte des décisions discrètes de l'opérateur et/ou des déplacements d'objets sont incluses dans cet état de l'art.

Pour réaliser ces tâches, la plupart des travaux de la littérature relèvent de deux grands types d'approches. La première est **l'interaction par langage gestuel**. Il s'agit de détecter des symboles gestuels et de les relier aux différentes commandes proposées par l'interface. Un dictionnaire de symboles gestuels remplace ainsi les clics des interfaces de type clavier-souris pour déclencher les actions.
La deuxième approche s'appuie sur une **manipulation directe** sans interprétation sémantique. Celle-ci consiste à inclure un effecteur, par exemple un curseur qui correspond à la main, dans une scène de réalité virtuelle. Un moteur physique calcule les collisions et les forces que cet effecteur applique sur les objets virtuels pour les déplacer.
Cette partie s'attache à étudier des systèmes représentatifs de chacune de ces approches. En particulier, les solutions apportées pour inclure les décisions discrètes de l'opérateur (saisie, sélection, dépose) et les déplacements sont analysées. Ce travail trace les limites de ces deux systèmes pour réaliser ces tâches et propose une approche alternative basée sur l'étude du **comportement naturel humain**.

4.1.1 L'interaction par langage gestuel

L'approche la plus répandue dans le domaine de l'IHM à partir des gestes consiste à exploiter un langage gestuel. Un dictionnaire de gestes est fourni à l'opérateur. Deux exemples de dictionnaires proposés dans la littérature sont montrés sur la figure 1.9. Chaque geste correspond à une commande spécifique.

Ces gestes sont détectés par des méthodes de vision par ordinateur à partir de capteurs RGB [Manresa 05] et plus récemment de capteurs 3D [Van den Bergh 11]. Le dictionnaire de gestes peut être composé de gestes statiques (poses de la main) [Manresa 05] ou dynamiques (mouvements) [Palacios 13] [Gillian 14b]. Ces gestes sont choisis de manière arbitraire par le concepteur de l'interface, il s'agit donc de symboles gestuels. Ceux-ci sont particulièrement adaptés à la reconnaissance de décisions discrètes comme la sélection, l'accès à une page précédente ou suivante.

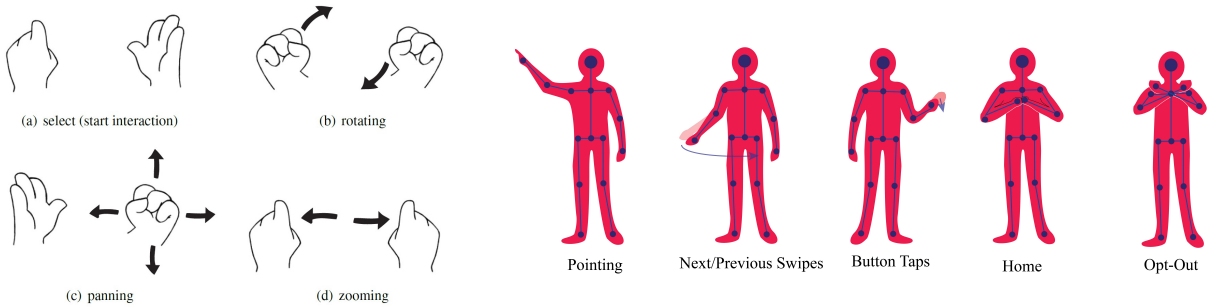


Figure 1.9 - *Exemples de dictionnaires de gestes symboliques pour l'IHM [Gillian 14b] [Van den Bergh 11]*

La plupart des travaux existants cherchent à remplacer l'interface clavier-souris pour naviguer dans des menus, parcourir des fichiers, etc [Gillian 14b]. Peu de travaux se concentrent sur des tâches de manipulation en réalité virtuelle. Van den Bergh et al. présentent une interface de navigation dans un modèle 3D de ville [Van den Bergh 11]. Cependant, les auteurs ne proposent pas de réelle manipulation d'objets et se concentrent sur la visualisation. Ainsi, le déplacement d'objets virtuels n'est généralement pas traité par ce type d'interfaces.

	Application	**Décisions**	**Déplacements**
[Manresa 05]	Jeu vidéo (aucune application montrée) Webcam	Gestes symboliques statiques (start, stop, move, no hand)	4 gestes symboliques statiques (gauche, droite, avant, arrière)
[Palacios 13]	IHM (aucune application montrée) Kinect	Gestes symboliques statiques	Gestes symboliques dynamiques (haut, bas, droite, gauche, avant, arrière)
[Gillian 14a] [Gillian 14b]	Bibliothèque pour l'IHM Multi capteurs (RGB, Kinect)	Gestes symboliques dynamiques (pointage, menu, sortie, clic, suivant/précédent)	-
[Van den Bergh 11]	Navigation visuelle dans un modèle 3D de ville	Gestes symboliques (position, rotation, zoom, start)	-

Tableau 1.2 - *Exemple d'interfaces typiques basées sur des langages gestuels.*

Plusieurs critiques sont formulées sur l'approche par langage gestuel. Norman note qu'il est nécessaire d'apprendre et de retenir un langage parfois complexe pour interagir [Norman 10]. De plus, les gestes employés sont peu naturels et intuitifs [Malizia 12].

Partant de ce constat, une solution consiste à déterminer les gestes les plus appropriés pour chaque commande à réaliser. À partir de l'étude des préférences de nombreux utilisateurs, Aigner et al. proposent de définir un **dictionnaire de gestes naturels** adapté à différentes tâches [Aigner 12]. Ainsi, le geste relevé pour la saisie et la sélection est une fermeture de la main. Pour le déplacement d'un objet, celui-ci est une translation jusqu'au point désiré. La dépose est associée à une ouverture de la main. Il s'agit donc de gestes qui imitent le geste réel utilisé pour réaliser la tâche. Le deuxième chapitre de ce travail explore cette solution appliquée à la micromanipulation. Le but est de déterminer si un dictionnaire de gestes naturels est suffisant pour créer une interface non symbolique.

Une autre critique des langages gestuels est formulée par Ardito et al. [Ardito 14]. Les auteurs distinguent les symboles gestuels, destinés à communiquer un sens, des gestes de manipulation non-sémantiques. Ils défendent l'idée que seuls les gestes de manipulation directe sont adaptés à la création d'interfaces naturelles. En effet, ces derniers ne dépendent pas de l'utilisateur, du contexte et des cultures. Les interfaces qui relèvent de cette approche sont étudiées dans la section suivante.

4.1.2 L'interaction par manipulation directe

La manipulation directe consiste à exploiter un moteur de réalité virtuelle qui reproduit de manière réaliste la physique du monde réel. Ce type d'approche n'implique pas d'interprétation sémantique des actions de l'utilisateur. Elle repose sur une interaction physique entre un effecteur qui suit les déplacements de la main de l'utilisateur et la scène 3D.

Un exemple représentatif de cette approche est l'Holodesk [Hilliges 12]. La surface de la main de l'utilisateur est détectée en 3D par un capteur Kinect. Elle est incluse dans le simulateur physique pour interagir directement avec des objets virtuels. Des calculs de collision sont réalisés par le moteur physique pour rendre possible la saisie d'objets.

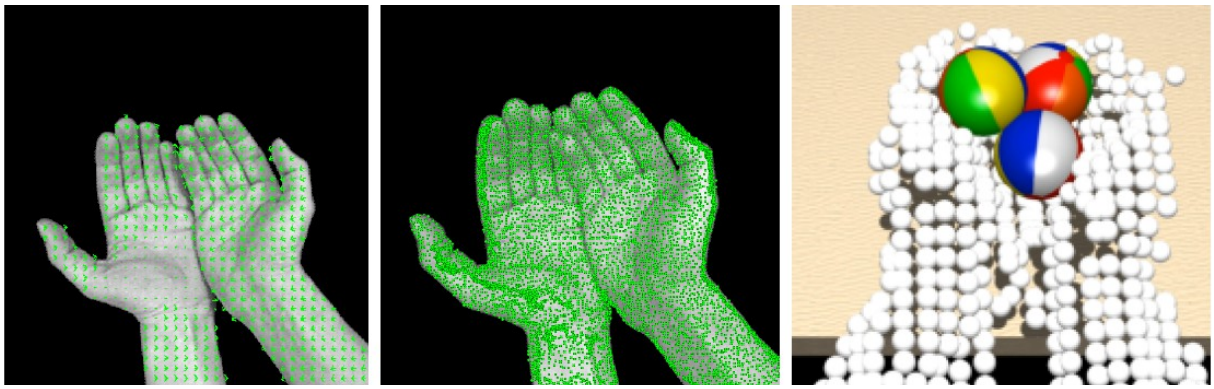


Figure 1.10 - *La surface de la main de l'utilisateur détectée avec la Kinect est représentée par des particules physiques qui interagissent avec des sphères virtuelles dans le simulateur. [Hilliges 12] [Van den Bergh 11]*

D'autres solutions exploitent un curseur qui suit les mouvements de la main. Une sélection automatique est réalisée lorsque le curseur est placé sur un objet [Poupyrev 98].

Contrairement à l'approche par langage gestuel, la manipulation directe facilite les tâches continues comme le déplacement d'un objet virtuel. Cependant, la saisie d'objets avec des dispositifs du type HoloDesk reste difficile car cette méthode est dépendante de calculs complexes de collision et elle limite les interactions possibles à la surface de la main visible par le capteur. Ainsi, elle n'apporte pas de solution pour détecter les décisions discrètes de l'utilisateur telles que la sélection ou la saisie. Pour valider la sélection certains travaux se reposent sur un symbole gestuel [Vogel 05] ou sélectionnent l'objet dès qu'il est en contact avec le curseur [Poupyrev 98].

En conclusion, les deux approches principalement représentées dans la littérature présentent des limites qui doivent être dépassées pour créer une interface naturelle non-symbolique :

- **Les langages gestuels** sont adaptés à la détection des décisions discrètes comme la saisie et la dépose d'objets. Cependant, ils reposent sur des gestes symboliques peu naturels que l'utilisateur doit apprendre.
- **La manipulation directe** est adaptée aux tâches continues de déplacement mais ne permet pas d'affecter de sens aux actions de l'utilisateur. Elle n'est donc pas adaptée à la détection des décisions.

Une piste de solutions peu explorée dans la littérature consiste à inférer les décisions de l'utilisateur d'après son comportement naturel. Ces approches sont principalement exploitées dans le domaine de la robotique humanoïde, où l'interaction physique et sociale avec l'être humain est un problème central [Melnyk 14]. Dans le cadre des interfaces homme-machine, ces approches ont pour objectif d'analyser le comportement de l'opérateur pour déterminer ses décisions de manière implicite. Il ne lui est donc plus nécessaire d'expliciter celles-ci par des symboles gestuels. Cette approche d'interaction basée sur le comportement est explorée dans la section suivante.

4.1.3 L'interaction basée comportement

Parmi les approches basée comportement, [Carrasco 10] propose d'exploiter la coordination du geste et du regard pour prédire l'intention de préhension et l'objet visé par l'utilisateur. [Stefanov 10] utilise la dynamique du geste avec un bras haptique pour inférer l'intention de l'opérateur pour des tâches de guidage. [Horvitz 03] utilise un modèle attentionnel probabiliste pour évaluer si l'ordinateur est la cible d'une commande vocale. D'autres travaux proposent d'exploiter la cinématique du geste pour inférer la tâche que l'utilisateur cherche à accomplir [Choumane 10]. Ces différentes solutions sont synthétisées dans le tableau 1.3 en mettant l'accent sur les signaux extraits et les informations haut niveau modélisées.

	Application	Signaux bas niveau	Haut niveau
[**Horvitz 03**]	Évaluation de la cible d'une commande vocale	Reconnaissance vocale Pose du visage	Engagement Statut attentionnel
[**Carrasco 10**]	Interaction avec un objet réel	Eyetracker : regard Caméra fixée au poignet	Intention : saisie Objet cible
[**Choumane 10**]	Manipulation de sphères en RV	Gant numérique : vitesse et accélération de la main	Intention : sélection, saisie, lâcher
[**Stefanov 10**]	Téléopération assistée par ordinateur	Bras haptique : force, vitesse et position	Intention : positionnement, transport

Tableau 1.3 - *Signaux de haut et bas niveau d'interfaces naturelles en IHM*

Les travaux recensés montrent que l'**analyse du comportement de l'utilisateur** est une piste pour une interaction plus naturelle. Un autre point notable est le fait que les informations de haut niveau telles que les décisions reposent sur l'extraction de **données de bas niveau** comme la pose du visage, le regard ou la cinématique du geste. Cette approche est une solution prometteuse pour reconnaître les décisions discrètes de l'utilisateur de manière non symbolique. En effet, elle repose sur le comportement naturel de l'utilisateur sans lui imposer l'apprentissage d'un langage gestuel. Celle-ci est donc adoptée dans ce travail.

4.2 Système proposé

4.2.1 Analyser le comportement naturel de l'opérateur

Dans le cadre des interfaces pour la micromanipulation, les tâches classiques impliquent de fournir à l'opérateur la capacité de saisir, déplacer et relâcher un objet donné. Pour améliorer le retour visuel, des simulateurs de systèmes de micromanipulation sont proposés. Ils reproduisent une plateforme réelle de micromanipulation en réalité virtuelle. Il est ensuite possible de les coupler à la tâche réelle. Les simulateurs sont une solution adaptée pour évaluer des solutions d'interfaces. Ils permettent de s'abstraire des contrainte d'une plateforme réelle afin de se concentrer sur l'interaction avec l'utilisateur.

Dans ce contexte, l'interface doit répondre à deux questions : "Quel objet ?", "Quelle tâche ?".
Déterminer l'objet visé par l'opérateur consiste à le sélectionner dans une scène virtuelle. Il s'agit par exemple de déterminer quel atome d'une molécule complexe l'utilisateur veut manipuler. Chez l'être humain, les mécanismes de sélection d'éléments dans une scène

visuelle sont gérés par l'attention visuelle [Kastner 04]. Ces processus ont pour but de réduire le coût cognitif d'une tâche en sélectionnant les informations visuelles pertinentes pour sa réalisation. Il existe des indices comportementaux du focus d'attention. Ainsi, l'orientation du regard est un indice important. Il est exploré dans de nombreux travaux de la littérature [Horvitz 03,Schiavo 13,Borji 13]. Ce travail propose de modéliser le **focus d'attention** de l'utilisateur à partir d'indices comportementaux. Il s'agit de déterminer la ou les cibles probables de la tâche dans des environnements complexes sans que l'utilisateur n'ait besoin de les expliciter.

Lorsque la cible est sélectionnée, la seconde étape consiste à lui associer la tâche à réaliser. L'approche par langage gestuel est centrée sur la forme du geste, qui est arbitraire. Pour dépasser les limites relevées dans la section 4.1.1, ce travail propose au contraire de se concentrer sur le but de l'utilisateur. Il s'agit donc de déterminer son **intention** d'après son comportement naturel.

4.2.2 Les modèles haut niveau du focus d'attention et de l'intention

Cette approche doit faire face à plusieurs difficultés :

- Les symboles gestuels sont faciles à reconnaitre, car ils correspondent à un dictionnaire préétabli et ils sont peu variables. L'approche basée comportement implique d'affecter du sens à des mouvements non contraints. Elle doit donc gérer des mouvements très variables. Dans ce but, des invariants du gestes caractéristiques doivent être déterminés.
- Ce travail se concentre sur des tâches génériques de manipulation. Ainsi, il n'est pas possible de se baser uniquement sur un contexte très spécifique pour déterminer la tâche.
- L'intention et le focus d'attention sont des notions de haut niveau. Celles-ci ne peuvent pas être extraites directement par les dispositifs d'acquisition. Elles reposent sur l'interprétation d'informations de bas niveau.

Dans ce but, ce travail propose d'intégrer une couche haut niveau modélisant le focus d'attention et l'intention de l'utilisateur à l'interface de micromanipulation (voir Fig. 1.11). Cette couche de haut niveau est construite à partir de signaux de bas niveau caractéristiques de l'intention et du focus d'attention.

4.2.3 Synthèse

La problématique de ce travail se concentre sur la **synthèse d'une interface naturelle et intuitive pour interagir avec le micromonde** par l'intermédiaire de la réalité virtuelle. Ici, une interface est considérée comme naturelle si elle ne nécessite pas

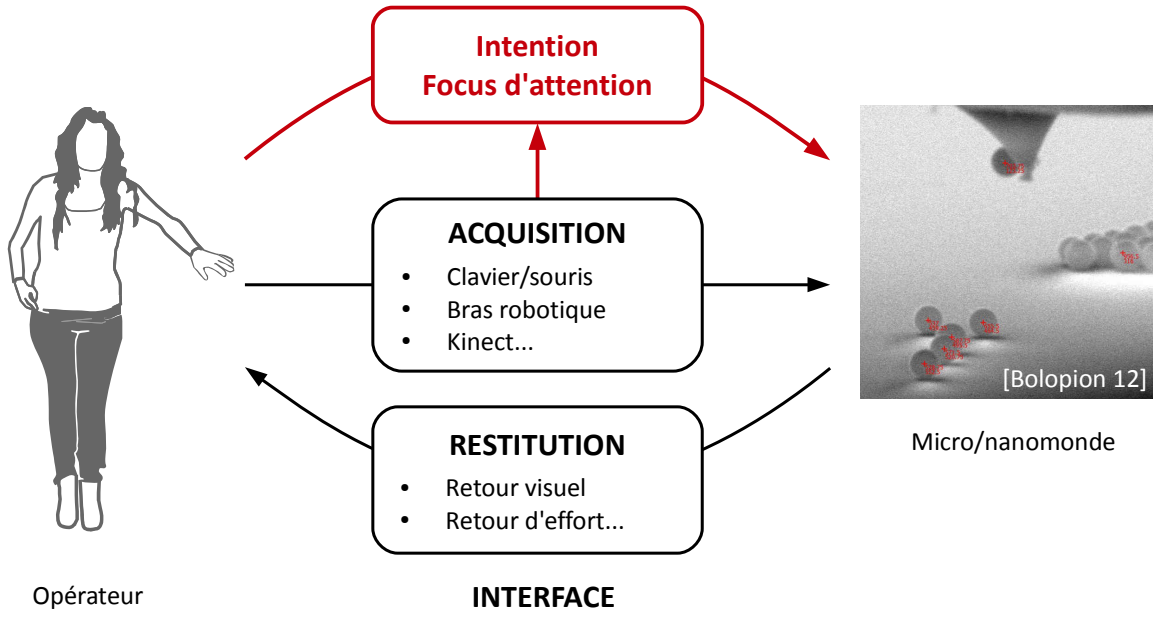


Figure 1.11 - *Schéma d'interface naturelle pour la micromanipulation.*

de fournir d'instruction à l'opérateur pour être fonctionnelle. Il s'agit donc d'interpréter le comportement naturel humain lors de tâches de manipulation.

Pour créer cette interface, ce travail propose de déterminer la cible et la tâche visées par l'utilisateur en s'inspirant de **modèles cognitifs computationnels de son focus d'attention et de son intention**. Ces modèles reposent sur des **signaux de bas niveau** extractibles par les capteurs. Dans ce but, des méthodes d'acquisition basées sur la vision par ordinateur sont mises en place.

Le deuxième chapitre de ce travail s'attache à décrire un simulateur intuitif de manipulation en réalité virtuelle. Pour interagir avec ce simulateur, une première méthode de détection des décisions de l'utilisateur est proposée. Elle repose sur un langage gestuel.
Pour réaliser une interface non symbolique, le troisième chapitre propose une approche basée comportement pour reconnaître les décisions. Cette méthode repose sur un modèle cognitif de l'intention.
Enfin, dans le quatrième chapitre, une méthode est proposée pour déterminer le focus d'attention de l'opérateur. Cette méthode a pour but d'élargir le champ d'application du système à des contextes complexes où l'objet cible ne peut pas être déterminé a priori.

Chapitre 2

Un système bas niveau d'évaluation de l'interface

Sommaire

1 Interagir avec le micromonde 6
1.1 La micromanipulation 6
1.2 Interagir avec une scène de micromanipulation 7
1.3 Interfaces de restitution en micromanipulation 7
1.4 Interfaces d'acquisition en micromanipulation 8
1.5 Vers une interface naturelle pour la micromanipulation 9
2 Les interfaces naturelles 10
2.1 Propriétés 10
2.2 Les symboles dans les interfaces 10
2.3 Sens et actions non symboliques 11
3 État de l'art des interfaces pour la micromanipulation 11
3.1 Interfaces de restitution 12
3.2 Les interfaces d'acquisition 14
3.3 Synthèse 17
4 Approche proposée 19
4.1 Les interfaces gestuelles à l'échelle macroscopique 19
4.2 Système proposé 23

Un enjeu des interfaces consiste à inclure les capacités décisionnelles de l'utilisateur dans la tâche de manipulation robotique. Afin de fournir une interface destinée à des non-spécialistes, la complexité de l'interaction réelle est masquée. L'entrée de ce système

correspond à la tâche que l'utilisateur veut réaliser. Ce chapitre décrit une application de ce système pour l'assistance à la téléopération. Ce système est appliqué à la manipulation de microbilles sous microscope à force atomique (AFM). Dans ce but, une interface de restitution visuelle est proposée. Cette interface inclut une main virtuelle afin de faciliter l'identification de l'effecteur avec la main réelle de l'utilisateur. Le système robotique réel ou simulé est téléopéré par l'intermédiaire de cette interface. L'enjeu est d'inclure les décisions de l'utilisateur dans ce système en automatisant les sous-tâches qui n'impliquent pas de décision. La difficulté principale de ce travail consiste à détecter la décision de l'utilisateur sans contraindre l'interaction. Une solution par reconnaissance de gestes "naturels" est proposée, qui servira de base de comparaison dans cette thèse.

1 La micromanipulation téléopérée par contact adhésif

A l'échelle micrométrique, les tâches typiques de manipulation consistent en la saisie, le déplacement et la dépose de micro objets sur le substrat. Un exemple de stratégie démontré expérimentalement est la micromanipulation de microsphères par adhésion avec une poutre d'un microscope à force atomique (AFM) [Haliyo 04].

Ce dernier est constitué d'une pointe de dimension micrométrique fixée à l'extrémité d'une poutre flexible. La poutre est déplacée pour effectuer un balayage de la surface de l'échantillon. L'interaction de la pointe avec la surface de l'échantillon provoque une déflexion du levier. Pour mesurer cette déflexion, la méthode la plus employée exploite un faisceau laser qui se réfléchit sur la poutre. La déviation de celui-ci est mesurée par une photodiode [Binnig 86] (fig.2.1). Il donne ainsi accès à une grande variété de propriétés des surfaces (mécaniques, électrique, magnétiques...). Lorsque le cantilever est utilisé sans pointe, il peut être exploité comme outil pour la manipulation de micro-objets. Lorsque l'AFM est utilisé pour manipuler un objet, la visualisation est réalisée par microscopie optique ou électronique.

1.1 Le principe de la micromanipulation par adhésion

Pour les objets de dimension inférieure à 100μm, les forces surfaciques (les forces de van der Waals, capillaires et électrostatiques) sont plus importantes que les forces volumiques. Il est possible de tirer parti de ces forces pour manipuler des micro objets. La poutre d'un AFM est utilisé pour la manipulation comme montré sur la figure 2.2. Les objets sont saisis par simple contact grâce à la supériorité de l'adhésion par rapport au poids. Ensuite, les objets sont déposés sur un substrat d'adhésion supérieure à celle de la poutre.

Pour les tâches non-automatisables, le savoir-faire d'un opérateur est nécessaire à la manipulation. Pour donner à l'utilisateur la possibilité de toucher à ces échelles, des bras

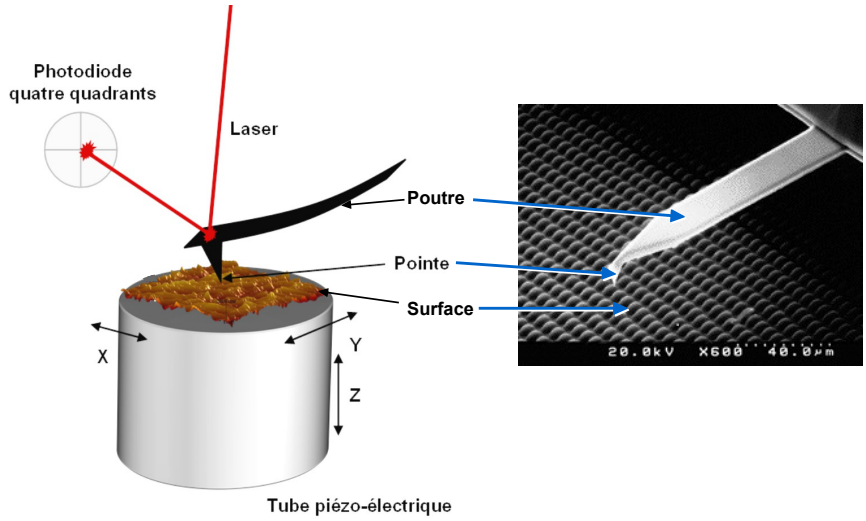


Figure 2.1 - *Principe de la caractérisation d'un échantillon par un microscope à force atomique et système réel*

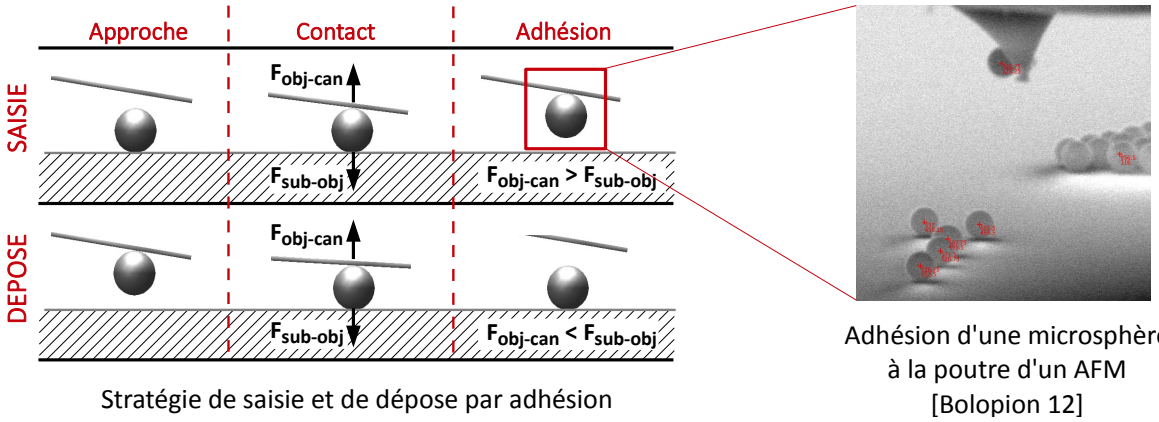


Figure 2.2 - *Saisie (en haut) et dépose (en bas) d'une microsphère avec la poutre d'un AFM par adhésion. F_{i-j} est la force d'adhésion entre i et j.*

à retour d'efforts sont proposés [Venture 06] [Boukhnifer 06] [Bolopion 13b]. Dans ce type de systèmes, le retour visuel est limité. Pour dépasser cette limite, certains travaux incluent un retour visuel en réalité virtuelle en 3D [Ferreira 04] [Sauvet 12]. Les simulateurs en réalité virtuelle sont des solutions intéressantes pour fournir à l'utilisateur une vue reconstruite en trois dimensions de la scène de manipulation. Elle peut ainsi être affichée à l'échelle de l'opérateur sous différents points de vue. La figure 2.3 illustre un système de téléopération basé sur un retour visuel en réalité virtuelle pour la micromanipulation par adhésion. Ces nouvelles interfaces donnent à l'opérateur la possibilité de manipuler des micro-objets intuitivement pour des applications variées [Millet 13].

Les outils de la réalité virtuelle permettent aussi d'inclure des simulateurs physiques pour compenser le faible nombre de capteurs [Ammi 06] [Bolopion 11]. Ces derniers reproduisent la plateforme réelle de manipulation. Ce type de simulateurs est une solution

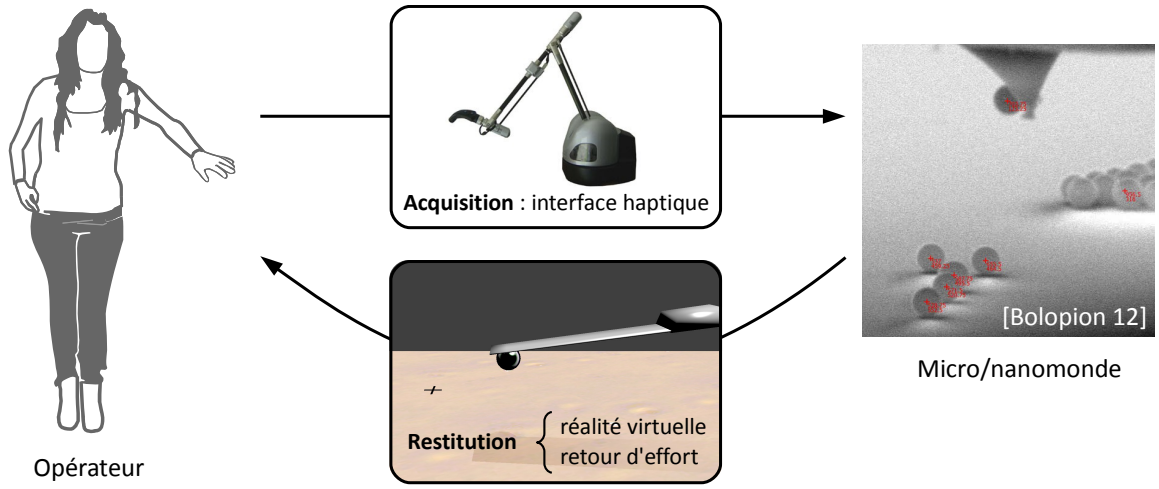


Figure 2.3 - *Téléopération d'un système de micromanipulation*

adaptée pour valider les interfaces proposées dans ce travail. De plus, le simulateur peut être couplé avec une plateforme réelle de micromanipulation. L'opérateur n'interagit directement qu'avec le simulateur, la complexité du système réel est ainsi masquée.

Dans des travaux antérieurs [Millet 08], les auteurs exploitent un bras haptique pour téléopérer une poutre d'AFM dans un simulateur en réalité virtuelle. Ce simulateur reproduit de manière réaliste les forces d'adhésion et la dynamique d'un système réel, et permet de valider la téléopération haptique. Il est composé d'une poutre virtuelle flexible qui peut être déplacé selon la cinématique du système. Les forces d'adhésion entre l'objet, le substrat et la poutre montrées sur la figure 2.2 sont modélisées de manière réaliste. L'interaction est mesurée à partir de la déflection de la poutre comme avec un système AFM réel. Le moteur physique rend possible l'interaction en temps réel. L'opérateur ressent les micro interactions telles que les phénomènes d'adhésion et de pull-off par un retour haptique.

Ce travail de thèse exploite ce simulateur pour explorer de nouveaux moyens d'interagir avec le micro monde pour fournir une assistance à la téléopération virtuelle.

1.2 Un simulateur physique pour l'évaluation

Le simulateur actuel est réalisé avec le moteur de jeu du logiciel Blender [1]. Il affiche une représentation 3D de la poutre, du substrat et des objets à manipuler. L'utilisateur commande la poutre en position avec une interface haptique et reçoit un signal proportionnel à la flexion de la poutre. Ce retour d'effort est l'unique information dont il dispose pour générer des trajectoires. Il réalise ainsi la saisie, le déplacement et la dépose en prenant

1. http ://www.blender.org

en compte la physique particulière de l'environnement.

Pour créer un lien plus intuitif entre la main de l'utilisateur et l'effecteur virtuel, ce travail propose d'inclure une interface naturelle dans le simulateur. Dans cette interface, l'opérateur interagit par le biais d'une main virtuelle, incluse dans le simulateur. Le système de téléopération proposé est montré sur la figure 2.4.

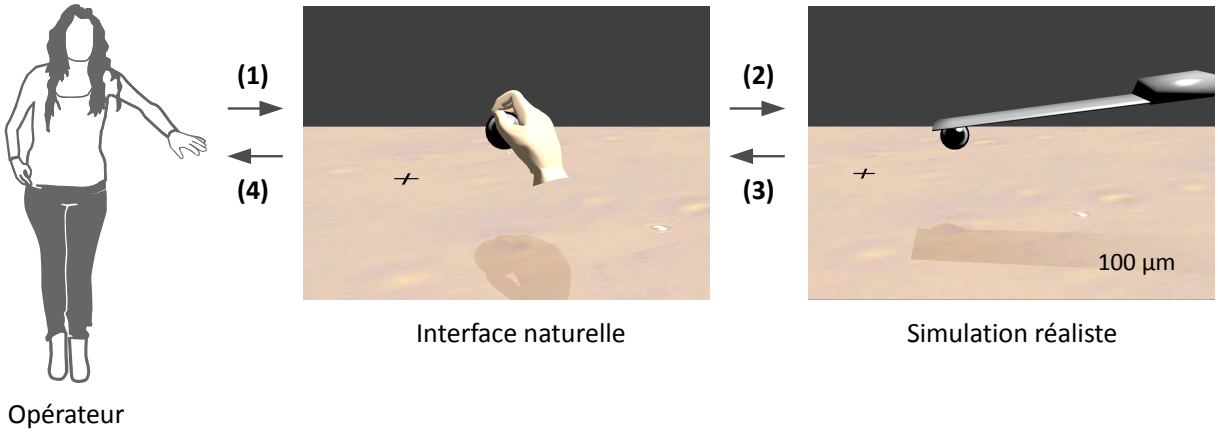


Figure 2.4 - *Téléopération d'un simulateur de micromanipulation en réalité virtuelle. (1) Les translations et rotations de la main de l'utilisateur sont appliquées à la main virtuelle ainsi que ses décisions (saisie, dépose) (2) Un système maître-esclave est utilisé pour téléopérer la poutre de l'AFM dans le simulateur réaliste. (3) La réussite ou l'échec de la tâche de saisie/dépose est retourné à l'interface naturelle. (4) Un retour visuel est donné à l'utilisateur.*

L'interface naturelle est réalisée avec le moteur de jeu du logiciel Blender. En particulier, la main virtuelle créée pour cette interface est présentée dans la suite de cette section.

1.2.1 Le logiciel Blender

Blender est un logiciel destiné à la création d'environnements virtuels dynamiques. Il dispose d'un outil de modélisation. À partir d'une armature en 3D, il est possible de créer un maillage déformable pour réaliser des animation. Dans ce but, il inclut un moteur de jeu adapté à des tâches interactives telles que les tâches de manipulation d'objets virtuels. Il dispose d'un affichage OpenGL en trois dimensions. Sa couche moteur physique générale (Bullet) est capable de réaliser la résolution des équations de Newton en temps réel ainsi que de détecter les collisions. Ce moteur de jeu est extensible par des scripts en Python. Blender est donc une bonne solution pour la simulation d'environnements réalistes interactifs.

1.2.2 La main virtuelle

Pour résoudre le problème de l'identification entre la main de l'utilisateur et l'effecteur virtuel, une main virtuelle est incluse dans le simulateur. Cette main est modélisée avec le logiciel Blender en respectant la morphologie de la main humaine. Des animations qui correspondent aux tâches de saisie et dépose sont créées pour fournir un retour visuel à l'opérateur lors de la manipulation.

1.2.2.1 Modélisation de la main

La modélisation de la main virtuelle est réalisée d'après la morphologie d'une main réelle. La main humaine est composée de 27 os. 19 d'entre eux constituent la paume et les doigts. Les 8 autres sont situés dans le poignet, comme montré sur la figure 2.5 ;

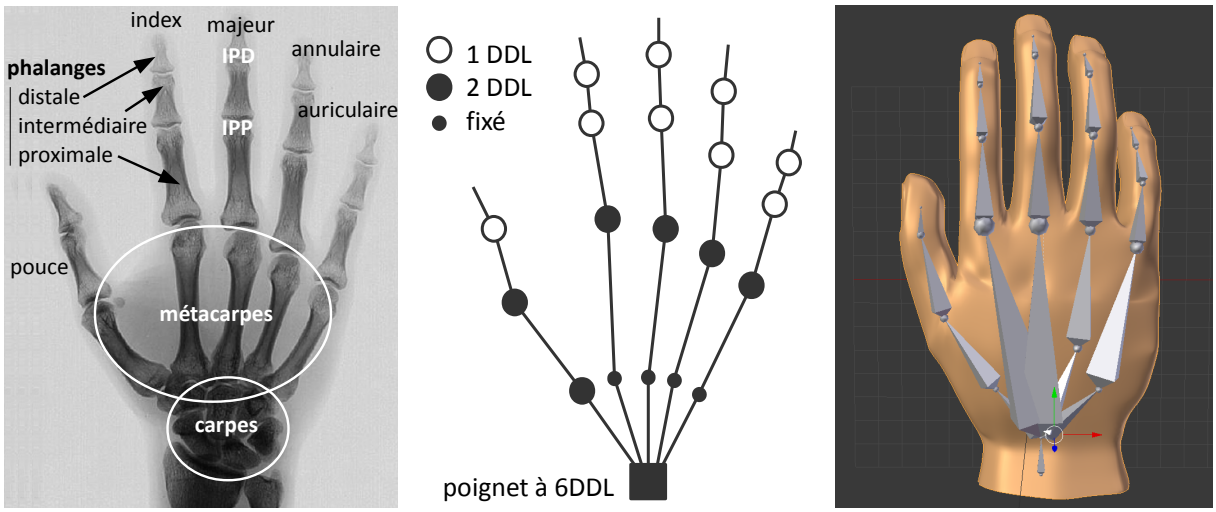


Figure 2.5 - *Anatomie de la main humaine et degrés de liberté des articulations. Les articulations interphalangienne sont notées IPP (inter-phalangienne proximale) et IPD (inter-phalangienne distale)*

Il s'agit d'un système redondant, la dimension des degrés de libertés est supérieure à la dimension du vecteur des positions atteignables. Le modèle inverse admet donc une infinité de solutions. Cependant, certaines configurations ne sont pas réaliste du fait des limites articulaires anatomiques. De plus, certains degrés de liberté sont couplés.

Les 19 os qui constituent la paume et les doigts sont modélisés en respectant les dimensions moyennes de la main humaine [Buryanov 10]. Les os du carpe sont modélisés de façon simplifiée par 3 os afin de ne conserver que les articulations les plus mobiles. La main virtuelle modélisée comporte donc une armature de 22 os. Elle est recouverte d'un maillage déformable. Les contraintes biomécaniques sont intégrées en fixant des seuils maximaux sur les angles articulaires.

La flexion de l'articulation PIP est couplée linéairement selon un rapport d'environ 0.7/° avec la flexion de l'articulation DIP [Nimbarte 08]. La flexion de l'articulation IPD est calculée selon ce rapport d'après la flexion de l'articulation IPP. La figure 2.5 représente le squelette virtuel ainsi que le maillage déformable.

La main virtuelle est commandée par une communication réseau en UDP. Les angles correspondant à chaque articulation sont commandés séparément par un port UDP distinct. Deux ports UDP sont dédiés à la translation et à la rotation de la paume de la main. Un script en Python est associé à chacun de ces ports afin de déplacer la main selon les données reçues.

1.2.2.2 Animations de la main

Dans le cadre de l'interface proposée, l'utilisateur doit pouvoir saisir et déposer des objets. Il est important de lui fournir un retour visuel correspondant à chacune de ces tâches afin de fermer la boucle du système comme montré sur la figure 2.3. Dans ce but, deux animations de la main virtuelle sont réalisées. Ces animations sont des transitions dynamiques entre deux poses statiques de la main :

- La transition d'une pose main ouverte à une pose main fermée pour la saisie
- La transition d'une pose main fermée à une pose main ouverte pour la dépose

Ces animations sont déclenchées lorsque la décision de saisie ou dépose correspondante est détectée.

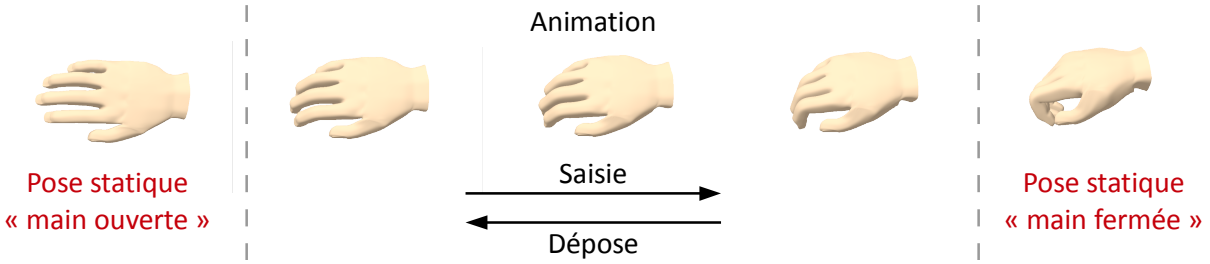


Figure 2.6 - *Animation de la main virtuelle pour la saisie et la dépose*

La suite de ce chapitre détaille chacun des deux couplages nécessaires à la téléopération du système proposé. La section 1.3 décrit le couplage entre la main virtuelle et le simulateur par une méthode de couplage bilatérale inspirée de la robotique classique. Ce couplage correspond aux étapes (2) et (3) de la figure 2.4. La section 1.4 s'intéresse au couplage entre l'utilisateur et la main virtuelle. Ce travail propose d'exploiter une caméra Kinect pour suivre la position de la main de l'opérateur. A partir de ces données, une méthode est proposée pour saisir, déplacer et déposer un micro-objet. Ce couplage correspond aux étapes (1) et (4) du schéma de téléopération de la figure 2.4.

1.3 Couplage main virtuelle-simulateur

La section précédente présente l'interface en réalité virtuelle proposée dans ce travail. Une interface destinée à la manipulation par AFM doit réaliser les tâches suivantes :

- **saisir** un objet par adhésion avec la poutre
- **déplacer** l'objet en évitant les contacts involontaires avec le substrat
- **déposer** l'objet sur un site cible par adhésion avec un autre substrat

Pour téléopérer ce système, cette section présente une méthode de couplage entre la main virtuelle et le simulateur réaliste de micromanipulation. Chacune des tâches de micromanipulation est étudiée.
Cette étape doit prendre en compte la fragilité de la poutre, les efforts qui lui sont imposés sont donc limités. De même, les accélérations sont limitées par les actionneurs robotiques. Une solution consiste à créer un système maître-esclave entre la main virtuelle et la poutre, équivalent à un système ressort-amortisseur virtuel [Hannaford 89].

1.3.1 Système maître-esclave pour la téléopération

Le système maître-esclave a pour avantage de rendre possible le réglage du suivi en position avec deux paramètres, la masse apparente de la poutre et la raideur du ressort. Le bruit de la trajectoire de la main est filtré. Les trajectoires de la poutre générées sont lissées, ce qui est plus adapté aux actionneurs micrométriques.

Cette méthode transforme une erreur de position entre la main virtuelle et la poutre en une force qui minimise cette erreur de position. Le simulateur physique calcule l'accélération et donc le déplacement de la poutre avec cet effort en entrée. Un amortisseur est ajouté pour éviter les vibrations du système.

Pour limiter les oscillations, la valeur du coefficient d'amortissement ζ est fixée à 1, qui correspond à l'amortissement critique :

$$\zeta = \frac{c}{2\sqrt{km}} = 1 \tag{2.1}$$

$$c = 2\sqrt{km} \tag{2.2}$$

où m est la masse de la poutre, k la constante de raideur du ressort virtuel, et c le coefficient d'amortissement

1.3.2 Méthode de saisie/dépose

Le comportement de ce système est adapté pour chacune des trois phases de la tâche de manipulation : déplacement, saisie et dépose de la microsphère sur le substrat.

1.3.2.1 Phase de déplacement

Au début de la tâche, le système maître est la main virtuelle. La poutre suit les déplacements de celle-ci selon deux axes (x,y) dans un plan parallèle au substrat. La coordonnée en z est fixée de façon à éviter les contacts involontaires avec le substrat.

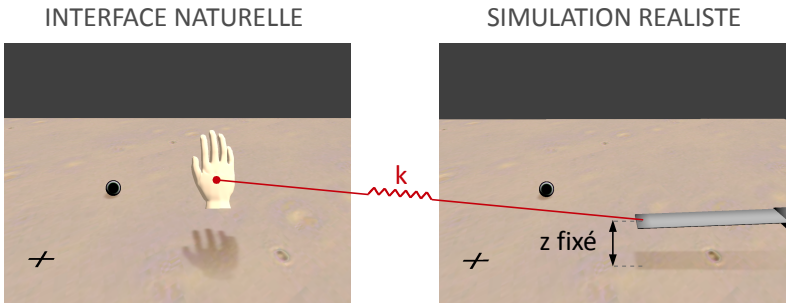


Figure 2.7 - *Phase de déplacement*

1.3.2.2 Phase de saisie

Lorsque la main virtuelle est suffisamment proche d'une microbille, le système maître est positionné sur cette microbille. La poutre vient donc se placer au dessus de celle-ci. Si la décision de saisie est détectée, la contrainte sur la coordonnée z est relâchée. La poutre est libre de descendre au contact de la microbille. Lorsque la main s'éloigne de la bille, le maître est à nouveau positionné sur celle-ci. La figure 2.8 illustre cette séquence.

1.3.2.3 Phase de dépose

Lorsque la main virtuelle est suffisamment proche d'une cible, le système maître est positionné sur cette cible. La poutre vient donc se placer au dessus de celle-ci. Si la décision de dépose est détectée, la contrainte sur la coordonnée z est relâchée. La poutre et la bille saisie sont libres de descendre au contact de la cible. Lorsque la main s'éloigne de la cible, le maître est à nouveau positionné sur celle-ci. Cette séquence est montrée sur la figure 2.9

1.3.3 Retour visuel sur l'échec/le succès de la tâche

Un retour visuel est fourni à l'utilisateur pour l'informer du succès ou de l'échec de la saisie/la dépose de la microbille. Pour cela, l'information de succès de la tâche est transmise à l'interface naturelle. Si la bille a adhéré à la poutre, la main virtuelle est visualisée

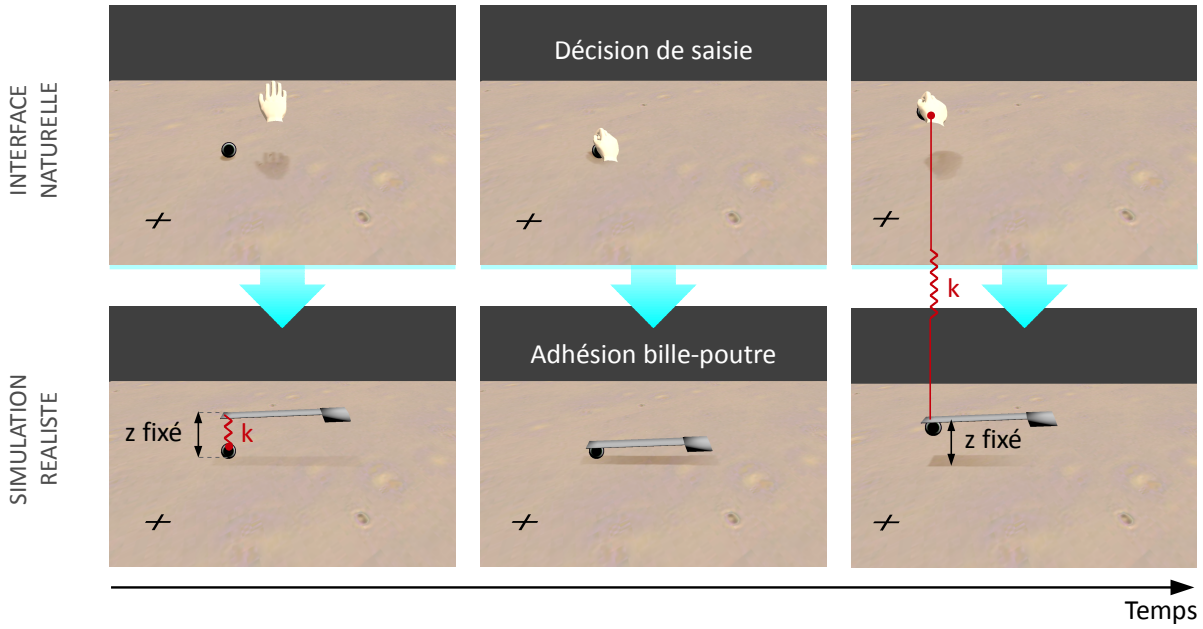


Figure 2.8 - *Phase de saisie*

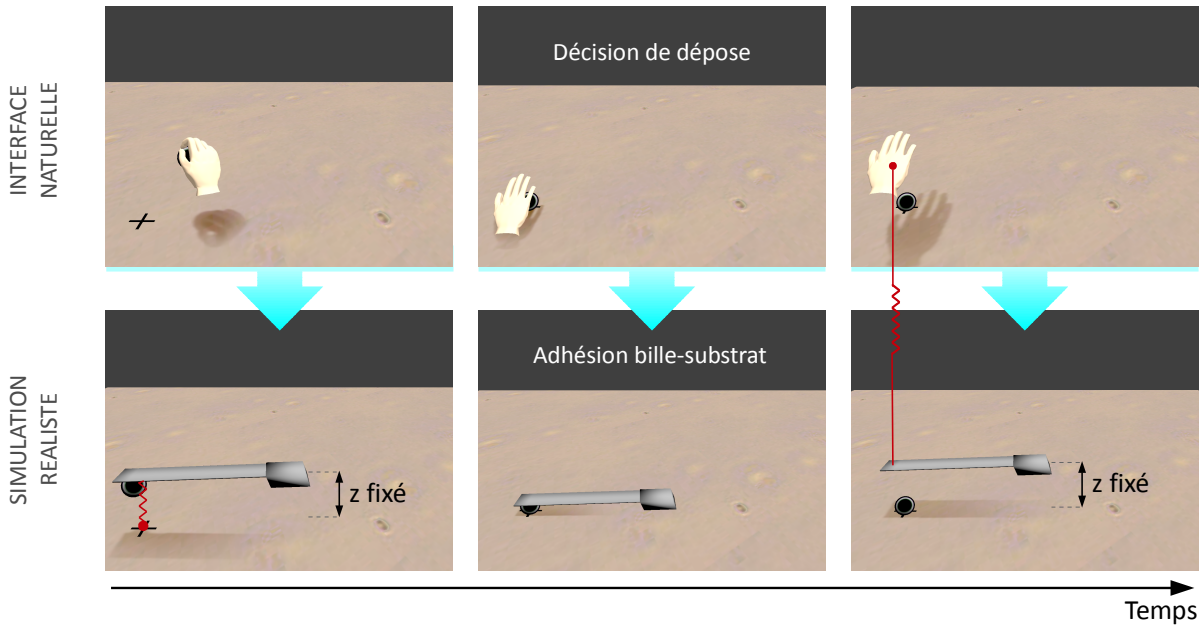


Figure 2.9 - *Phase de dépose*

tenant la bille. Dans le cas contraire, la bille ne change pas de position. L'utilisateur peut retenter l'action de saisie. De même pour la dépose, tant que la bille n'a pas adhéré à la cible sur le substrat, elle reste fixée à la main virtuelle.

1.4 Couplage utilisateur-main virtuelle

Pour réaliser le couplage entre l'utilisateur et la main virtuelle lors de chacune des phases de la téléopération, il est nécessaire de détecter ses actions.
La phase de déplacement correspond à une action continue de la main de l'opérateur. L'état de l'art proposé dans le chapitre I montre que les approches par manipulation directe de bas niveau sont particulièrement adaptées à ce type de tâches.
Contrairement à celles-ci, les tâches de saisie et dépose correspondent à des actions discrètes. Elles nécessitent une reconnaissance des décisions de l'utilisateur.
La méthode adoptée pour réaliser chacune de ces tâches à partir des données acquises par la Kinect est présentée dans les parties suivantes.

1.4.1 La méthode d'acquisition avec le capteur Kinect

Le capteur de vision le plus classique est la caméra RGB. De nombreux algorithmes sont développés pour reconnaître les gestes sur des images et vidéos RGB. Pour une interaction dans un monde en 3 dimensions, le capteur choisi doit être capable de restituer des informations dans les 3 directions de l'espace.

Une solution consiste à utiliser des capteurs 3D qui renvoient directement l'information de distance, sous la forme de cartes de profondeur. Dans ce type d'images, la valeur d'un pixel correspond à la distance du point correspondant de l'objet filmé par rapport à la caméra. De nombreux capteurs 3D de ce type sont disponibles, par exemple le capteur Kinect Microsoft (voir figure 2.10). Ce capteur est initialement destiné au jeu vidéo. Il dispose d'outils logiciels dédiés optimisés pour l'interaction.

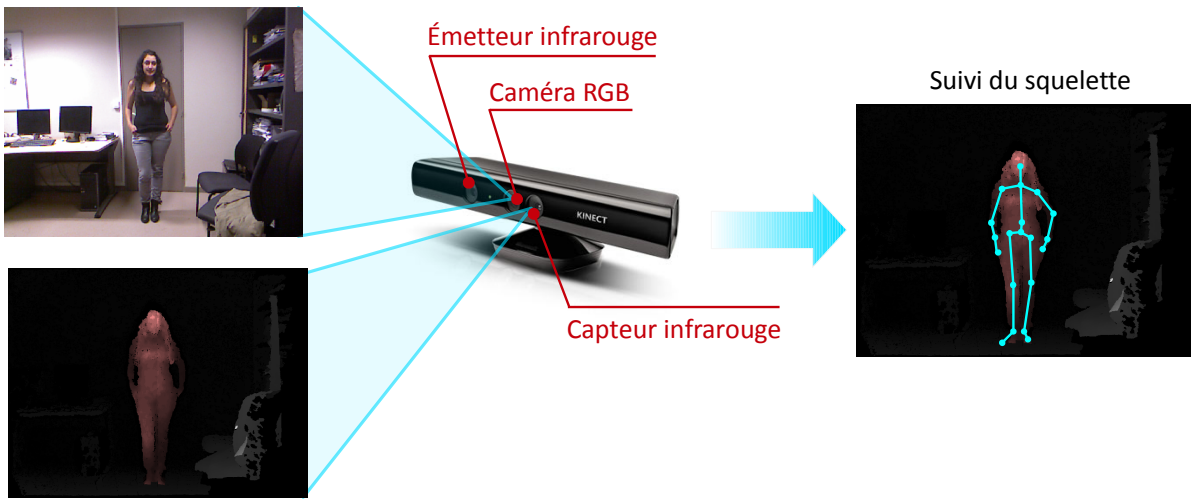


Figure 2.10 - *Articulations suivies par le SDK Micosoft Kinect*

Le capteur Kinect optimisé pour une interaction en trois dimensions et est disponible à bas coût. Il est donc une solution d'acquisition adaptée pour interagir avec la réalité virtuelle.

1.4.1.1 Choix de la librairie

Il existe plusieurs librairies adaptées à l'acquisition des images issues du capteur Kinect. Parmi celles-ci, deux en particulier disposent d'algorithmes de capture du mouvement de l'utilisateur. Le SDK Kinect Microsoft est la solution commerciale fournie avec le capteur. Son alternative principale est la librairie OpenNI qui offre des fonctionnalités similaires.

Ces deux librairies comportent les fonctionnalités suivantes :

- Capture d'une image couleurs (RGB)
- Capture d'une image de profondeur
- Suivi des articulations du squelette de l'utilisateur en 3D (20 articulations)

Le SDK Kinect Microsoft dispose de deux fonctionnalités supplémentaires, c'est donc celui-ci qui est sélectionné :

- Estimation de la pose du visage
- Reconnaissance de gestes de saisie (main fermée) et dépose (main ouverte) rapide (2 ms)

À partir des données acquises, il est possible d'extraire différentes caractéristiques de bas niveau adaptées à l'analyse du geste.

La **position de la main** est approximée par la position de l'articulation du poignet. Elle est extraite à partir de l'outil de suivi du squelette. Chaque articulation est représentée par ses coordonnées en 3 dimensions dans le repère de la Kinect.

Il est possible d'estimer la **vitesse** et l'**accélération** de la main en dérivant la position de l'articulation entre deux images consécutives. Le pas de temps entre deux images successives extraites par le capteur est considéré comme constant.

L'estimation de l'**orientation de la main** est peu robuste avec la Kinect. Elle peut cependant être approximée par la direction coude-poignet. Le SDK Kinect dispose d'une fonction d'estimation de l'orientation des articulations qui est exploitée ici.

1.4.2 La détection des déplacements avec la Kinect

Lors des phases de déplacement, la main virtuelle reproduit les translations et les rotations de la main de l'opérateur. Les caractéristiques de bas niveau extraites sont exploitées.

La position de la main est mise à l'échelle pour déplacer la main virtuelle. Les gains à appliquer sont définis de façon à éviter la fatigue due à des mouvements de grande amplitude. De plus, des études ergonomiques recommandent de ne pas imposer une flexion supérieure à 90°au coude [Cail 12].

À chaque image reçue de la Kinect, la position 3D et la rotation de la main sont envoyées en UDP sur les ports correspondants. La pose 3D de la main est ainsi reconstruite en temps réel d'après les mouvements de l'utilisateur.

1.4.3 La détection des décisions de l'utilisateur

Les déplacements continus de la main virtuelle sont gérés de manière directe d'après les informations de bas niveau extraites par le capteur Kinect. La question principale de ce travail porte sur l'étape suivante, qui consiste à détecter les actions discrètes de l'utilisateur. Dans ce but, deux approches sont comparées pour déterminer les décisions de saisie et de dépose. La première méthode proposée est basée sur la reconnaissance de deux gestes "naturels". Elle fait l'objet de la section suivante. Dans le chapitre III, une nouvelle méthode de prédiction de l'intention de l'utilisateur est proposée. Le simulateur de micromanipulation est utilisé pour mettre en place une expérience utilisateur afin d'analyser et de comparer les performances de ces deux approches de manière qualitative et quantitative.

2 Détection des décisions par reconnaissance de gestes

Le chapitre I montre que l'approche par manipulation directe est insuffisante pour détecter les décisions. Il est nécessaire d'interpréter les gestes de l'opérateur. Dans ce but, l'approche par langage gestuel est couramment exploitée dans le domaine de l'IHM. Cette partie s'intéresse à la taxinomie du geste humain afin de déterminer un dictionnaire de gestes naturels adapté aux tâches de manipulation. Les gestes reconnus sont exploités pour déclencher les actions de saisie et de dépose dans le simulateur de micromanipulation.

2.1 Taxinomie du geste humain

2.1.1 Définition du geste

Mitra définit le geste dans le cadre de l'interaction homme machine comme étant "des mouvement expressifs et significatifs du corps qui impliquent des mouvements physiques des doigts, des mains, des bras, de la tête et du corps avec l'intention de 1) transmettre de l'information significative ou 2) interagir avec l'environnement" [Mitra 07].

2.1.2 Le geste dans l'IHM

Au niveau du geste humain, la plupart des travaux de recherche ont été menés autour des gestes symboliques. Certains travaux portent par exemple sur la reconnaissance de gestes de la langue des signes [Zafrulla 11]. Dans le domaine de l'IHM, l'approche par langage gestuel consiste à définir un dictionnaire de gestes qui correspondent à des commandes données. La figure 2.11 montre un exemple de dictionnaire de gestes proposé par Zeller pour l'interaction avec une simulation moléculaire [Zeller 97] . Ren et O'Neill proposent une méthode de sélection d'objets virtuels en 3D selon l'objet vers lequel se dirige le geste. La sélection est confirmée en levant l'autre main [Ren 13].

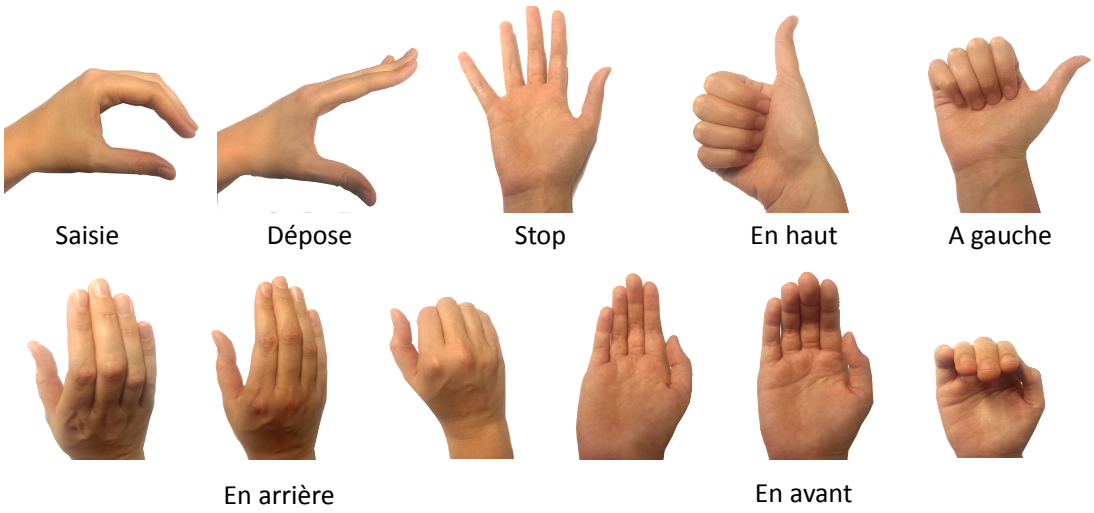


Figure 2.11 - *Exemple de dictionnaire de gestes proposé par [Zeller 97]*

Ces gestes symboliques qui correspondent à un dictionnaire sont dits sémiotiques. Ils ne constituent qu'une partie des gestes réalisés par l'homme. Ils correspondent au point 1) de la définition donnée dans le paragraphe 2.1.1. Dans la taxinomie des gestes, on distingue les gestes ergodiques, qui concernent la manipulation des objets et n'ont pas de but communicatif. Ces derniers correspondent au point 2) de la définition donnée dans le paragraphe 2.1.1.

La figure 2.12 synthétise les résultats de plusieurs travaux sur la classification du geste humain. Pour une interface destinée à la manipulation d'objets virtuels, les gestes impliqués n'ont pas pour vocation de communiquer une information. Il sont donc de type ergodiques. Ces gestes n'ont pas de forme symbolique prédéfinie. Ils peuvent avoir une grande variabilité.

Les approches citées précédemment exploitent des gestes sémiotiques pour réaliser des tâches de manipulation. Ces solutions assimilent donc les gestes ergodiques à des gestes sémiotiques. La première conséquence est que l'utilisateur ne peut pas se comporter de manière naturelle lors de l'interaction avec l'interface, puisque ses gestes doivent se confor-

mer aux entrées du dictionnaire. L'interaction est donc contrainte. La seconde est que ces solutions ne sont pas adaptées à des utilisateurs naïfs, puisqu'ils doivent apprendre et retenir ces différents symboles.

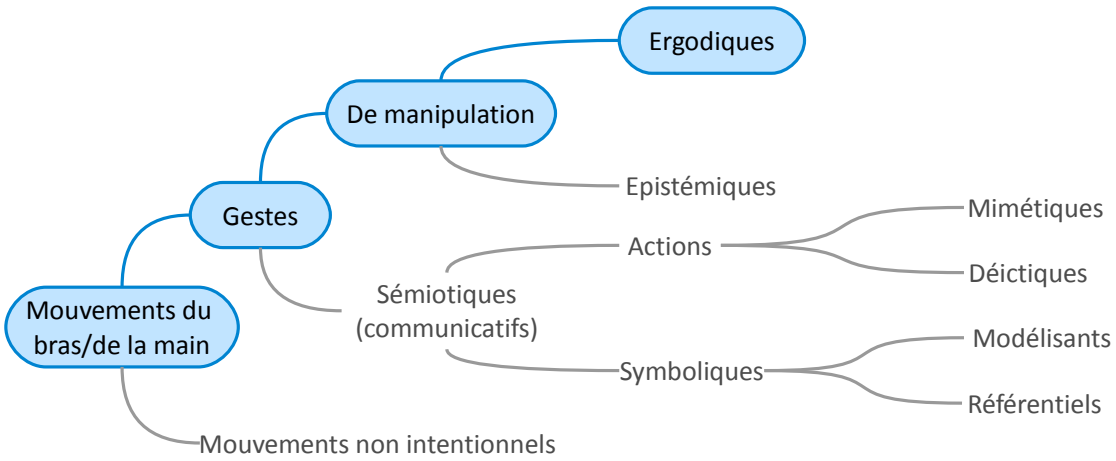


Figure 2.12 - *Taxinomie des gestes*

La plupart des approches existantes se concentrent sur les gestes sémiotiques. En effet, contraindre le geste de l'utilisateur permet de limiter sa variabilité, ce qui facilite sa reconnaissance. De plus, en limitant les gestes à un dictionnaire prédéfini, il est plus facile de segmenter les mouvements qui sont porteurs de sens. Pour créer une interface naturelle, l'enjeu consiste à faire face à deux contraintes :

- reconnaitre des gestes naturels de manipulation (ergodiques) et pas un symbole gestuel (sémiotique), c'est-à-dire ne pas définir de dictionnaire a priori que l'utilisateur doit mémoriser
- ne pas contraindre la pose de l'utilisateur, donc être capable de reconnaître des gestes non contraints malgré leur variabilité

2.2 La méthode par reconnaissance de gestes

Pour éviter de contraindre l'utilisateur à apprendre des gestes symboliques, certains travaux cherchent à établir un dictionnaire de gestes naturels [Aigner 12]. Les auteurs proposent d'étudier les préférences gestuelles d'un grand nombre d'utilisateurs lors de différentes tâches.
En se basant sur des gestes définis directement par les utilisateurs, ces travaux ont pour but de créer des interfaces plus naturelles. Ce point de vue est adopté dans cette partie. Cette approche est par la suite évaluée dans le contexte de l'interface de micromanipulation d'un point de vue qualitatif et quantitatif.

La tâche de micromanipulation sous AFM implique la saisie et la dépose de microsphères. D'après les travaux d'Aigner et al., les gestes associés à ces actions par la majorité

des sujets sont la fermeture de la main pour saisir un objet et l'ouverture pour la dépose. Il s'agit donc de reconnaître des gestes statiques "main ouverte" et "main fermée".
La bibliothèque du SDK Microsoft de reconnaissance de gestes permet une reconnaissance robuste et invariante en fonction de la pose de la main, qui correspond à l'état de l'art dans le domaine de la reconnaissance de gestes statiques. Elle repose sur la méthode des forêts d'arbres décisionnels [Shotton 13].

La figure 2.13 montre une séquence de téléopération qui exploite cette approche pour détecter la décision de saisie.

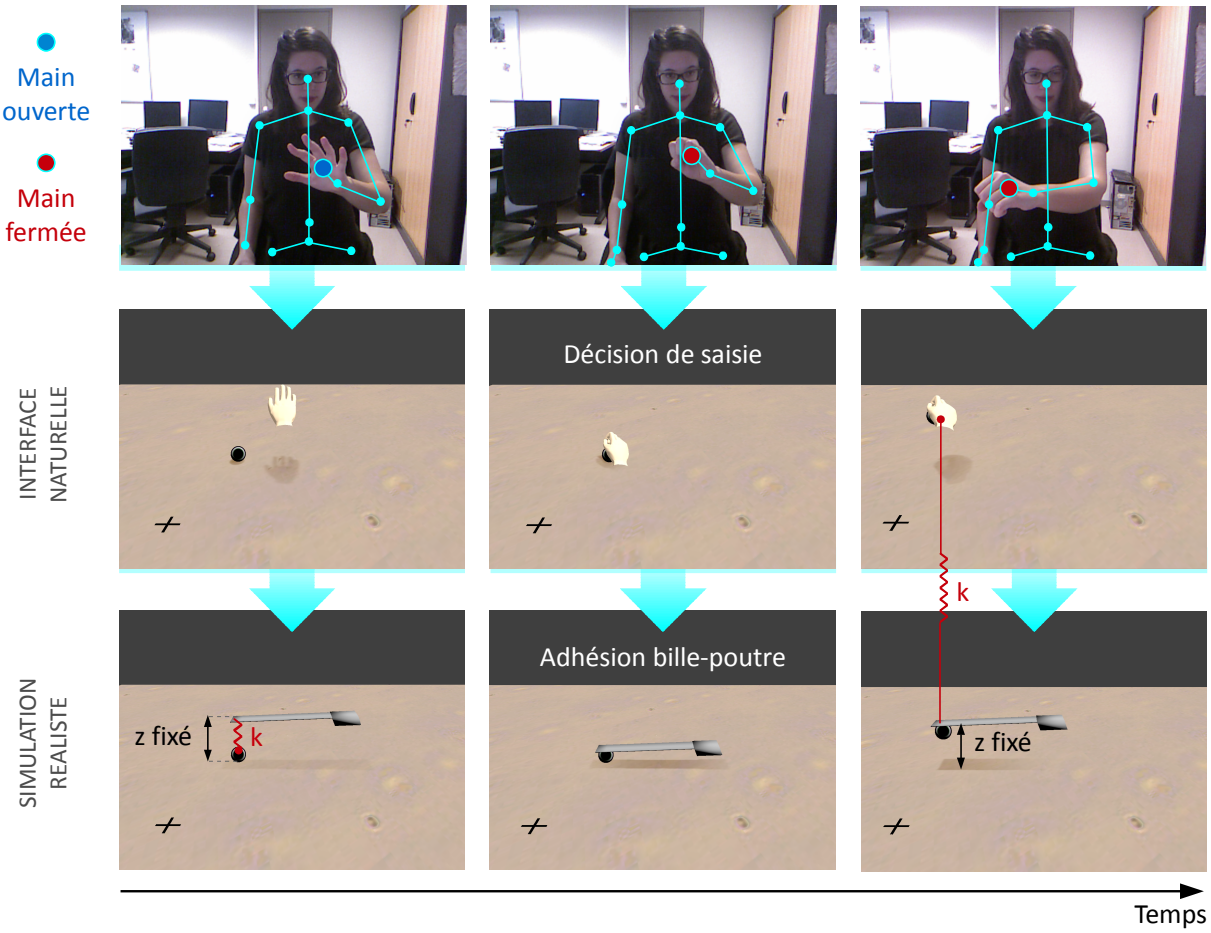


Figure 2.13 - *Séquence de téléopération avec l'approche par reconnaissance de gestes main ouverte et main fermée*

3 Expériences utilisateur

Pour évaluer la méthode de reconnaissance des décisions basée sur un dictionnaire de gestes naturels, un protocole expérimental est mis en place. En particulier, cette expérience

a pour but de déterminer si ce système est adapté pour créer une interface naturelle du point de vue de l'utilisateur.

3.1 Méthode d'évaluation

La tâche consiste à saisir une microsphère et la déplacer jusqu'à une cible marquée par une croix sur le substrat. Une nouvelle configuration de la microsphère et de la cible est tirée aléatoirement à la fin de chaque tour. Pour valider l'aspect naturel de l'approche proposée dans le cadre de la tâche de micromanipulation sous AFM, une nouvelle méthode d'évaluation expérimentale est proposée. Cette méthode repose sur 3 principes :

- **L'absence de consigne à l'utilisateur**
  Aucune indication ne lui est donnée sur la méthode à utiliser pour manipuler les objets. La variabilité du geste ergodique n'est donc pas contrainte. De plus, cette méthode valide le fait qu'aucun dictionnaire de geste ne doit être appris pour interagir.
- **La réussite de tâches spécifiques de manipulation**
  L'approche proposée doit permettre à l'utilisateur de réaliser les sous-tâches de micromanipulation. Le taux de réussite de ces sous-tâches ainsi que la durée nécessaire à leur réalisation sont des indicateurs significatifs de la pertinence de l'approche évaluée.
- **Le retour utilisateur**
  L'évaluation par des questionnaires utilisateurs donne un retour subjectif complémentaire à l'évaluation quantitative. Elle rend possible le dialogue avec l'utilisateur final pour mettre au point une interface qui lui est adaptée.

3.2 Protocole du test utilisateur

Un protocole de test utilisateur est mis en place pour évaluer la méthode d'interaction par reconnaissance de gestes.

Le protocole est conduit en deux phases :

- **La détection de la proximité**
  Dans le protocole expérimental choisi, la cible de l'utilisateur et sa décision sont connus à chaque instant. Il est donc possible de mettre en place un cas témoin dans lequel la saisie et la dépose sont déclenchées simplement par un critère de proximité entre la main virtuelle et la cible qui est connue. Dans le cadre de ce cas témoin, la tâche est réalisable indépendamment du geste et de la pose de la main.
- **L'approche par reconnaissance de gestes**
  Dans la seconde phase, la méthode classique de reconnaissance de gestes décrite dans la section 2.2 est utilisée.

L'ordre des phases présentées à chaque utilisateur est modifié de manière aléatoire afin d'éviter un effet d'apprentissage dans les résultats. Chaque phase est constituée de 15 tours de saisie et dépose. A la fin de chaque phase, un questionnaire est soumis à l'utilisateur. Ce questionnaire est basé sur le System Usability Scale (SUS) [Brooke 96]. Il explore plusieurs axes d'évaluation : la facilité d'utilisation, le confort, l'immersion dans l'interface et la rapidité à maîtriser l'interface. Neuf sujets adultes ont participé à l'expérience.

3.3 Résultats quantitatifs

Le succès de la tache est évalué sur un critère basé sur la proximité et la durée : si la main reste proche de la cible pendant plus de $0.5s$ sans que la saisie/dépose soit détectée, la tâche est considérée comme un échec. Le pourcentage de réussite est montré sur la figure 2.14 à gauche.

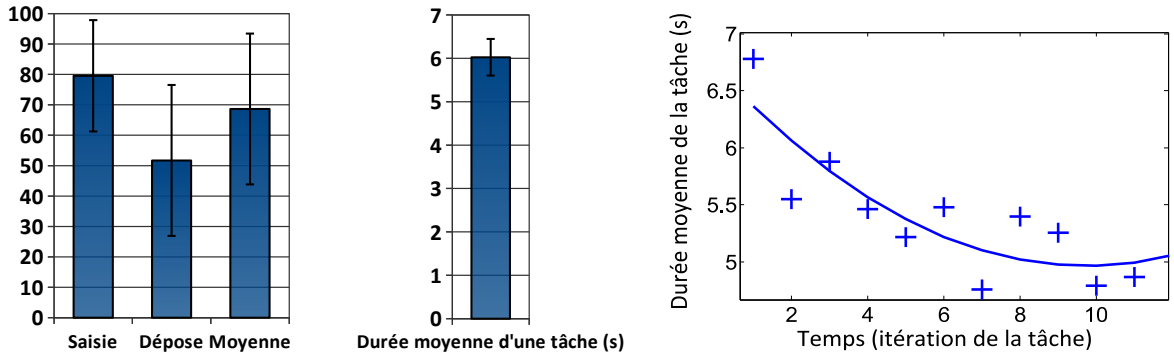


Figure 2.14 - *Pourcentage de succès des deux tâches avec la méthode par reconnaissance de gestes (à gauche) et durée moyenne d'une tâche (au milieu). Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. La courbe de droite montre l'évolution de la durée moyenne de l'ensemble des utilisateurs pour réaliser une tâche en fonction du temps*

Le pourcentage de succès de la saisie avec l'approche par reconnaissance de gestes est de 79.5%. La saisie est donc bien reconnue avec cette méthode. Cependant, la reconnaissance de la dépose est peu fiable. Son pourcentage de succès est 1.5 fois plus faible (51.7%). Ce résultat est expliqué par le fait que sans consigne préalable, l'utilisateur a tendance à ne pas ouvrir complètement la main pour déposer l'objet. De nombreux faux positifs de main ouverte sont donc observés. L'approche par reconnaissance de gestes nécessite $6.1s$ en moyenne pour réaliser la tâche, comme montré sur la figure 2.14 au milieu.

Une indication de l'effet d'apprentissage consiste à observer l'évolution de la durée nécessaire pour accomplir une tâche lors des répétitions de celle-ci. Les résultats obtenus sont illustrés par la figure 2.14 à droite.

Après une interpolation polynomiale d'ordre 3, la courbe obtenue est décroissante, ce

qui indique que plusieurs répétitions sont nécessaires à l'utilisateur pour apprendre à réaliser la tâche. De plus, après une dizaine de répétitions, la durée moyenne atteint une limite de 4.7s.

3.4 Résultats qualitatifs

Les notes attribuées par l'ensemble des utilisateurs sont moyennées pour évaluer un score global du test SUS sur 100. Ces résultats sont illustrés sur la figure 2.15.

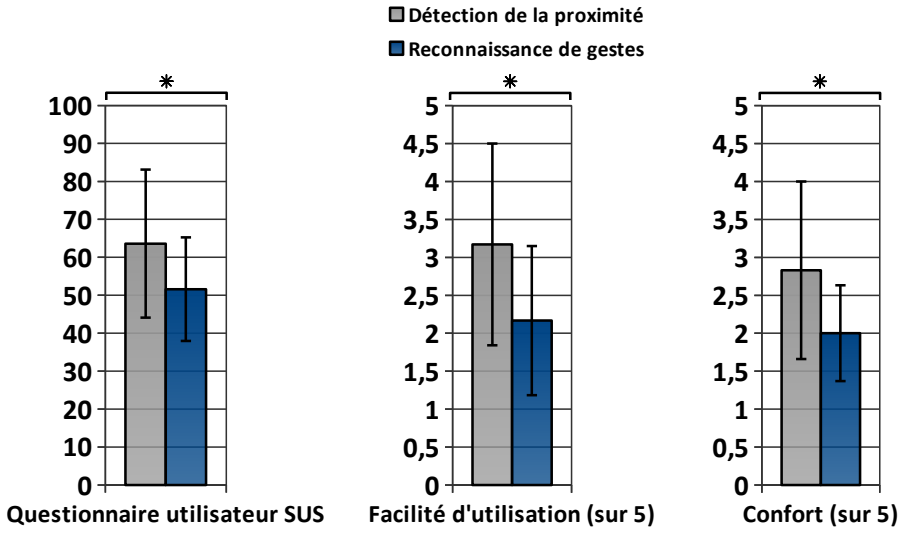


Figure 2.15 - *Résultats du test utilisateur SUS. Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA.*

Le questionnaire montre une préférence de 10% pour l'approche témoin. La reconnaissance de gestes obtient un score proche de 50%. En particulier, la facilité d'utilisation et le confort ont obtenu des notes inférieures à la moyenne dans le cas de l'approche par reconnaissance de gestes. Ainsi, une approche simplement basée sur la détection de la proximité avec l'objet cible semble plus naturelle et intuitive aux utilisateurs que la méthode par reconnaissance de gestes.

3.5 Analyse des résultats

Les résultats obtenus indiquent que l'approche par reconnaissance de gestes implique une phase d'apprentissage, puisque 10 répétitions sont nécessaires à l'utilisateur pour apprendre à réaliser la tâche. De plus, les utilisateurs évaluent cette approche comme peu naturelle et intuitive, comme l'indiquent les résultats qualitatifs obtenus dans le paragraphe 3.4. La méthode classique de reconnaissance de geste ne répond donc pas aux enjeux décrits dans le chapitre I.

Certaines remarques peuvent expliquer des résultats. Une première observation est que le geste n'est reconnu que lorsqu'il est terminé. L'animation correspondante de la main virtuelle ne peut donc être déclenchée qu'après la fin du geste. Ainsi, les utilisateurs mentionnent une impression de retard entre le geste réalisé et le retour visuel, bien que la reconnaissance de geste fonctionne en temps réel. Une autre remarque est que l'opérateur a tendance à ne pas fermer totalement la main lors de la saisie, comme pour adapter l'ouverture à la taille de l'objet virtuel qu'il veut saisir. L'interface semble ainsi assez proche d'une interaction réelle pour que l'opérateur se comporte comme s'il interagissait avec de vrais objets.

Ces indices conduisent à penser qu'il existe une dualité entre le comportement naturel de l'utilisateur envers l'interface et le langage symbolique que celle-ci utilise. Cette observation montre qu'il n'est pas suffisant de réduire un comportement naturel comme la saisie/dépose d'un objet à la reconnaissance d'une main ouverte/fermée. Alors que le caractère naturel de ces gestes semble attesté, cette méthode n'est pas suffisante pour obtenir une interface non symbolique. Le dictionnaire de geste discrets main ouverte/fermée reste un dictionnaire symbolique, et impose à l'utilisateur l'apprentissage de ces symboles pour communiquer. Il ne s'agit donc pas de la reconnaissance des gestes ergodiques précédemment définis, mais de la reconnaissance de gestes sémiotiques.

4 Conclusion

Ce chapitre propose d'inclure une interface naturelle en réalité virtuelle entre l'opérateur et un système de micromanipulation par contact adhésif. Cette interface masque la complexité du système réel. Elle inclut une main virtuelle afin de faciliter le lien entre la main de l'utilisateur et l'effecteur dans la scène de manipulation. Des méthodes de couplage utilisateur-main virtuelle et main virtuelle-scène de micromanipulation sont proposées. L'acquisition est réalisée par des méthodes de vision par ordinateur avec un capteur Kinect.

Au niveau du couplage utilisateur-main virtuelle, deux types de tâches sont distinguées. Les actions continues telles que les déplacements sont détectées directement à partir des informations de position et de rotation de la main de l'opérateur extraites.
Les décisions de l'utilisateur qui correspondent à des actions discrètes (saise, dépose) ne sont pas extractible directement à partir du capteur. Pour les détecter, une première approche consiste à reconnaître des gestes "main ouverte" et "main fermée". Une expérience utilisateur est mise en place pour évaluer le caractère naturel de cette méthode d'un point de vue qualitatif et quantitatif dans le cadre de la micromanipulation. Les résultats obtenus montrent que cette approche est considérée comme peu naturelle par les utilisateurs. D'un point de vue quantitatif, le taux de reconnaissance des décisions est faible. Il semble ainsi qu'établir un dictionnaire de gestes symboliques ne soit pas suffisant pour créer une interface naturelle.

Le chapitre III propose une approche basée sur l'analyse du comportement naturel de l'opérateur. Pour détecter les décisions de manière non symbolique, une méthode de prédiction de l'intention est proposée.

Chapitre 3

Une interface basée sur la prédiction de l'intention

Sommaire

1 La micromanipulation téléopérée par contact adhésif 28
1.1 Le principe de la micromanipulation par adhésion 28
1.2 Un simulateur physique pour l'évaluation 30
1.3 Couplage main virtuelle-simulateur 34
1.4 Couplage utilisateur-main virtuelle 37
2 Détection des décisions par reconnaissance de gestes 39
2.1 Taxinomie du geste humain 39
2.2 La méthode par reconnaissance de gestes 41
3 Expériences utilisateur 42
3.1 Méthode d'évaluation 43
3.2 Protocole du test utilisateur 43
3.3 Résultats quantitatifs 44
3.4 Résultats qualitatifs 45
3.5 Analyse des résultats 45
4 Conclusion 46

Le chapitre précédent montre qu'il est insuffisant d'utiliser une reconnaissance de gestes "main ouverte" et "main fermée" pour obtenir une interface naturelle de micromanipulation. En conséquence, le taux de reconnaissance des décisions de l'utilisateur est faible. Au niveau qualitatif, cette méthode est peu naturelle et intuitive du point de vue des utilisateurs. Enfin, il est nécessaire à l'utilisateur de répéter plusieurs fois l'action pour

apprendre à maîtriser l'interface. Cette approche ne répond donc pas aux enjeux d'une interface non symbolique tels que définis dans le chapitre 1.

En particulier, deux problèmes principaux sont relevés. Le dictionnaire de gestes discrets main ouverte/fermée reste un dictionnaire symbolique, et impose à l'utilisateur l'apprentissage de ces symboles pour communiquer. Cette approche n'est donc pas adaptée à la reconnaissance de gestes ergodiques. D'autre part, la reconnaissance de gestes n'implique aucune possibilité d'anticiper les actions de l'utilisateur. Ainsi, il existe un décalage temporel entre ces actions et leurs effets dans l'interface. La maîtrise de l'interaction en devient difficile.

L'objectif de ce chapitre est la proposition d'une nouvelle approche non symbolique et prédictive pour dépasser ces limites. Cette approche est basée sur un modèle cognitif computationnel de reconnaissance de l'intention [Oztop 05]. Il s'agit de s'inspirer du fonctionnement du cerveau humain, capable d'affecter du sens aux actions non symboliques d'autrui. De plus, l'être humain reconnaît le but d'une action de manière prédictive, avant que celle-ci ne soit terminée. Cet aspect prédictif est une propriété intéressante pour créer une interface qui anticipe les actions de l'utilisateur. Une telle interface serait capable d'annuler le délai observé entre l'action réelle de l'utilisateur et l'animation de la main virtuelle.

L'intention est une notion de haut niveau, qui ne peut pas être extraite directement à partir des capteurs. Elle est donc modélisée à partir de signaux comportementaux extractibles de bas niveau. La première difficulté consiste à sélectionner des signaux de bas niveau à la fois suffisamment invariants et caractéristiques de l'intention. Dans ce but, une étude des invariants du geste ciblé est réalisée dans la section 1. Dans la section 2, les signaux sont détectés avec le capteur Kinect et validés expérimentalement dans le simulateur présenté dans le chapitre II. Un modèle computationnel de prédiction de l'intention est proposé dans la section 3 à partir des signaux validés. Une évaluation de l'interface basée sur ce modèle complète ce chapitre.

1 Sélection des signaux de bas niveau pour modéliser l'intention

1.1 Définition fonctionnelle de l'intention

Une action est composée de deux parties : une intention et un mouvement. Dans le cas d'une action délibérée, le mouvement est causé par une **intention a priori** [1]. Il s'agit d'une intention d'agir, formée avant la réalisation de l'acte en lui-même [Searle 83]. Un

1. Le terme "intention a priori" est une traduction qui fait référence à la notion de "prior intention" proposée par J. Searle. La notion d'"intention in action" est ici traduite par "intention dans l'action" [Searle 83].

même mouvement peut être causé par différentes intentions. Par exemple, une personne peut saisir une pomme pour la manger ou pour la donner à quelqu'un. Ce processus, initié par l'intention a priori et finalisé par l'action, est illustré sur la figure 3.1. Becchio et al. montrent qu'un observateur humain peut inférer l'intention d'un acteur en observant ses actions [Becchio 12]. Dans la suite de ce travail, cette capacité est appelée **prédiction de l'intention**, puisque l'observateur peut prédire le but de l'action avant même que celle-ci ne soit terminée.

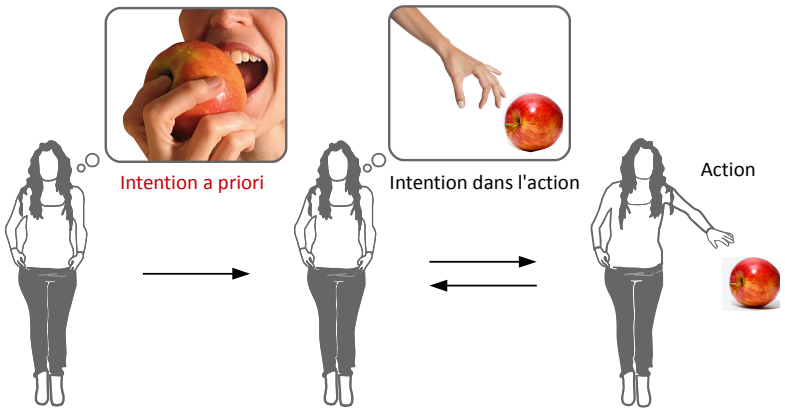


Figure 3.1 - *L'intention a priori "Je veux manger cette pomme" déclenche une intention dans l'action de saisie qui provoque une action motrice du bras en direction de la pomme.*

L'intention est une notion de haut niveau. Elle nécessite une interprétation des données de bas niveau issues des capteurs. Être capable de spécifier les signaux pertinents qui encodent l'intentionnalité est un pré-requis pour pouvoir modéliser l'intention. Cette partie s'intéresse à déterminer ces signaux comportementaux caractéristiques de l'intention dans le contexte de l'interface de micromanipulation.

L'identification de ces signaux doit satisfaire deux conditions.

1. Ils doivent être **caractéristiques de l'intention** de l'utilisateur. Différents travaux qui s'intéressent à cet aspect sont étudiés dans la sous-section 1.2
2. Les signaux sélectionnés doivent être suffisamment **invariants** pour une intention donnée pour que la reconnaissance soit possible malgré la variabilité du geste. Ce point est traité dans la sous-section 1.3.

1.2 État de l'art des signaux caractéristiques de l'intention

Dans le cadre de la micromanipulation, deux intentions sont considérées : les intentions de saisie et de dépose. Ces actions sont deux cas particuliers du geste ciblé chez l'être humain. L'objectif de cette sous-section est de déterminer les signaux pertinents pour modéliser l'intention par une étude du comportement intentionnel humain.

Pour déterminer les signaux caractéristiques de l'intention, deux types d'expériences sont proposées dans la littérature. La première catégorie consiste à observer le compor-

tement d'un acteur qui réalise ces tâches. Il est ainsi possible de déterminer les signaux comportementaux qui varient selon son intention. Ces expériences sont présentées dans la sous section suivante. Dans la seconde catégorie d'expériences, un acteur réalise la tâche en présence d'un observateur. Ces expériences ont pour but d'identifier les signaux exploités par l'observateur pour prédire l'intention de l'acteur. Pour isoler la contribution d'un signal donné, ces travaux réalisent une occlusion spatiale ou temporelle du geste de l'acteur. Les signaux identifiés suivant cette approche sont présentés dans la sous section 1.2.2 page 53.

1.2.1 Étude de l'acteur

Becchio et al. proposent de distinguer deux phases dans le geste de saisie d'un objet. La première est la phase d'atteinte et la seconde la phase de saisie avec fermeture des doigts [Becchio 10]. Ces travaux montrent que ces deux phases sont influencées par l'intention. L'expérience proposée compare des signaux cinématiques lors de tâches individuelle (prendre un objet pour soi) ou sociale (prendre un objet pour le donner à quelqu'un) ainsi que des tâches de compétition ou de coopération. Les résultats obtenus montrent une amplitude de la vitesse du poignet supérieure pour les tâches individuelles par comparaison aux tâches sociales (fig. 3.2).

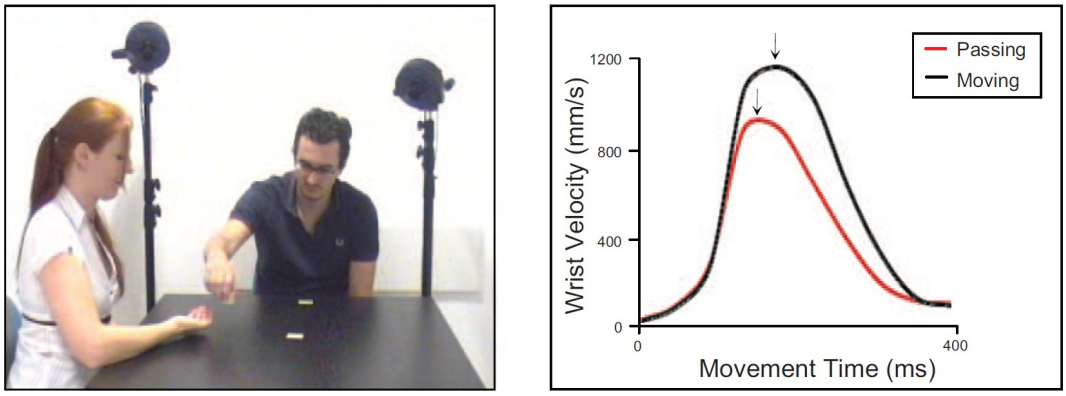


Figure 3.2 - *Tâche d'atteinte et de saisie pour évaluer si la cinématique du geste est dépendante de l'intention sociale ou individuelle (à gauche) et résultats expérimentaux (à droite) [Becchio 10]*

Ansuini et al. comparent les mouvements des doigts à l'approche d'une bouteille d'eau lors de différentes tâches [Ansuini 08]. Cinq conditions sont explorées : une simple saisie sans action ultérieure, déplacer la bouteille, jeter la bouteille, remplir un récipient d'eau et donner la bouteille à l'expérimentateur. Il montre que le positionnement des doigts à l'approche et au contact varie selon la tâche planifiée. De plus, la durée totale de la phase d'atteinte dépend de l'intention. Par exemple, lorsqu'une action ultérieure n'est pas requise (simple saisie), la durée de la phase d'atteinte est plus longue, comparée à toutes les autres conditions.

L'intention de communiquer une information influence les caractéristiques cinématiques

du geste [Sartori 09]. Sartori et al. proposent une expérience de saisie d'objets en distinguant deux conditions. Dans le premier cas, l'objet est saisi pour communiquer une information à un partenaire. Le second cas est une condition contrôle sans interaction avec un partenaire. En particulier, les résultats obtenus montrent que le maximum d'ouverture de la main est atteint plus tardivement lorsque le sujet a une intention communicative. La même observation est effectuée pour le point le plus haut de la trajectoire du poignet. Les auteurs concluent qu'il existe une tendance à planifier plus minutieusement l'approche de l'objet lorsque l'action est destinée à être reconnue par le partenaire.

Les différents signaux issus de ces travaux sont synthétisés dans le tableau 3.1.

	Signaux bas niveau	**Signaux haut niveau**
[**Becchio 12**]	Vitesse du poignet : amplitude, instant du pic	Intention sociale et individuelle
[**Sartori 09**]	Cinématique du geste : trajectoire, vitesse d'ouverture des doigts	Intention communicative et individuelle
[**Ansuini 08**]	Cinématique des doigts (angles, vitesses) Durée du geste	Intention lors de tâches de manipulation d'une bouteille (boire, jeter, déplacer, donner)

Tableau 3.1 - *Signaux de bas niveau impliqués dans les tâches de manipulation*

Les signaux relevés donnent une idée des modalités pertinentes à exploiter pour construire un modèle de prédiction de l'intention. En particulier, deux types de caractéristiques sont distinguées : les caractéristiques cinématiques du mouvement du poignet lors de la phase d'approche de l'objet, et les caractéristiques cinématiques des doigts lors de la saisie.

1.2.2 Étude de l'observateur

Pour déterminer les signaux de bas niveau caractéristiques de l'intention, une autre approche consiste à s'interroger sur les signaux exploités par l'homme pour inférer l'intention de l'autre.

Sartori et al. proposent d'évaluer l'impact respectif de l'observation des mouvements du bras et du visage pour discriminer les intentions [Sartori 11]. Dans ce but, les sujets observent des vidéos d'acteurs qui réalisent des tâches de saisie avec différentes intentions : la coopération, la compétition ou un but individuel. Sur ces vidéos, le bras ou le visage de l'acteur est masqué. Pour chaque type d'occultations, l'observateur doit reconnaître l'intention. Les résultats montrent que des **indices sur le bras** sont plus informatifs lorsqu'il s'agit de discriminer les intentions individuelles. D'autre part, le visage est plus

informatif pour reconnaitre les intentions de coopération et de compétition.
Une autre hypothèse est la capacité de l'être humain à **déterminer l'intention de manière prédictive**, avant que le but soit observé. Une seconde expérience est alors proposée. Les vidéos sont présentées avec une occultation temporelle de la fin du geste. Les résultats obtenus montrent que l'observateur est capable de discriminer l'intention lors de la phase d'atteinte de l'objet.

Les données cinématiques donnent aussi à un observateur extérieur la possibilité d'obtenir des informations sur l'objet cible, par exemple sa taille, même lorsque cet objet n'est pas visible [Campanella 11]. L'observateur semble exploiter principalement des données cinématiques. De nouvelles expériences sont menées avec des vidéos où des points lumineux sont placés au niveau des articulations (fig. 3.3) [Manera 10]. Les performances ainsi obtenues sont proches des résultats sur des vidéos classiques. Cette remarque montre l'importance de la **cinématique du geste** pour la prédiction de l'intention chez l'homme. Les auteurs montrent que ces points lumineux sont suffisants à l'observateur pour distinguer des intentions communicatives et non communicatives.

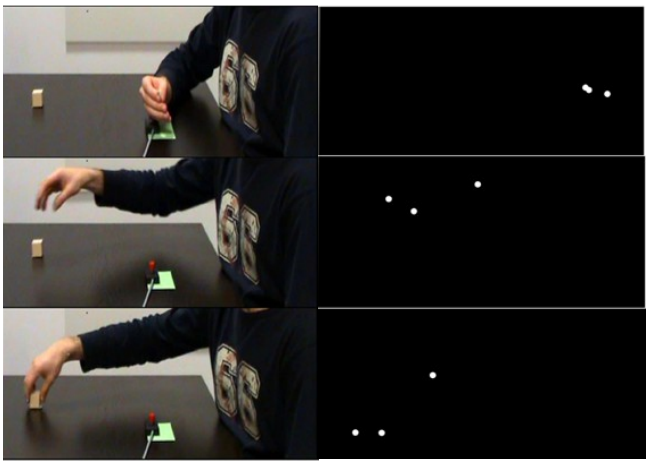


Figure 3.3 - *Images extraites des vidéos montrées à l'observateur. A droite, seuls des points lumineux sont visibles. Ils correspondent aux articulations du pouce, de l'index et du poignet [Manera 10].*

D'autres travaux [Stapel 12] évaluent l'importance du **contexte de l'action** pour prédire l'intention. L'observateur est capable de prédire l'intention de manière plus précise si l'action est contrainte par une cible visible et un contexte précis. Cependant, les auteurs montrent que les observateurs humains exploitent d'avantage la cinématique du geste que des informations visuelles directes sur l'objet cible et le contexte.

Ces différents résultats sont synthétisés dans le tableau 3.2.

Le bras semble porter plus d'informations que le visage pour reconnaître les intentions individuelles. Il est ainsi essentiel d'être capable de détecter des informations cinématiques du mouvement du bras pour déterminer l'intention dans le cadre de tâches de manipulation. D'autre part, il est suffisant d'observer les articulations pour reconnaître l'intention. **La détection des articulations du squelette par le SDK Kinect apparaît ainsi**

	Signaux bas niveau	Haut niveau
[**Sartori 11**]	Mouvements du bras, visage	Intention de coopération, compétition et individuelle
[**Campanella 11**]	Cinématique des doigts	Taille de l'objet visé
[**Manera 10**]	Cinématique des doigts	Intention communicative et non communicative
[**Stapel 12**]	Contexte de l'action, cinématique du geste, objet cible	Intention de saisie

Tableau 3.2 - *Signaux de bas niveau exploités par l'être humain pour l'inférence de différents types d'intention*

comme une solution adaptée à l'acquisition des signaux de bas niveau. Les travaux cités montrent en particulier l'importance de la position des doigts et du poignet pour reconnaître l'intention. Cependant, les solutions de l'état de l'art en vision par ordinateur pour détecter les doigts sont peu précises ou ne fonctionnement pas en temps réel. **Ce travail se concentre donc sur la position du poignet détecté avec le capteur Kinect.**

1.3 Les signaux invariants du geste ciblé humain

Pour réaliser un modèle prédictif de l'intention, la section précédente relève les signaux de bas niveau pertinents pour reconnaître l'intention. En particulier, les informations cinématiques de la main sont relevées. Cette partie s'attache à déterminer les invariants de la cinématique de la main lors du geste ciblé chez l'homme dans le cadre de tâches de manipulation.

1.3.1 Profil gaussien de la vitesse

Lors d'un geste dirigé vers une cible, la vitesse de la main suit une loi log-normale. Cette dernière peut être approximée par une gaussienne pour des mouvements de vitesse intermédiaire [Nagasaki 89]. Ce profil classique est illustré sur la figure 3.4. La courbe est tracée à partir d'un geste ciblé dans le simulateur en réalité virtuelle décrit dans le chapitre précédent.

À partir de ce profil classique, l'équation suivante exprime l'amplitude de la vitesse en

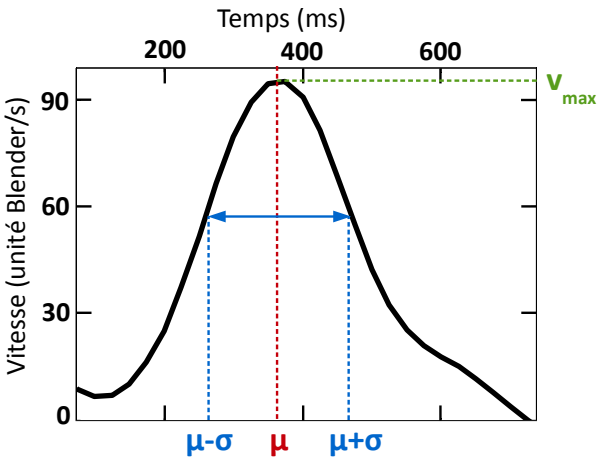


Figure 3.4 - *Le profil gaussien invariant de la vitesse de la main pour les gestes ciblés. La moyenne est notée μ, l'écart type σ et le maximum v_{max}. L'unité de longueur par défaut du logiciel Blender est appelée "unité Blender". Elle n'a pas d'équivalent dans le monde réel.*

fonction du temps lors de l'atteinte d'une cible :

$$v(t) = v_{max} \cdot \mathrm{e}^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \tag{3.1}$$

avec μ la moyenne, σ l'écart type, v la vitesse de la main et v_{max} la vitesse maximum (fig. 3.4).

Dans ce travail, la vitesse est définie comme la dérivée de la distance entre la main et la cible. Elle est ainsi invariante quelle que soit la position relative de la cible par rapport à la main lors de l'interaction :

$$v(t) = \frac{d(dist(t))}{dt} \tag{3.2}$$

avec $dist(t)$ la distance entre la main virtuelle et la cible à un instant t.

1.3.2 Loi d'isochronie du mouvement

Une autre propriété des geste ciblés est la loi d'isochronie du mouvement [Viviani 95]. Il est démontré que la durée d'un geste d'atteinte d'une cible est une constante. Ainsi, la vitesse varie linéairement en fonction de la distance initiale à la cible. Ce résultat est validé avec le simulateur en réalité virtuelle d'après 60 gestes d'atteinte d'une cible. Le résultat est montré sur la figure 3.5.

L'équation linéaire suivante exprime la variation du maximum de la vitesse de la main en fonction de la distance initiale à la cible.

$$v_{max}(d_0) = a \cdot d_0 + b \tag{3.3}$$

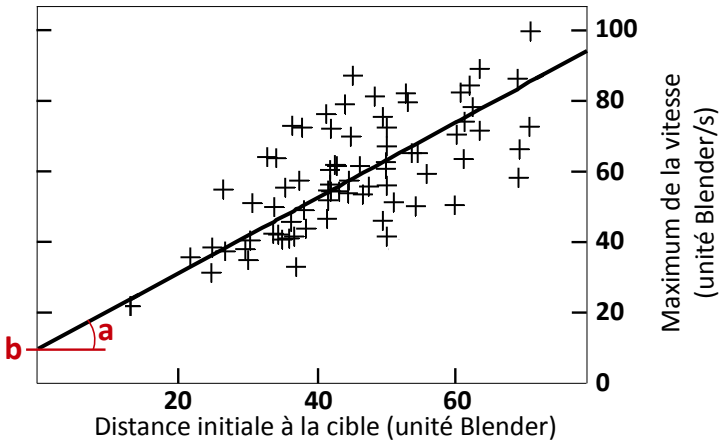


Figure 3.5 - *Illustration du principe d'isochronie du mouvement pour les gestes ciblés. La courbe est tracée à partir de 60 gestes d'atteinte d'une cible dans le simulateur.*

avec a le coefficient directeur, b l'ordonnée à l'origine et d_0 la distance initiale entre la main et la cible (fig. 3.5).

Parmi les signaux de bas niveau caractéristiques de l'intention, ce travail se concentre sur la **vitesse de la main**. En effet, ce signal est **caractéristique de l'intention**. De plus, il est directement **extractible à partir du capteur Kinect** de manière robuste. Enfin, il comporte des **invariants** qui sont exploités dans la suite de ce travail.

La saisie et la dépose sont deux exemples de tâches qui correspondent à un geste ciblé. Les invariants relevés ne dépendent pas de ces tâches. Cependant, il est intéressant de noter que l'être humain est capable de prédire l'intention malgré ces invariants du geste. Une hypothèse consiste à supposer qu'il existe des modulations de la forme gaussienne de la vitesse et de la loi d'isochronie de mouvement qui dépendent de l'intention. Il serait ainsi possible de reformuler les lois invariantes décrites en prenant en compte la tâche. La validation de cette hypothèse avec le simulateur proposé dans le chapitre précédent fait l'objet de la section suivante.

2 Étude de l'influence de l'intention sur les invariants du geste

Pour évaluer l'influence de l'intention sur les invariants du geste ciblé, un protocole expérimental est mis en place. La tâche consiste à saisir une microsphère et la déplacer jusqu'à une cible marquée par une croix sur le substrat. Une nouvelle configuration de la position de l'objet et de la cible est tirée aléatoirement à la fin de chaque essai.
La saisie et la dépose sont déclenchées automatiquement lorsque la main est proche de la cible. La vitesse de la main et la distance initiale à la cible sont enregistrées pour chaque

tâche à partir des données fournies par le capteur Kinect. La figure 3.6 montre le protocole expérimental mis en place.

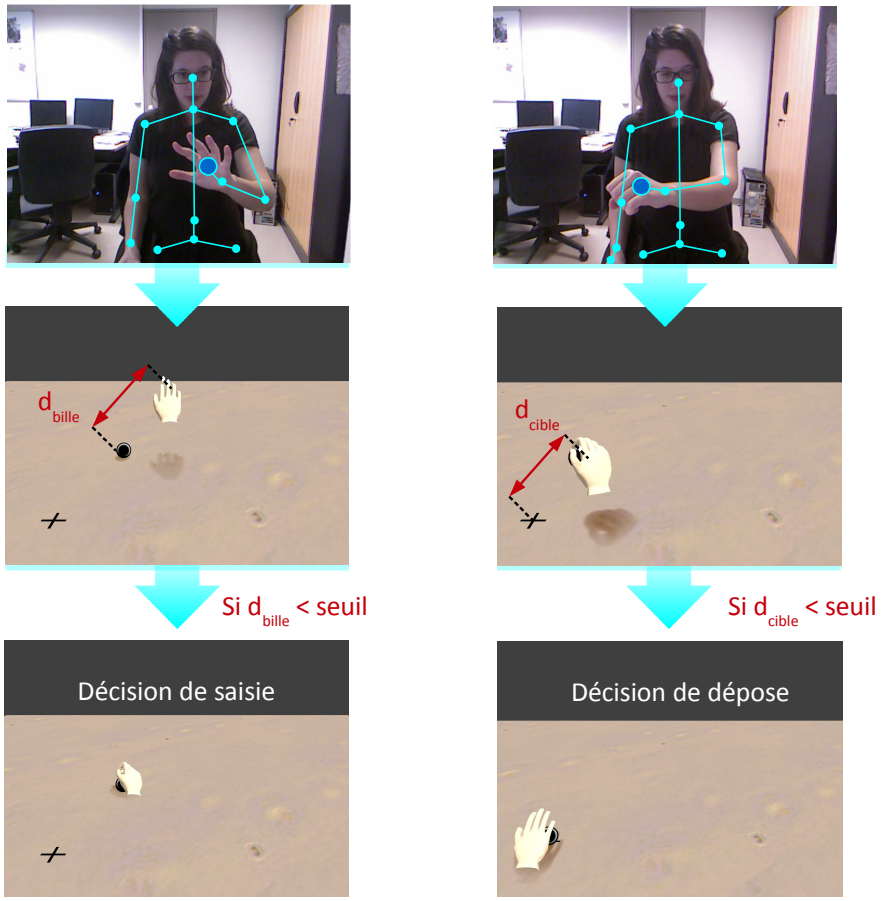


Figure 3.6 - *Le squelette est suivi avec le SDK du capteur Kinect. Lorsque la main est suffisamment proche de la bille, la saisie est déclenchée automatiquement. Une fois la bille saisie, lorsque la main est suffisamment proche de la cible, elle est déposée automatiquement.*

Suivant la méthode d'évaluation proposée dans le chapitre II, aucune information n'est donnée à l'opérateur sur la méthode à utiliser pour saisir et déposer l'objet afin de ne pas contraindre le geste. Le test est réalisé par neuf sujets adultes naïfs. Chaque tâche est répétée 15 fois. Une base de donnée de 270 gestes est ainsi constituée.

2.1 Influence de l'intention sur la cinématique du geste

Pour évaluer l'influence de l'intention sur les invariants du mouvement, un ajustement gaussien est effectué sur la courbe des vitesses de la main pour chaque tâche. L'amplitude, l'écart-type et la moyenne sont ainsi estimés. La figure 3.7 montre un exemple de résultat pour un utilisateur. Les trois paramètres de la gaussienne (amplitude (A), moyenne (EV) et écart-type (SD)) sont tracés en fonction de la distance initiale à la cible. Chaque point

correspond à la réalisation d'une tâche (en noir pour la tâche de saisie et en bleu pour la tâche de dépose). Une interpolation polynomiale est ensuite réalisée sur les points tracés pour chaque paramètre et chaque tâche.

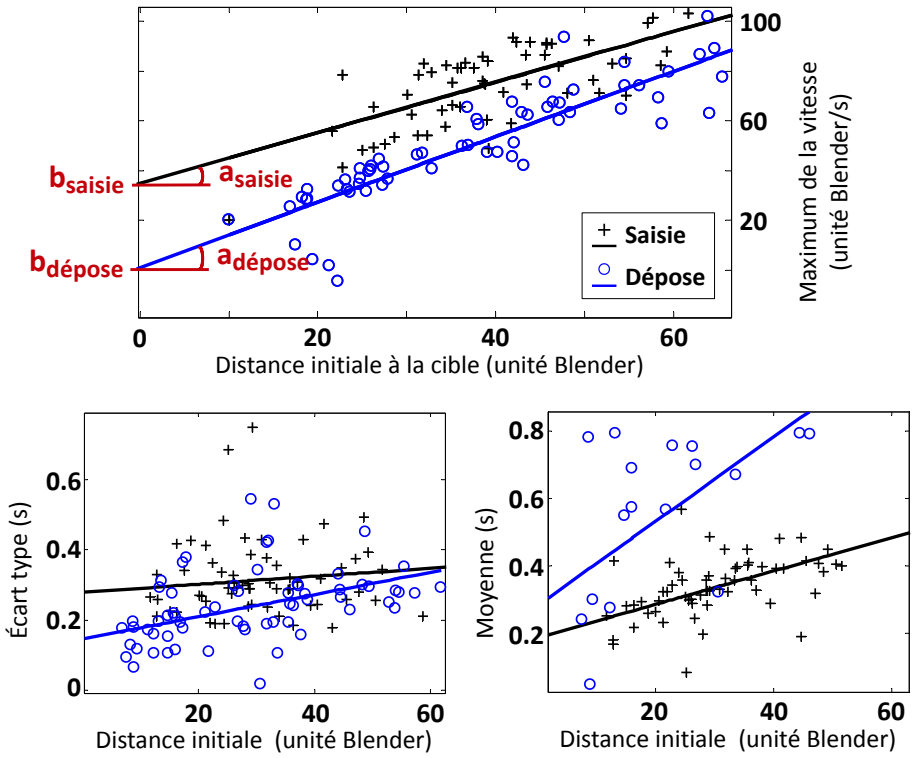


Figure 3.7 - *Influence de la tâche sur les paramètres du profil gaussien des vitesses lors du geste ciblé.*

L'amplitude dépend linéairement de la distance initiale à la cible. Seul le premier ordre de l'interpolation polynomiale est significatif. Ce résultat est cohérent avec la loi d'isochronie du mouvement. L'écart-type ne semble pas dépendre de la distance à la cible. De plus, la moyenne dépend aussi linéairement de la distance initiale à la cible.

Il est aussi intéressant de remarquer que les courbes associées aux tâches de saisie et de dépose sont différentes. Bien que les deux tâches correspondent à des gestes d'atteinte d'une cible, l'amplitude et la moyenne varient selon l'intention de l'opérateur. Un test d'analyse de la variance (ANOVA) est réalisé pour évaluer la signification statistique des résultats. Pour l'amplitude, la valeur-p est inférieure à 0.001%. Cette dernière indique que la différence entre l'amplitude de saisie et celle de la dépose est très significative. La différence entre les écart-types est peu significative.

Selon ces courbes, l'intention semble modifier la cinématique du geste, même pour un même geste d'atteinte d'une cible [Becchio 12] [Sartori 11]. De plus, l'hypothèse formulée précédemment est validée puisque les paramètres de la gaussienne sont différents selon les deux tâches. L'intention est ainsi un paramètre supplémentaire à prendre en compte pour décrire le mouvement ciblé chez l'être humain.

2.2 Influence de l'utilisateur sur la cinématique du geste

Les paramètres obtenus lors de l'ajustement gaussien sont comparés pour les différents utilisateurs. La figure 3.8 montre un exemple de l'influence de deux utilisateurs sur les paramètres de la gaussienne ajustée en fonction de la distance initiale à la cible.

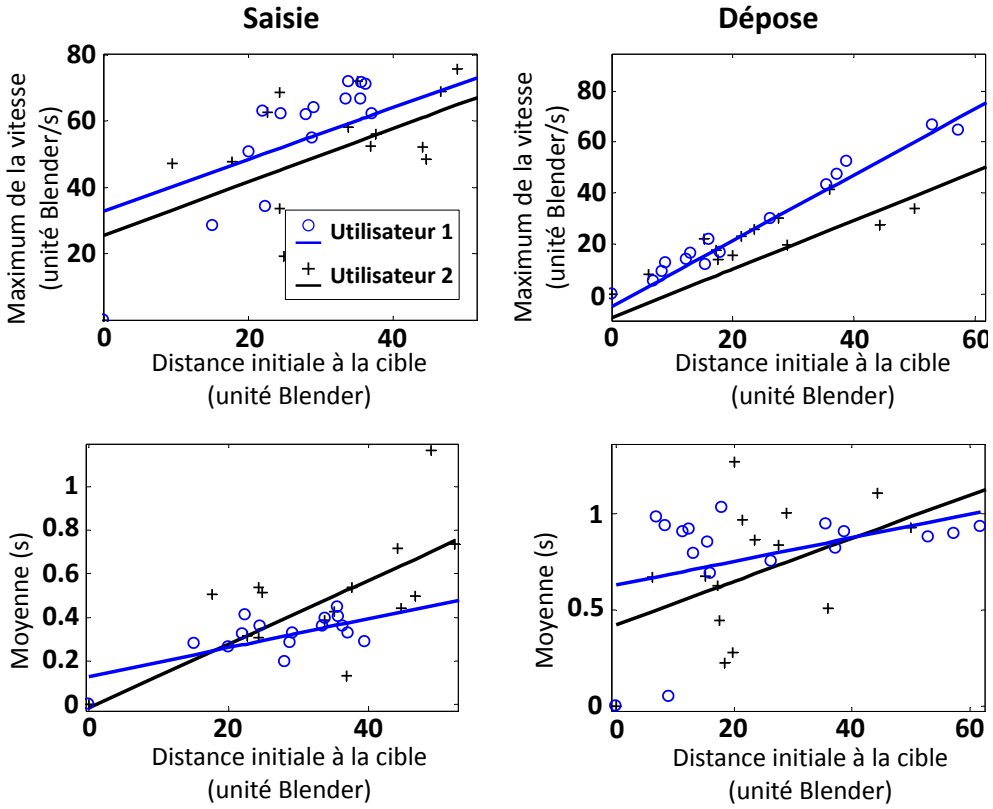


Figure 3.8 - *Exemple de l'influence de deux utilisateurs sur les paramètres des gaussiennes pour les actions de saisie et de dépose de micro-objets virtuels*

Les courbes sont croissantes et linéaires pour tous les utilisateurs. De plus, les amplitudes de la tâche de saisie sont plus grandes que celles de la dépose pour tous les utilisateurs. A l'inverse, la moyenne est inférieure lors de la saisie. Ainsi, les courbes gaussiennes ont des caractéristiques similaires, indépendamment de l'utilisateur. Cette observation montre la généricité du modèle.

2.3 Reconnaissance de l'intention

Les résultats obtenus conduisent à penser qu'il est possible de classifier des gestes ciblés selon l'intention de l'acteur. Dans ce but, un apprentissage par la méthode des k plus proches voisins est réalisé. Le test est effectué par une validation croisée sur la base de données de gestes. Des matrices de confusion sont estimées en faisant varier les caractéristiques utilisées pour la classification : l'amplitude et/ou l'écart-type. Le tableau 3.3

montre les résultats obtenus.

		Maximum		**Écart-type**		**Combinaison**	
		CE		CE		CE	
		Saisie	**Dépose**	**Saisie**	**Dépose**	**Saisie**	**Dépose**
CR	**Saisie**	86.52	13.48	73.03	26.97	82.02	17.98
	Dépose	14.44	85.56	30.00	70.00	16.67	83.33
		Sensibilité : 0.86 Spécificité : 0.86		Sensibilité : 0.71 Spécificité : 0.72		Sensibilité : 0.83 Spécificité : 0.82	

Tableau 3.3 - *Matrices de confusion normalisées (en pourcentage) obtenues par classification (KPPV) à partir des caractéristiques suivantes : l'amplitude (gauche), l'écart type (milieu) et ces deux caractéristiques (droite). La classe estimée est notée CE et la classe réelle CR.*

Il est important de noter que les performances ne sont pas comparables aux résultats obtenus par une reconnaissance de gestes classique. Il s'agit dans ce cas de reconnaître deux intentions qui diffèrent pour un même geste d'atteinte d'une cible. Malgré cette difficulté, les résultats indiquent qu'il est possible de reconnaître l'intention avec de bonnes performances. La sensibilité et la spécificité sont maximisées lorsque seule l'amplitude est utilisée pour la classification. Elle est 15% supérieure au taux obtenu avec les écarts-types. De plus, exploiter les deux caractéristiques n'apporte pas d'amélioration. Ces résultats vont dans le sens des observations statistiques données dans la section 2.1. Celles-ci montrent que seule l'amplitude est statistiquement significative de l'intention.

2.4 Conclusion

La validation expérimentale présentée dans cette partie confirme l'influence de l'intention sur les paramètres de la gaussienne des vitesses lors du geste ciblé. Cette remarque valide l'hypothèse formulée dans la section 1. En particulier, l'amplitude dépend significativement de la tâche.

Les lois invariantes du geste (équations 3.3 et 3.1 page 56) sont ainsi reformulées en prenant en compte ce nouveau paramètre :

$$v_{max}(d_0, tâche) = a_{tâche} \cdot d_0 + b_{tâche} \tag{3.4}$$

avec $a_{tâche}$ et $b_{tâche}$ les paramètres appris à partir de l'interpolation linéaire de l'amplitude en fonction de la distance à la cible, selon la tâche de saisie ou de dépose (fig. 3.7).

$$v(t) = v_{max}(d_0, tâche) \cdot \mathrm{e}^{-\frac{(t - \mu_{tâche}(d_0))^2}{2 \cdot \sigma^2}} \tag{3.5}$$

avec $\mu_{tâche}(d_0)$ la valeur moyenne estimée pour chaque tâche et pour une distance initiale d_0 donnée à partir de l'interpolation linéaire proposée sur la figure 3.7. L'écart-type σ ne dépend pas de la tâche.

Ces indices montrent qu'il est possible de reconnaître l'intention sans passer par des symboles gestuels appris. La section suivante propose un modèle de haut niveau pour prédire l'intention en tenant compte de ces paramètres.

3 Modèle haut niveau de prédiction de l'intention

La section précédente propose une reformulation des lois invariantes du geste ciblé pour inclure l'influence de la tâche. Ces dernières décrivent le comportement naturel de l'être humain lorsqu'il réalise des tâches de manipulation. Cette section propose d'exploiter un modèle de prédiction de l'intention qui repose sur les lois définies. Ce dernier a pour objectif de détecter l'intention de l'opérateur de manière prédictive et non symbolique. Pour construire ce modèle, une piste consiste à s'intéresser aux modèle cognitifs proposés dans la littérature pour prédire l'intention.

3.1 Modèle cognitif de prédiction de l'intention par un observateur humain

Rizzolati et al. montrent que les mêmes aires cérébrales sont activées lors de l'exécution d'une action et lors de son observation [Rizzolatti 96]. Ce système miroir pourrait être au centre des capacités de reconnaissance de l'intention de l'autre [Keysers 06]. Oztop et al. proposent un modèle computationnel qui relie la planification motrice et l'inférence de l'intention [Oztop 05]. Lors de la planification motrice de l'acteur, la conséquence sensorielle de l'action programmée est prédite. Cette prédiction est réalisée par un contrôleur d'anticipation (feedforward). Une fois que cette conséquence sensorielle est observée, une erreur est calculée par le cerveau entre la prédiction et l'observation réelle. Avec cette méthode, l'acteur compense les occlusions et les délais temporels dûs au traitement de l'information sensorielle.

Lors de la prédiction de l'intention, le même modèle est utilisé tout en inhibant la commande motrice. À partir d'une hypothèse d'intention, l'observateur prédit la conséquence sensorielle d'après son propre contrôleur d'anticipation. Ensuite, il évalue l'erreur entre le geste qu'il observe et sa prédiction. Si l'erreur de prédiction est trop grande, une nouvelle hypothèse d'intention est sélectionnée. Dans le cas contraire, l'hypothèse initiale est maintenue. Le modèle d'Oztop est donc prédictif. Il apparaît comme un candidat intéressant

pour réaliser une interface prédictive et non symbolique.

Dans ce travail, l'interface est considérée comme un observateur, et celle-ci doit inférer l'intention de l'acteur/opérateur. Cependant, le modèle original est contrôlé en position. **Il ne permet pas de distinguer un geste ciblé d'un geste aléatoire si ceux-ci ont la même trajectoire.** Cette remarque limite son application au problème posé. Pour dépasser cette limite, ce travail propose de **reformuler ce modèle en exploitant la vitesse de la main**. En effet, la section précédente montre que celle-ci comprend des invariants lors du geste ciblé. La figure 3.9 illustre le formalisme en vitesse des modèles cognitifs de l'acteur et de l'observateur proposé.

La partie suivante s'intéresse à la construction du contrôleur d'anticipation à partir des lois invariantes de la vitesse.

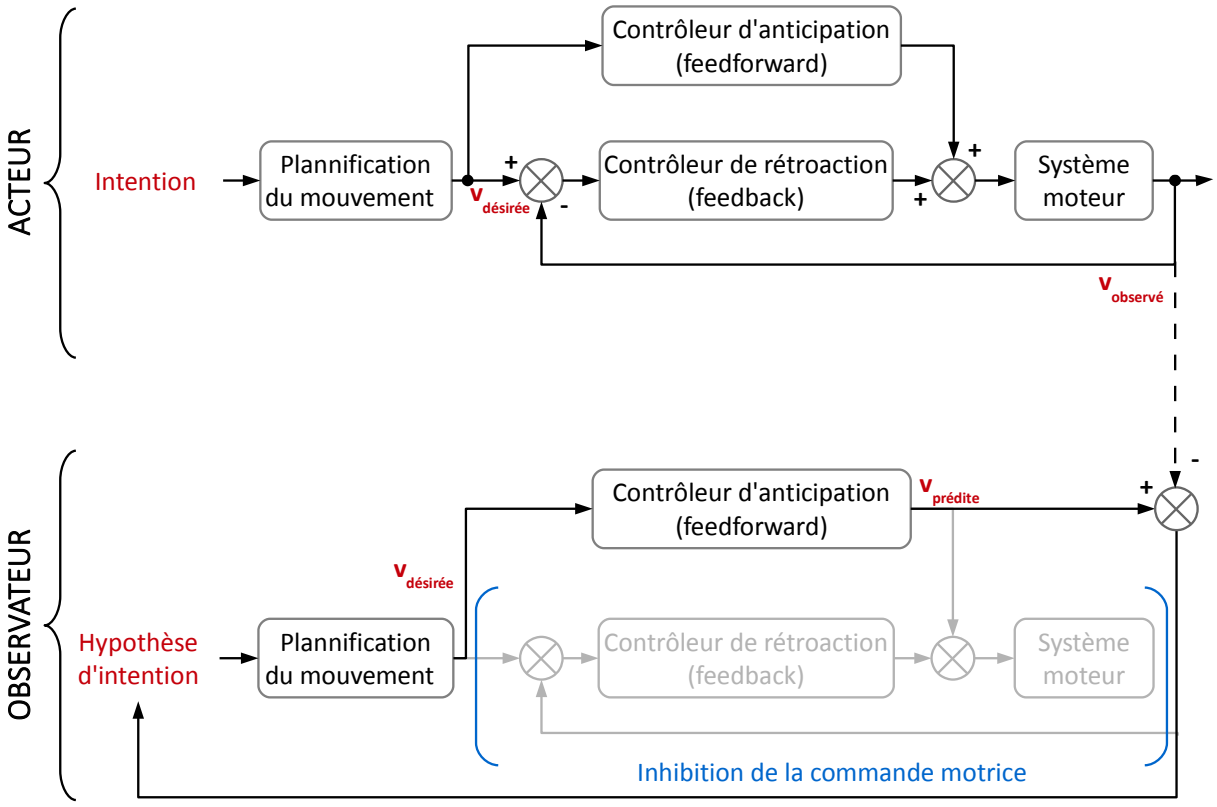


Figure 3.9 - *Modèle cognitif computationnel de prédiction de l'intention*

3.2 Modèle de prédiction de l'intention de saisie/dépose pour la micromanipulation

3.2.1 Influence du contexte de la micromanipulation

Dans le contexte de la téléopération, deux intentions sont considérées : la saisie d'une microsphère, et sa dépose sur un site cible spécifique sur le substrat. Ainsi, l'espace des états est réduit à deux intentions possibles. De plus, l'hypothèse d'intention dépend aussi du contexte (les objets qui sont présents), et des actions possibles. Dans l'interface de micromanipulation, lorsque la main est libre, sa seule possibilité d'action est de saisir un objet. Lorsqu'une microsphère est saisie, la seule intention possible est de la déposer sur le substrat. Ainsi, le système doit prendre en compte ce contexte pour établir une hypothèse d'intention pertinente. Cette observation est cohérente avec l'hypothèse suivante : la prédiction de l'intention par un être humain est plus précise lorsque le contexte est connu [Stapel 12].

3.2.2 Estimation des prédicteurs de la vitesse

Pour mettre en place ce modèle prédictif, la première étape consiste à construire des prédicteurs du mouvement. Ces derniers correspondent au comportement moteur appris par l'observateur. Il s'agit du contrôleur d'anticipation proposé sur la figure 3.9. Pour rendre possible la prédiction, les contrôleurs d'anticipation de l'acteur et de l'observateur doivent être les mêmes. Ce travail propose de les construire à partir des lois invariantes du geste ciblé introduites dans la section précédente. À partir des équations 3.5 et 3.4, un prédicteur de la vitesse de la main est estimé :

$$v_{pred}(t) = v_{max}(d_0, tâche) \cdot \mathrm{e}^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \tag{3.6}$$

où σ et μ sont les moyennes respectives de l'écart-type et de la valeur moyenne (fig. 3.7) estimées sur l'ensemble des échantillons, $v_{max}(d_0, tâche)$ représente le maximum du prédicteur gaussien calculé à partir de l'équation linéaire (3.4) et v_{pred} est le prédicteur de la vitesse.

Les résultats de la section 2 page 59 montrent que l'écart-type ne dépend pas de la distance à la cible. Sa valeur est donc une moyenne sur l'ensemble des échantillons correspondant à chaque tâche.
Un problème central dans le domaine de la reconnaissance de gestes est la segmentation. Il s'agit de déterminer les limites temporelles d'un geste. Dans les approches classiques, cette phase est préalable à la classification. Dans le cadre du modèle proposé, **la segmentation du geste est émergente**. En effet, un geste débute lorsqu'il est suffisamment bien prédit. La classification et la segmentation sont ainsi considérées comme interdépendantes.

De plus, cette remarque montre que le centrage des gaussiennes prédites est lui aussi émergent du modèle. Il n'est pas nécessaire de prendre en compte la moyenne. Celle-ci est simplement calculée par une moyenne sur l'ensemble de la base d'apprentissage.

L'architecture globale du système proposé est montré sur la figure 3.10.

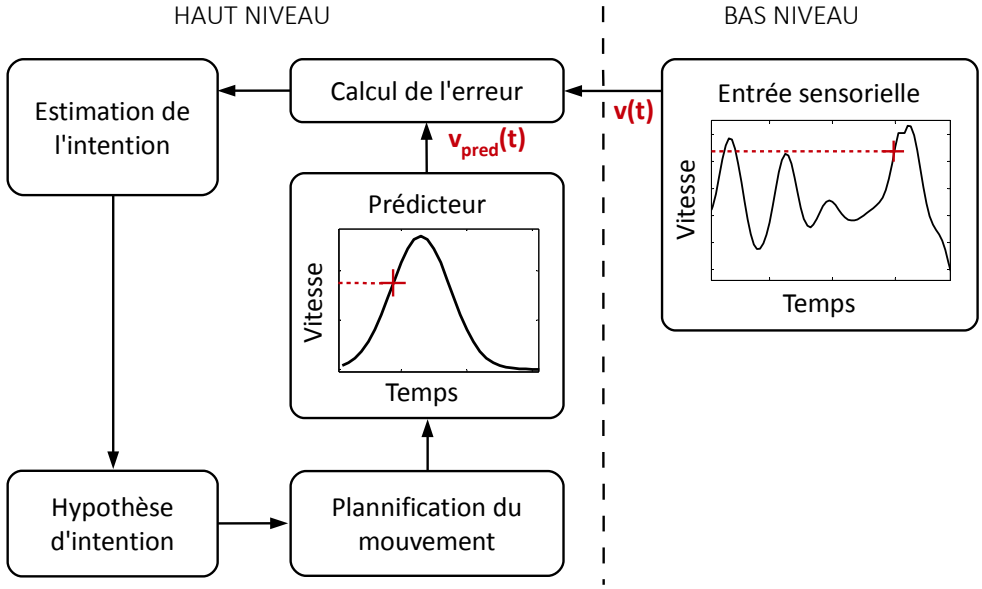


Figure 3.10 - *Modèle de prédiction de l'intention basé sur les invariants de la vitesse de la main lors du geste ciblé. $v_{pred}(t)$ est la vitesse prédite et $v(t)$ la vitesse réelle de la main acquise avec la Kinect.*

3.2.3 Application à la saisie et la dépose d'une microsphère sur un substrat

Lors de la première phase de la tâche, l'utilisateur saisit la microsphère. D'après le modèle proposé, l'hypothèse d'intention est fixée sur l'intention de saisie, et le prédicteur correspondant est calculé. L'algorithme 1 présente la méthode de prédiction proposée pour la tache de saisie. Une hypothèse d'intention est validée si le prédicteur est suffisamment proche de la courbe de vitesses réelle pendant une durée définie. Ce seuil est choisi selon la fréquence d'image du capteur, après que le maximum de la gaussienne est atteint. Une fois que l'intention de saisie est prédite, l'hypothèse d'intention est fixée sur la dépose et les étapes sont répétées.

La figure 3.11 montre un exemple de prédiction de l'intention lors d'une tache de saisie. Les deux premières secondes correspondent à des mouvements aléatoires non ciblés. L'erreur de prédiction est alors supérieure au seuil, et de nombreux nouveaux prédicteurs sont calculés. Après deux secondes, le geste correspond à la phase d'atteinte de la microsphère. L'erreur moyenne entre le prédicteur et la vitesse réelle de la main de l'utilisateur est

Algorithme 1 Prédiction de l'intention de saisie

Extraction de la distance à la cible courante $dist(t)$ à partir du capteur
$v(t) = \frac{d(dist(t))}{dt}$
if $v(t) > 0$ **then**
  $\overline{err} = \frac{1}{N} \cdot \sum_{i=1}^{N}(v_{pred}(N) - v(t))^2$
  if $\overline{err} > errThresh$ **then**
    $v_{pred}(t) = v_{max}(d_0, saisie) \cdot \mathrm{e}\left(-\frac{(t-\mu)^2}{2 \cdot \sigma^2}\right)$
    $N = 1$
  else
    $N = N + 1$
  end if
end if
if $N > Npred$ **then**
  Intention de saisie prédite
end if

où $\overline{err}$ est l'erreur moyenne estimée entre la vitesse observée et le prédicteur, $errThresh$ est le seuil maximal d'erreur acceptée et N_{pred} le nombre de points nécessaires à la reconnaissance de l'intention.

inférieure au seuil, et après nombre de points défini, l'intention de saisie est prédite (en pointillés rouges sur la figure).

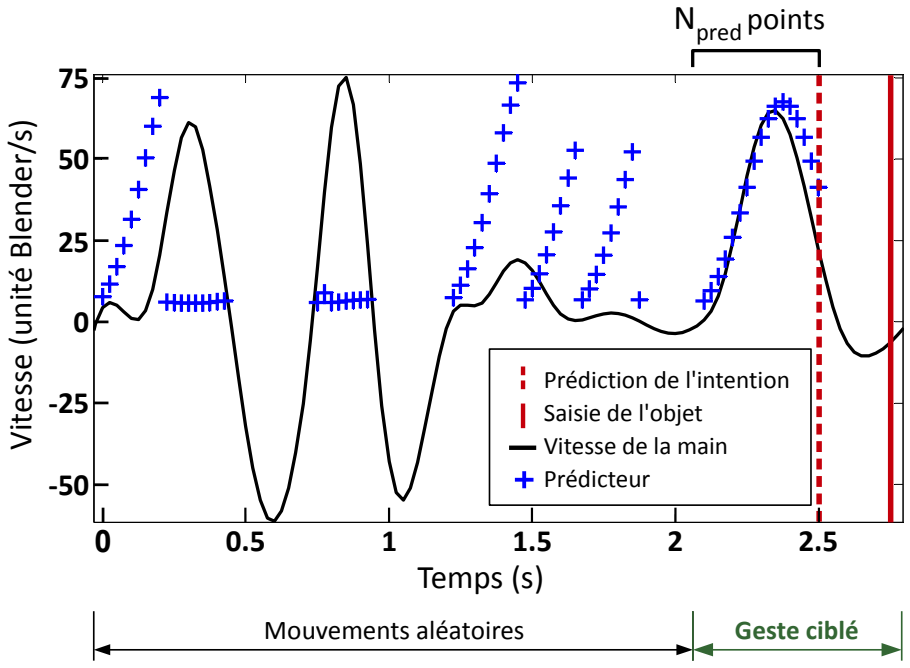


Figure 3.11 - *Exemple d'une séquence de prédiction de l'intention*

Une fois que l'intention de saisie est prédite, l'interface déclenche l'animation correspondante de la main virtuelle afin d'éviter le délai observé avec la méthode classique de reconnaissance de gestes.

3.3 Conclusion

Dans cette section, un modèle de prédiction de l'intention est proposé à partir des caractéristiques cinématiques de bas-niveau sélectionnées dans la section précédente. Cette approche cherche à résoudre les problèmes relevés dans le chapitre II :

- éviter l'apprentissage d'un dictionnaire de gestes symboliques (sémiotiques) par l'utilisateur sans contraindre ses mouvements et en déterminant des caractéristiques invariantes du geste ciblé (ergodique) chez l'être humain.
- construire un système prédictif récurrent qui dispose d'une causalité circulaire en s'inpirant du fonctionnement de l'être humain, expert pour prédire l'intention de l'autre.

Pour valider expérimentalement l'apport de cette approche, la section suivante présente le protocole mis en place ainsi que les résultats obtenus en suivant la méthode d'évaluation présentée dans le chapitre 2.

4 Évaluation de l'interface pour la micromanipulation

4.1 Protocole du test utilisateur

Un protocole de test utilisateur est mis en place pour évaluer le système interactif proposé et pour le comparer à l'approche classique par reconnaissance de gestes. La tâche consiste à saisir une microsphère et la déplacer jusqu'à une cible marquée par une croix sur le substrat. Une nouvelle configuration de la microsphère et de la cible est tirée aléatoirement à la fin de chaque essai. Pour valider l'aspect naturel de l'interaction, aucune information n'est fournie à l'opérateur sur la méthode à utiliser pour réaliser la tâche.

Le protocole est conduit en trois phases. La première est un cas témoin pour lequel la saisie et la dépose sont déclenchées simplement par un critère de proximité entre la main et la cible. Dans la seconde phase, la méthode classique de reconnaissance de gestes décrite dans le chapitre II est utilisée. La troisième phase exploite l'approche de prédiction de l'intention proposée dans ce travail. Le protocole est montré sur la figure 3.12 pour la prédiction de l'intention.

L'ordre des phases présentées à chaque utilisateur est modifié de manière aléatoire afin d'éviter un effet d'apprentissage dans les résultats. Chaque phase est constituée de

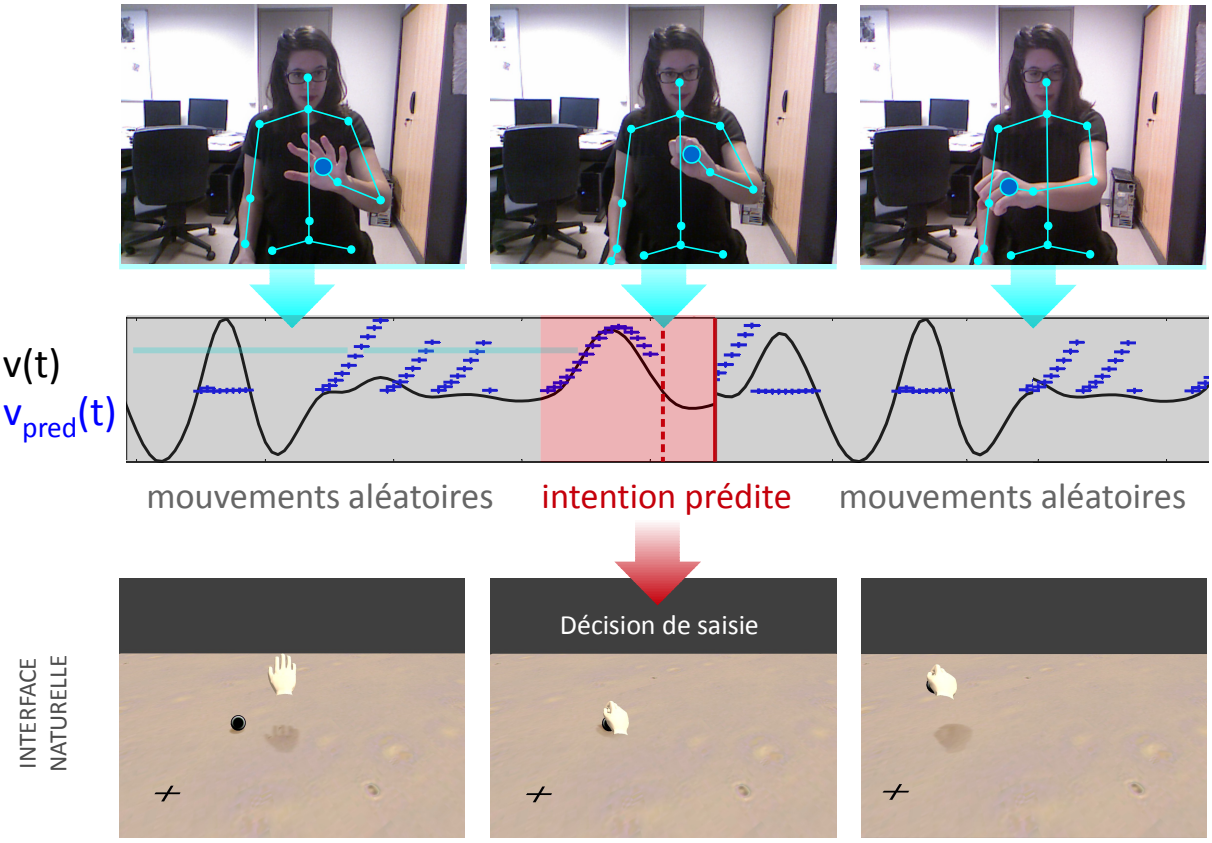


Figure 3.12 - *Expérience utilisateur d'évaluation de l'interface basée sur la prédiction de l'intention.*

15 essais de saisie et dépose. A la fin de chaque phase, un questionnaire est soumis à l'utilisateur pour évaluer de manière qualitative l'approche d'une telle interface sur le confort et la facilité d'utilisation. Ce questionnaire est basé sur le System Usability Scale (SUS) [Brooke 96]. Neuf sujets naïfs adultes ont participé à l'expérience.

4.2 Résultats comparatifs de la reconnaissance de gestes et de la prédiction de l'intention

4.2.1 Résultats quantitatifs

Le succès de la tache est évalué sur un critère basé sur la proximité et la durée : si la main reste proche de la cible pendant plus de $0.5s$ sans que la saisie/dépose soit détectée, la tâche est considérée comme un échec. Le pourcentage de réussite pour chaque méthode est montré sur la figure 3.13 à gauche. La méthode de prédiction de l'intention proposée montre une amélioration significative de 33.8% par rapport à l'approche classique pour la tâche de dépose.

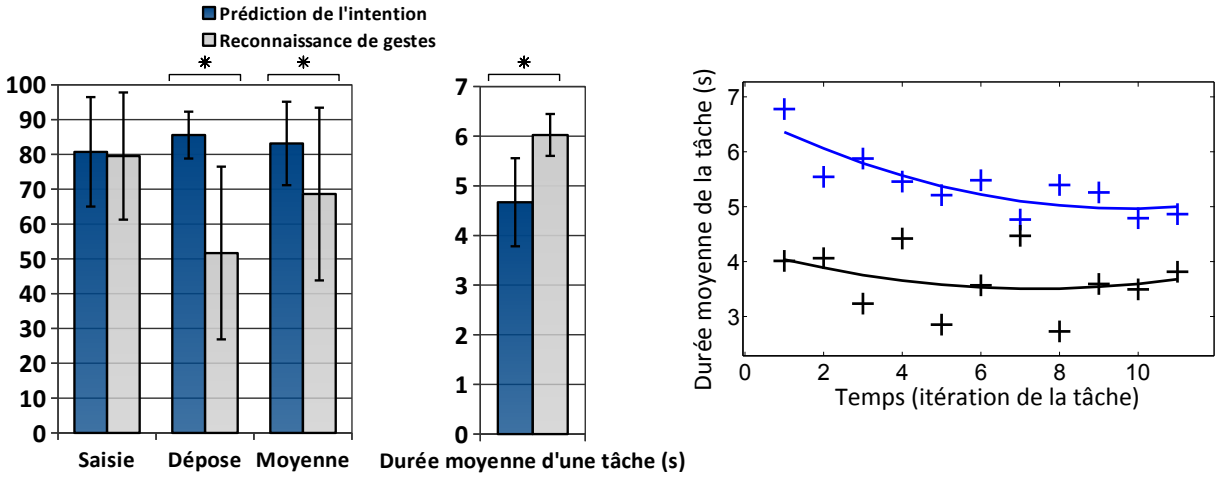


Figure 3.13 - *Pourcentage de succès des deux tâches avec la méthode de prédiction de l'intention et la reconnaissance de gestes (à gauche) et durée moyenne d'une tâche (au milieu). Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA. La courbe de droite montre l'évolution de la durée moyenne de l'ensemble des utilisateurs pour réaliser une tâche en fonction du temps. La méthode de prédiction de l'intention est représentée en noir et la reconnaissance de gestes en bleu*

Un autre critère d'évaluation est la durée moyenne d'un essai. La durée moyenne avec la méthode de prédiction de l'intention est de $4.7s$, alors que l'approche classique nécessite $6s$ en moyenne pour réaliser la tâche (fig. 3.13). L'approche proposée améliore significativement la durée de la tâche de manipulation.

Une indication de l'effet d'apprentissage consiste à observer l'évolution de la durée nécessaire pour accomplir une tâche lors des répétitions de celle-ci. Les résultats obtenus sont illustrés par la figure 3.13 droite. Après une interpolation polynomiale d'ordre 3, la méthode par prédiction de l'intention réalise la tâche en un temps inférieur à la reconnaissance de gestes. Cette remarque est vraie pour l'ensemble des expériences. De plus, la pente de la courbe de la reconnaissance de gestes est plus importante que celle de la prédiction de l'intention. Cette pente constitue un indicateur de l'effet d'apprentissage au cours des répétitions sur la durée de la tâche. L'approche par prédiction de l'intention a une pente faible, elle est donc maîtrisée dès la première itération. Cette propriété semble indiquer que l'interface proposée est naturelle selon la définition proposée dans le chapitre 1.

4.2.2 Résultats qualitatifs

Le questionnaire utilisateur SUS montre une préférence de 30.8% de plus pour l'approche proposée par rapport à l'approche classique (fig. 3.14). En particulier, l'évaluation de la facilité d'utilisation montre de meilleurs résultats pour la prédiction de l'intention

par rapport à la reconnaissance de geste et au cas-témoin. L'évaluation du confort montre aussi une préférence significative pour l'approche proposée.

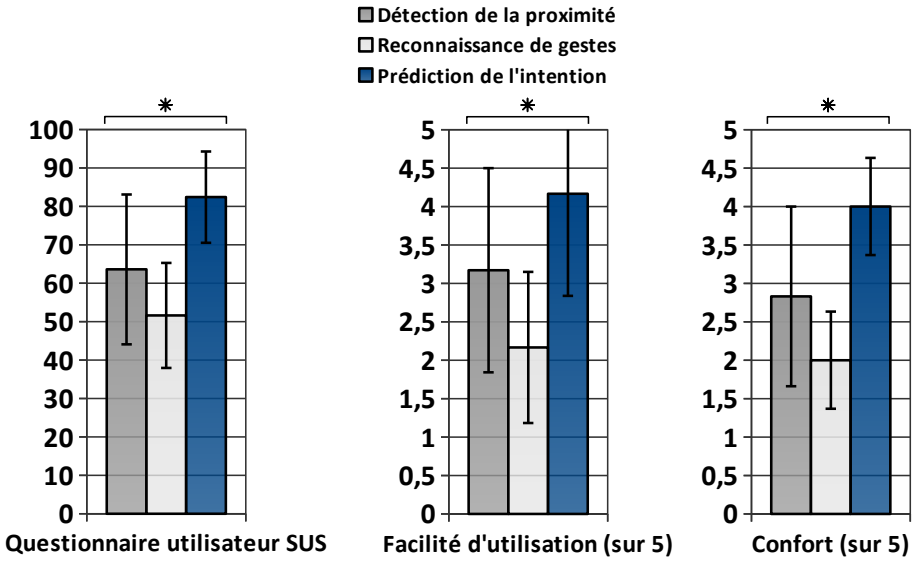


Figure 3.14 - *Résultats du questionnaire utilisateur SUS. Une astérisque (*) indique une valeur p inférieure à 0.05 ($p < 0.05$). L'analyse statistique utilisée est un test ANOVA.*

Les résultats montrent l'amélioration obtenue avec la méthode proposée dans un contexte d'interaction naturelle sans consigne donnée à l'utilisateur, à la fois quantitativement et qualitativement. L'approche de prédiction de l'intention semble validée comme une solution intéressante pour fournir une assistance à l'opérateur pour la micromanipulation d'objets sous AFM. Cependant, certains utilisateurs mentionnent une contrainte : la détection semble parfois trop prédictive. Cette remarque peut nuire à l'impression de contrôler la tâche. Il pourrait être intéressant d'étudier de manière précise à quel point de la courbe la saisie/dépose doit être déclenché pour améliorer cet aspect.

5 Conclusion

Le chapitre II montre que la reconnaissance des gestes main ouverte/fermée est insuffisante pour créer une interface non symbolique. L'utilisateur doit toujours apprendre un dictionnaire de gestes pour interagir. Pour créer une interface naturelle non symbolique, ce chapitre propose une nouvelle approche de prédiction de l'intention. Il est montré qu'il est possible de reconnaître l'intention à partir de la cinématique du geste ciblé. Cette observation conduit à une reformulation de la loi d'isochronie du mouvement et du profil gaussien de la vitesse pour intégrer ce paramètre. Un modèle haut niveau récurrent est construit en s'inspirant de la prédiction de l'intention par un être humain. Un modèle cognitif prédictif est adopté [Oztop 05]. Ce dernier propose un contrôle en position. Pour distinguer un geste ciblé d'un geste aléatoire lorsque ceux-ci ont la même trajectoire, ce travail propose de reformuler ce modèle en vitesse. Il repose sur la construction

d'un prédicteur à partir des invariants en vitesse du geste ciblé. Sans consigne donnée à l'utilisateur, les performances obtenues sont significativement améliorées en termes de succès et de durée de la tâche. De plus, le questionnaire utilisateur SUS montre une préférence significative des utilisateurs par rapport à l'approche classique par reconnaissance de gestes. L'approche basée comportement apparaît comme une solution intéressante pour une interaction plus naturelle pour l'opérateur. Il est important de noter que l'approche proposée est indépendante de la pose de la main et des doigts, puisqu'il suffit d'évaluer la vitesse de la main. Ainsi, cette approche est **généralisable à de nombreux autres types d'interfaces**, par exemple à une souris ou un bras haptique.

L'ensemble de ce chapitre considère que la cible du geste est connue a priori. La méthode proposée est donc limitée à des contextes où elle peut être déterminée facilement. Dans des contextes plus complexes, il est nécessaire de réaliser une étape préalable de sélection de la cible. D'un point de vue cognitif, la sélection est réalisée par les mécanismes attentionnels. Le chapitre suivant a pour objectif de modéliser ces mécanismes pour déterminer la cible de l'opérateur de manière prédictive.

Chapitre 4

Estimation du focus d'attention pour des scènes multicibles

Sommaire

1 Sélection des signaux de bas niveau pour modéliser l'intention 50

1.1 Définition fonctionnelle de l'intention 50

1.2 État de l'art des signaux caractéristiques de l'intention 51

1.3 Les signaux invariants du geste ciblé humain 55

2 Étude de l'influence de l'intention sur les invariants du geste 57

2.1 Influence de l'intention sur la cinématique du geste 58

2.2 Influence de l'utilisateur sur la cinématique du geste 60

2.3 Reconnaissance de l'intention 60

2.4 Conclusion 61

3 Modèle haut niveau de prédiction de l'intention 62

3.1 Modèle cognitif de prédiction de l'intention par un observateur humain 62

3.2 Modèle de prédiction de l'intention de saisie/dépose pour la micromanipulation 64

3.3 Conclusion 67

4 Évaluation de l'interface pour la micromanipulation 67

4.1 Protocole du test utilisateur 67

4.2 Résultats comparatifs de la reconnaissance de gestes et de la prédiction de l'intention 68

5 Conclusion 70

Le chapitre précédent propose une approche non symbolique basée sur le comportement de l'opérateur pour reconnaître ses décisions. L'objet cible est considéré comme connu dans son environnement. Cependant, certaines tâches de manipulations impliquent des scènes plus complexes dans lesquelles il est impossible de déterminer a priori la cible de l'opérateur. Une stratégie de sélection sans a priori apparaît comme une solution prometteuse pour se rapprocher d'une scène réelle de micromanipulation.

L'objectif de ce chapitre est la proposition d'une méthode pour réaliser cette sélection dans le cadre d'une interface naturelle. L'enjeu consiste à déterminer la cible visée sans qu'elle soit formulée de manière explicite. Une solution consiste à analyser le comportement naturel de l'opérateur lors de tâches de sélection. D'un point de vue cognitif, la sélection est effectuée par les mécanismes attentionnels. L'être humain est capable d'inférer le focus d'attention de l'autre pour réaliser des tâches collaboratives qui impliquent une attention conjointe. La modélisation de cette capacité constitue une piste pour déterminer le focus d'attention à partir d'indices comportementaux.

Pour construire ce modèle de haut niveau, la première partie de ce chapitre s'attache à sélectionner les signaux de bas niveau caractéristiques du focus d'attention. À partir des signaux extraits, un modèle d'estimation du focus d'attention est proposé dans la section 3. Ce modèle inclut les contraintes liées aux tâches disponibles et à la configuration de la scène de micromanipulation. Une évaluation de la solution proposée est ensuite présentée dans la section 4. Elle montre l'influence des indices du comportement de l'utilisateur et de la méthode d'activation du modèle sur l'estimation du focus d'attention sur une grille d'objet.

1 Les mécanismes attentionnels

1.1 Définitions

1.1.1 L'attention sélective et le focus d'attention

Le cerveau humain traite un flux de données très important chaque seconde [Koch 06]. Cette propriété est rendue possible par la mise en place de mécanismes qui réduisent la quantité des données en filtrant les informations non pertinentes. Ce processus de sélection est appelé **attention sélective**. Duncan montre que l'être humain fixe son attention sur un seul objet à la fois [Duncan 84]. L'auteur distingue deux étapes. La première est un processus pré-attentionnel de bas niveau, effectué de manière parallèle sur l'ensemble de la scène. La seconde étape est réalisée de manière sérielle sur un objet unique. L'attention est centrée sur cet objet. Sélectionner une cible consiste ainsi à estimer la zone où se concentre l'attention. Cette définition trace le cadre du **focus d'attention**. Ce dernier dépend de propriétés de la scène et du contexte, généralisées sous le terme de saillance.

1.1.2 La notion de saillance

La saillance caractérise les zones de la scène qui se distinguent de leur voisinage pour un observateur. Celle-ci peut être de nature visuelle, auditive ou encore tactile. Pour étudier l'attention de l'opérateur dans une scène de réalité virtuelle, ce travail se concentre sur la saillance visuelle. À partir d'une scène visuelle, statique ou dynamique, plusieurs travaux de la littérature proposent de créer des **cartes de saillance**. Il s'agit d'une image 2D sur laquelle l'intensité de chaque pixel encode son caractère saillant. Ces cartes ont généralement pour but de prédire les saccades visuelles d'un observateur. Le regard est considéré dans la plupart des travaux comme une métrique idéale du focus d'attention visuelle.

Différentes approches sont proposées dans la littérature pour expliquer les saccades visuelles [Borji 13]. Deux types de mécanismes sont principalement envisagés. La saillance orientée "bottom-up" correspond aux traitements réflexes de bas niveau d'une image. La seconde approche est orientée "top-down". Elle est volontaire et dépendante de la tâche.

1.1.2.1 La saillance orientée "bottom-up"

Cette approche repose principalement sur l'analyse des **stimulus visuels d'une scène** donnée. Elle consiste à chercher les zones d'une image qui attirent l'attention en détectant leur saillance visuelle. Par exemple, dans une scène où un seul trait horizontal est présent au milieu de traits verticaux, l'attention est immédiatement centrée sur celle-ci [Treisman 80]. Un exemple de cartes de saillance établies à partir de la couleur et de l'orientation est donné sur la figure 4.1. Cependant, ces modèles ne permettent de prédire qu'une partie des saccades visuelles. En effet, il est montré que la plupart des zones de fixations dépendent surtout de la tâche que l'on cherche à accomplir [Henderson 99].

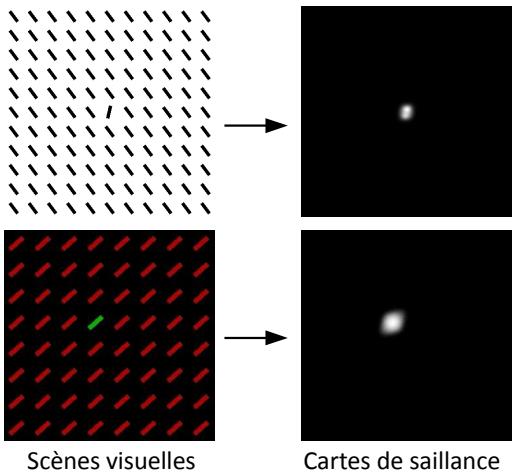


Figure 4.1 - *Exemples de cartes de saillance bottom-up établies à partir de deux types de caractéristiques : l'orientation (en haut) et la couleur (en bas) [Gao 07]*

1.1.2.2 La saillance orientée "top-down"

Contrairement au contrôle bottom-up, cet aspect de l'attention est **orienté tâche**. Il s'agit d'un processus volontaire [Itti 01]. Un exemple classique de l'attention top-down est donné par Yarbus et al. [Yarbus 67]. Les auteurs montrent expérimentalement que les mouvements des yeux dépendent de la tâche en posant différentes questions à des observateurs face à une même scène (fig. 4.2 a). Selon la tâche demandée, par exemple estimer l'âge des personnes présentes (fig. 4.2 b) ou se souvenir de leurs vêtements (fig. 4.2 c), les saccades visuelles diffèrent fortement.

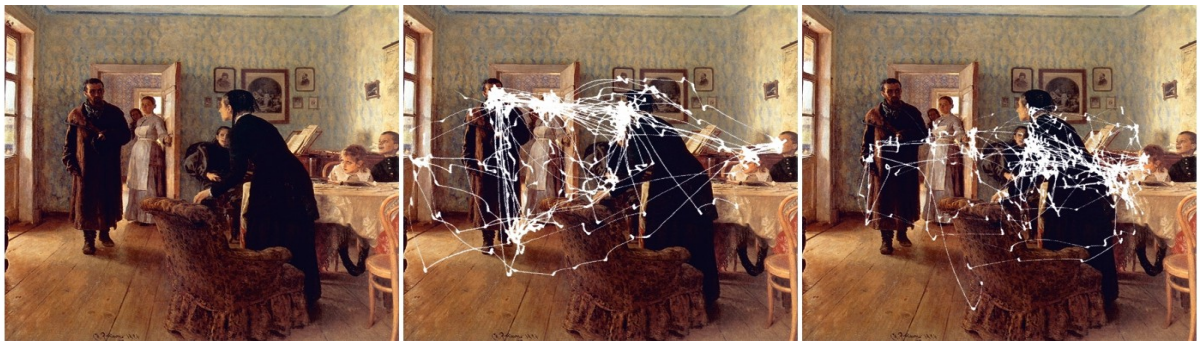


Figure 4.2 - *Les saccades visuelles sur une même scène dépendent de la tâche. Des saccades différentes sont observées selon la question posée à l'observateur [Yarbus 67].*

Dans le simulateur proposé dans ce travail, les objets présents sont connus et les tâches disponibles sont identifiées. Dans ce cadre contrôlé, estimer les objets saillants selon la tâche ne présente pas de difficulté. Cependant, les approches "bottom-up" et "top-down" n'offrent pas de solution pour discriminer la cible de l'utilisateur parmi ces objets.

1.2 Approche proposée

Les approches "bottom-up" ont pour objectif d'extraire des objets saillants dans une scène visuelle. Elles sont basées sur des caractéristiques telles que l'orientation et la couleur. L'estimation de l'attention "top-down" inclut la notion de tâche à réaliser. Ces travaux considèrent ainsi que le focus d'attention ne dépend que de la scène et de la tâche. Pour une scène et une tâche données, il est considéré comme statique. Ces approches semblent limitées pour déterminer le focus d'attention de l'opérateur de manière dynamique lors de son interaction avec l'interface. Pour dépasser cette limite, une solution consiste à exploiter le **comportement de l'opérateur** pour déterminer sa cible.

L'être humain s'appuie sur le comportement de l'autre pour déterminer son focus d'attention lors de tâches collaboratives. Cette capacité est appelée l'**attention conjointe**. Des modèles de cette dernière sont proposés en sciences cognitive et en robotique hu-

manoïde pour réaliser des tâches de coopération homme-robot. Cette solution apparaît comme une piste intéressante pour résoudre le problème posé.

Pour estimer le focus d'attention de l'opérateur sur l'ensemble de la scène de manière continue, ce travail propose de construire une **carte de saillance basée comportement**. Cette dernière est construite à partir d'informations sur la scène (attention "bottom-up"), de la tâche (attention "top-down") et d'indices de bas niveau du comportement de l'opérateur. Le point le plus saillant correspond au focus d'attention dans la scène. La figure 4.3 synthétise le modèle complet développé dans ce chapitre.

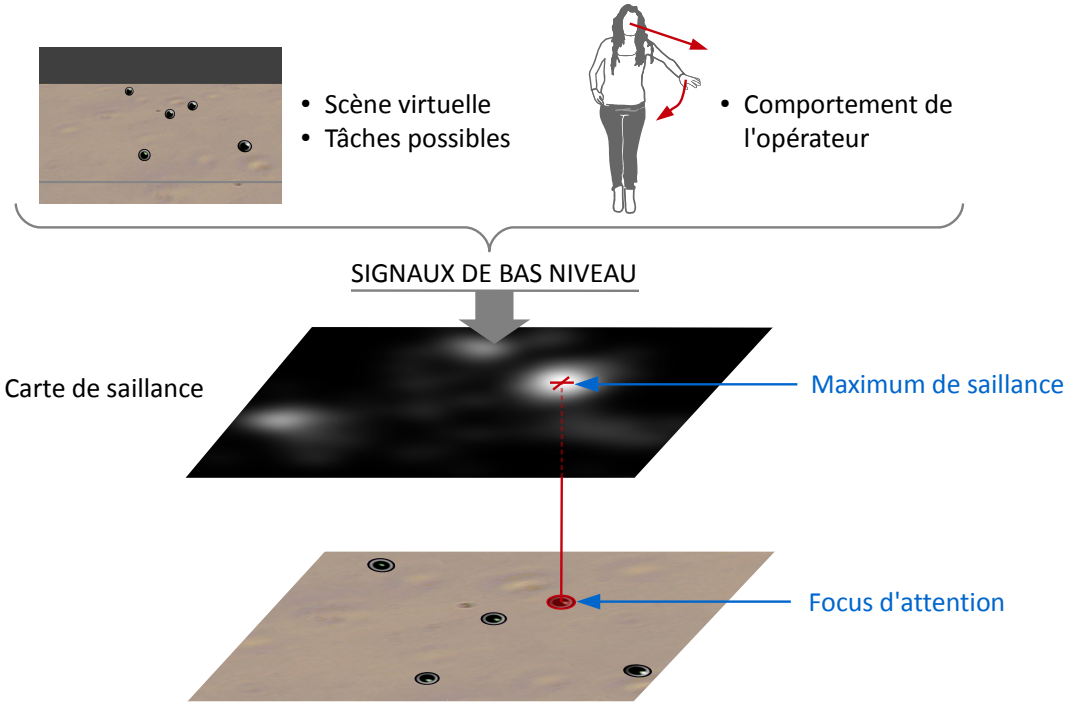


Figure 4.3 - *Modèle du focus d'attention basé sur le comportement de l'opérateur proposé.*

Pour construire ce modèle, la première étape consiste à déterminer les signaux comportementaux caractéristiques du focus d'attention. Le choix de ces derniers est contraint par les informations extractibles à partir des capteurs. En particulier, l'acquisition doit être réalisée sans nuire au naturel de l'interaction.

2 Les indices bas niveau du focus d'attention

2.1 Le regard et la pose du visage

Différentes études désignent le regard comme indice essentiel de l'attention conjointe [Frischen 07]. Nummenmaa et al. synthétisent des données d'imagerie cérébrale et des études

de lésions qui montrent que l'être humain détecte l'attention de l'autre à partir de la **direction du regard, de la tête et du corps** [Nummenmaa 09]. D'autres travaux montrent que **la direction de la tête est suffisante pour déterminer le focus d'attention** [Nuku 08]. Dans le protocole expérimental proposé, les sujets portent des lunettes noires. Sans voir la position du regard, les observateurs suivent spontanément leur focus d'attention. La pose de la tête apparaît comme un indice important de l'attention. Dans le cadre d'une tâche de micromanipulation, la pose de la tête constitue ainsi un signal de bas niveau pertinent pour discriminer la cible de l'opérateur. Ce signal est adopté dans ce travail.

2.2 Le geste

Lors de leur première année, les enfants observent principalement le visage de l'adulte avec lequel ils interagissent. Dans la deuxième moitié de la première année, leur attention est déplacée vers la main de l'adulte et l'objet manipulé. Amano et al. décrivent cette transformation comme un précurseur des mécanismes de l'attention conjointe [Amano 04]. Le geste semble ainsi être un autre indice important du focus d'attention. Langton et al. étudient l'influence de signaux visuels contradictoires (le geste et la direction de la tête) présentés à des observateurs [Langton 00]. Dans l'expérience proposée, une image d'un homme qui réalise un geste de pointage vers le haut ou vers le bas est montrée au sujet. La tête de l'acteur est orientée soit dans la direction du geste soit dans la direction opposée. La consigne consiste à ignorer la pose de la tête pour un groupe, et la direction du geste pour l'autre groupe. Les réponses des sujets sont notées par des clics rapides sur des boutons "haut" et "bas". Les réponses des utilisateurs sont plus affectées par l'incohérence du geste que par celle de la tête. De plus, la réponse des sujets au signal gestuel est plus rapide. Ces indicent montrent que le geste est un signal essentiel du focus d'attention de l'acteur. Le geste semble être privilégié par les observateurs par rapport à la pose du visage lorsque ces deux signaux sont présents. L'expérience proposée s'intéresse au geste de pointage. Il s'agit donc d'un geste dédié à la communication symbolique. Ce travail propose d'exploiter le geste dans un autre contexte : la manipulation d'objets. Dans ce contexte, **la position de la main à un instant donné** et **la direction du geste** sont deux indices bas niveau qui informent un observateur de la cible visée. Contrairement au geste de pointage, ces signaux ne sont pas symboliques. Ils constituent une piste intéressante pour déterminer la cible de l'opérateur sans nécessiter l'apprentissage de symboles. Ces indices sont donc retenus dans le cadre de ce travail.

2.3 Choix des signaux pour notre interface

Deux signaux principaux sont relevés dans la partie précédente : le geste et le regard. Suivre le regard de manière précise nécessite un eye tracker. Ces derniers impliquent de contraindre les mouvements de l'utilisateur, ce qui peut nuire au naturel de l'interaction. De plus, la pose de la tête semble être un indice suffisant à un observateur humain pour

déterminer le focus d'attention. Dans ce travail, le regard est ainsi approximé par la **pose de la tête**.

La détection de celle-ci peut être réalisée avec des méthodes de vision par ordinateur. Cependant, les erreurs de détection sont supérieures à 8^o avec les méthodes actuelles de la littérature [Fanelli 11]. Face à un écran de taille classique, les déplacements de la tête sont de faible amplitude. La précision des méthodes de vision par ordinateur semble donc insuffisante. Une solution consiste à exploiter un dispositif dédié tel que le TrackIR [1]. Ce dernier nécessite de porter un réflecteur qui peut être posé sur un chapeau. Il permet d'extraire trois angles qui définissent la pose de la tête : le tangage, le roulis et le lacet (fig. 4.4), ainsi que la position de la tête en 3D. Le roulis est négligé. Le point regardé par l'utilisateur sur l'écran est estimé à partir des angles de tangage et de lacet et de la position 3D de la tête :

$$x_{regard} = x_{tête} + z_{tête} * tan(\alpha_{lacet}) \tag{4.1}$$

$$y_{regard} = y_{tête} + z_{tête} * tan(\alpha_{tangage}) \tag{4.2}$$

avec (x_{regard}, y_{regard}) la position du point regardé sur l'écran, $(x_{tête}, y_{tête})$ la projection orthogonale de la position de la tête sur l'écran, $z_{tête}$ la distance entre la tête et l'écran, $\alpha_{tangage}$ l'angle de tangage, α_{lacet} l'angle de lacet.

L'autre indice essentiel du focus d'attention relevé est le **geste**. Pour suivre celui-ci, la position de la main est détectée à partir du squelette Kinect (voir chapitre 2). La position 3D extraite est projetée orthogonalement sur l'écran. Un curseur est affiché à cette position (fig. 4.4).

À partir des signaux extraits (point regardé et curseur de position de la main), un modèle haut niveau du focus d'attention est proposé dans la partie suivante.

3 Modélisation haut niveau du focus d'attention

La section précédente relève le geste et la pose de la tête comme signaux comportementaux caractéristiques du focus d'attention. L'attention est une notion de haut niveau. Elle nécessite l'intégration et l'interprétation de ces signaux de bas niveau pour déterminer la cible de l'opérateur. La méthode proposée pour réaliser cette intégration doit répondre aux contraintes liées à l'interface. En particulier, elle doit prendre en compte la configuration de la scène de manipulation et le contexte. Cette partie s'attache à spécifier ces contraintes pour proposer une méthode adaptée à la sélection de cibles dans une scène de micromanipulation.

1. http ://www.trackir.fr

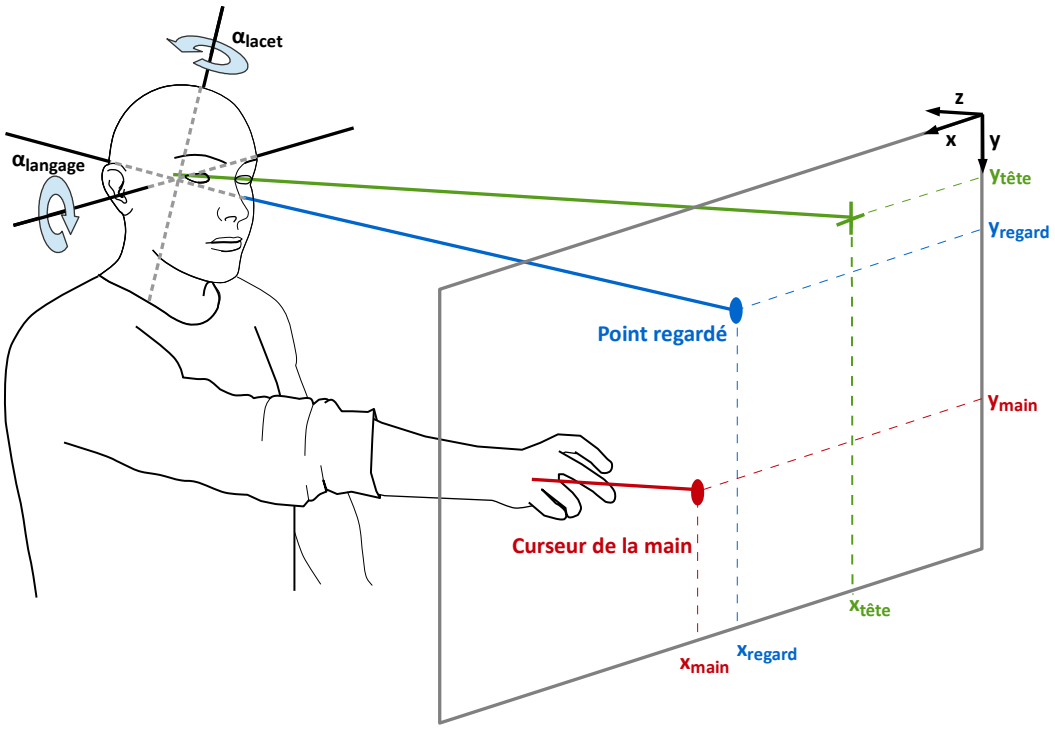


Figure 4.4 - *La pose de la tête*

3.1 Contraintes et objectifs

3.1.1 Contraintes liées à la scène et au contexte

Dans le cadre de l'interface proposée, deux cas sont distingués selon la tâche. Le focus d'attention peut porter soit sur un objet discret donné, soit sur une région quelconque sur le substrat. Ces phases sont présentées sur la figure 4.5.

- **cas discret** : lors de la phase de saisie, plusieurs sphères sont présentes dans la scène. Dans ce cas, déterminer le focus d'attention de l'opérateur consiste à sélectionner la sphère cible que l'opérateur souhaite saisir. Seul un nombre fini d'objets discrets sont donc éligibles.
- **cas continu** : lorsqu'une microsphère est saisie, l'utilisateur doit pouvoir la déposer sur l'ensemble de la surface du substrat. Cette fois, le focus d'attention n'est pas contraint par un nombre fini d'objets. Il s'agit donc de déterminer le focus d'attention sur une surface continue.

La solution choisie dans ce travail pour détecter le focus d'attention doit être capable d'appréhender ces deux possibilités de l'opération de micromanipulation.

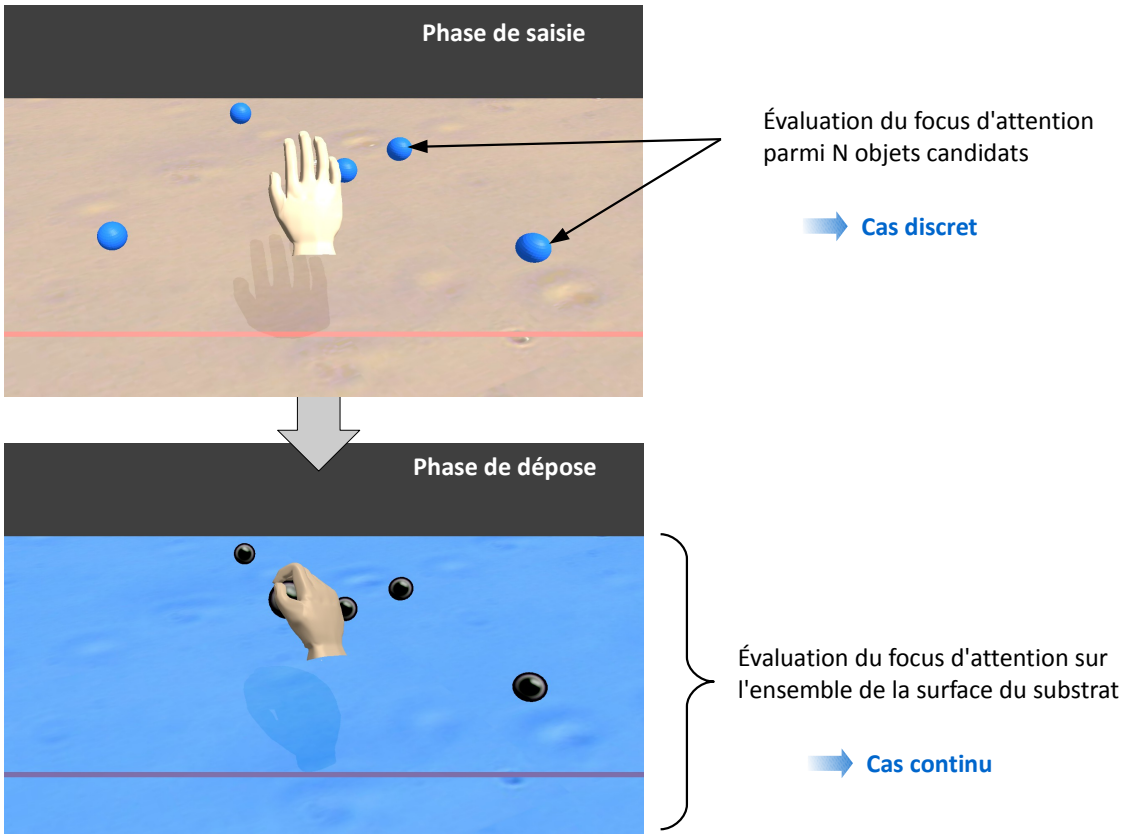


Figure 4.5 - *Cas discret et cas continu du focus d'attention selon la tâche.*

3.1.2 Contrainte de prédictivité

Le chapite précédent montre l'apport d'un système prédictif de l'intention pour déterminer la tâche. Ce principe est appliqué ici pour sélectionner la cible de manière naturelle. Une solution prometteuse consiste à proposer un **modèle prédictif du focus d'attention**. En particulier, la détection du focus est une étape préalable à la détection de l'intention puisque cette dernière nécessite de connaître la cible. Pour réaliser un modèle avec une détection de la tâche dans un contexte multicible, la prédictivité de la sélection est un objectif essentiel.

Dans la partie suivante, différentes solutions de la littérature pour déterminer le focus d'attention à partir du comportement sont étudiées en mettant l'accent sur les contraintes de notre contexte.

3.2 État de l'art des modèles de l'attention conjointe

Pour détecter le focus d'attention à partir du comportement, une piste consiste à s'inspirer du fonctionnement du cerveau humain. Cette approche est exploitée pour la

détection de l'intention dans le chapitre précédent. Les mécanismes en jeu dans l'attention conjointe induisent la capacité humaine à déterminer le focus d'attention de l'autre. Différents modèles de cette capacité sont proposés dans la littérature.

Hoffman et al. proposent un modèle d'attention conjointe dédié à l'interaction homme-robot [Hoffman 06]. Une carte de saillance est estimée à partir des objets de la scène et de la direction du regard du sujet. Ce système reste limité à des cas discrets. D'autres travaux exploitent la pose du visage pour inférer le focus d'attention dans un contexte similaire [Yucel 13]. Un vecteur qui correspond à la direction du regard est estimé à partir de la pose 3D du visage. L'objet de la scène le plus proche de ce vecteur est sélectionné. Les auteurs ne proposent pas de méthode pour construire une carte de saillance. Parks et al. exploitent la pose du visage pour déterminer une carte de saillance qui prédit les saccades oculaires réelles [Parks 14]. Le système fonctionne sur des images 2D fixes. La carte obtenue présente l'avantage de donner un résultat continu.

Il existe peu de systèmes qui exploitent le geste pour déterminer le focus d'attention. Schauerte et al. proposent un système basé sur la direction d'un geste de pointage [Schauerte 14]. Ce système repose sur un geste déictique symbolique. Il propose de créer une carte de saillance à partir de la scène et du geste.

Le tableau 4.1 synthétise ces résultats en mettant l'accent sur les contraintes relevées dans la partie précédente.

	Application	**Prédictif**	**Cas discret**	**Cas continu**
[**Hoffman 06**]	Interaction homme-robot	non	oui	non
[**Yucel 13**]	Interaction homme-robot	non	oui	non
[**Parks 14**]	Prédiction de saccades occulaires	non	oui	oui
[**Schauerte 14**]	Interaction homme-robot	non	oui	non

Tableau 4.1 - *Synthèse des modèles du focus d'attention basés sur le comportement humain*

Parmi les solutions analysées, seules les propositions à base d'une carte de saillance semblent possibles pour estimer le focus d'attention de manière continue. Le cas discret est alors un cas particulier de celui-ci. Construire une carte de saillance semble donc une solution adaptée au problème posé. L'ensemble de ces approches estiment le focus d'attention selon la position du regard ou la direction du geste à un instant donné. Ainsi, **le caractère prédictif est absent**. Dans ce travail, ces indices sont appelés

"**caractéristiques locales**" du focus d'attention. Pour dépasser cette limite, ce travail propose d'inclure la direction de déplacement du point regardé et du geste, considérés comme des "**caractéristiques prédictives**". Pour réaliser la fusion de l'ensemble de ces caractéristiques, le modèle choisi doit donc être capable d'intégrer des **données multimodales** à partir du geste et du regard. De plus, les données locales et prédictives **ne sont pas synchronisées temporellement**. La partie suivante propose un modèle basé sur les champs neuronaux dynamiques pour réaliser la fusion de ces caractéristiques.

Le modèle haut niveau proposé doit ainsi avoir les propriétés suivantes :

- la capacité d'intégrer des données multimodales, potentiellement contradictoires et non synchronisées temporellement
- l'estimation continue de la probabilité attentionnelle sur l'ensemble de la scène pour construire une carte de saillance
- la capacité à faire émerger des points de focus stables
- l'évaluation dynamique du focus lors de l'interaction

Pour proposer ce modèle, une solution consiste à s'inspirer du fonctionnement du cerveau humain. Le tissu nerveux est constitué de neurones interconnectés selon des schémas complexes. Rougier et al. montrent que l'attention peut être vue comme une propriété émergente d'une population de neurones [Rougier 06]. Ces populations interagissent selon des mécanismes d'excitation et d'inhibition. Ces derniers font émerger une compétition entre groupes de neurones. Cette compétition neuronale possède des propriétés dynamiques adaptées pour réaliser le caractère sélectif du processus d'attention. Exploiter un modèle de la dynamique corticale semble ainsi une solution prometteuse pour déterminer le focus d'attention.

3.3 Les champs neuronaux dynamiques

Les champs neuronaux dynamiques sont des modèles qui décrivent l'évolution spatio-temporelle de la dynamique corticale. Ils modélisent le taux de décharge moyen d'une population de neurones. Ils sont un bon modèle des processus cognitifs attentionnels grâce à leurs propriétés :

- **l'intégration de consignes contradictoires**
  des consignes contradictoires peuvent coexister sur un même champ de neurones
- **la bifurcation**
  la dynamique du champ permet de fusionner ou de scinder des consignes proches, ce qui permet d'éliminer des régimes oscillants
- **l'hystérésis de la décision**
  des effets mémoire permettent une stabilité du système malgré d'éventuelles occultations temporelles, par exemple des erreurs de détection
- **le maintien de bulles d'activité** [2]

2. Dans le domaine des champs neuronaux dynamiques, une zone d'activité importante du champ est

les champs neuronaux sont capables de faire émerger et de maintenir des bulles d'activité en lien avec l'excitation fournie par un stimulus

Ils présentent l'avantage de pouvoir intégrer des consignes non synchronisées temporellement. De plus, ils permettent de retrouver une continuité entre les points de discrétisation. Ce modèle semble donc adapté pour réaliser la fusion de données multimodales et non synchronisées. Leur continuité est une propriété intéressante pour gérer le cas continu du focus d'attention.

La dynamique de ces populations de neurones est régie par des équations intégro-différentielles. Ces équations sont introduites par Amari en 1977, qui a démontré la stabilité de ce modèle et étudié sa dynamique dans le cas unidimensionnel [Amari 77]. Un champ neuronal à une dimension est constitué de neurones caractérisés par leur abscisse. L'auteur propose de décrire l'évolution temporelle du potentiel d'un neurone i d'abscisse x_i selon la loi suivante :

$$\tau \cdot \frac{\partial u(x_i,t)}{\partial t} = -u(x_i,t) + \int_{-\infty}^{+\infty} w(x_i - x_k) \cdot f[u(x_k,t)]\mathrm{d}x_k + h + s(x_i,t) \tag{4.3}$$

avec τ une constante de temps, $u(x_i,t)$ l'activité du neurone i d'abscisse x_i à l'instant t, $w(x_i - x_k)$ le noyau d'une fonction qui caractérise l'influence du neurone voisin k d'abscisse x_k sur le neurone i d'abscisse x_i, f une fonction de seuillage, h le potentiel de repos du champs neuronal et $s(x_i,t)$ le stimulus en entrée sur le neurone i.

L'entrée du champ neuronal est un stimulus sensoriel. Ce stimulus active les neurones du champ. Les neurones activés interagissent entre eux. Cette interaction peut être excitatrice ou inhibitrice. L'activité d'un neurone à un instant donné dépend aussi de l'activité du champ aux instants précédents.

Leur dynamique peut être simulée numériquement par une discrétisation temporelle et spatiale des équations continues, pour une population de N neurones, pendant une période T pas. La démonstration est donnée en annexe A.

$$u(x_i,T) = \sum_{n=1}^{T} \frac{(\tau-1)^{n-1}}{\tau^n} \cdot s_{tot}(x_i, T-n) \tag{4.4}$$

où $s_{tot}(x_i, T-n)$ caractérise la somme du stimulus en entrée et de l'influence des neurones voisins selon l'équation suivante :

$$s_{tot}(x_i,T-n) = s(x_i,T-n) + \sum_{k=1}^{N} w(x_i - x_k) \cdot f[u(x_k,T-n)] \tag{4.5}$$

La section suivante détaille l'estimation de chacun de ces paramètres pour construire une carte de saillance de la scène à partir du comportement de l'opérateur.

appelée "bulle d'activité".

3.4 Estimation d'une carte de saillance basée comportement

Cette partie propose d'exploiter un champ de neurones dynamique pour réaliser une carte de saillance basée sur le comportement de l'utilisateur. Cette estimation de la saillance est continue sur la scène. Ce travail considère que dans le cadre du simulateur proposé, le focus d'attention ne porte que sur un point donné sur le plan du substrat. Le problème devient alors bidimensionnel.

3.4.1 Le stimulus d'activation du champ

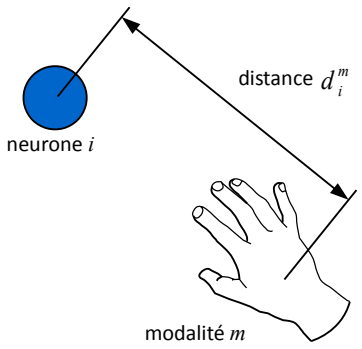
L'entrée d'un champ de neurone est un stimulus d'activation. Ce travail propose d'activer le champ par un stimulus basé sur les signaux de bas niveau du comportement humain caractéristiques du focus d'attention relevés dans la section 2. En particulier, deux types d'indices sont exploités. Les **indices locaux** reposent sur une estimation de la position du regard ou de la main à un instant donné. Ce type d'indice est employé dans toutes les méthodes de la littérature relevées dans la partie 3.2.

D'autre part, l'objectif est de déterminer le focus de manière prédictive. La direction de déplacement du regard et du geste donne une indication prédictive sur leur position. Ce travail propose ainsi d'inclure ces **indices directionnels**.

- **Indices locaux**

Ces indices sont basés sur la distance d entre le curseur [3] et chaque neurone du champ. L'activation est de type gaussien. L'écart type σ permet de gérer la distance limite d'activation des neurones autour de la position du curseur. À partir de la distance, une probabilité normalisée est estimée pour chaque neurone. La probabilité locale ${}^{loc}P_i^m(t)$ pour le neurone i d'être le focus d'attention à l'instant t, évaluée à partir de la modalité $m = \{main, regard\}$ est définie suivant l'équation :

$$ {}^{loc}P_i^m(t) = \frac{e^{-\frac{d_i^m(t)}{2\sigma^2}}}{\sum_{i=1}^{N} e^{-\frac{d_i^m(t)}{2\sigma^2}}} \tag{4.6} $$



avec σ l'écart type de la gaussienne d'activation, estimé de façon à activer le voisinage

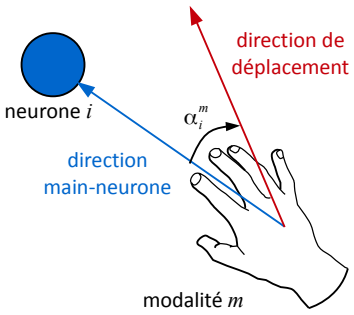
3. La position de la main ou du regard projetée sur l'écran est appelée curseur.

proche du curseur en tenant compte de la précision du capteur. Dans ce travail, il est fixé à 2 fois la distance entre deux neurones consécutifs.

- **Indices directionnels**

Ces indices sont basés sur l'angle α entre la trajectoire du curseur et la direction curseur-neurone. La probabilité directionnelle ${}^{dir}P_i^m(t)$ pour le neurone i d'être le focus d'attention à l'instant t, évaluée à partir de la modalité m est définie suivant l'équation :

$$ {}^{dir}P_i^m(t) = \frac{e^{-\frac{\alpha_i^m(t)}{2\sigma^2}}}{\sum_{i=1}^{N} e^{-\frac{\alpha_i^m(t)}{2\sigma^2}}} \tag{4.7}$$



- **Calcul du stimulus $s(x_i, t)$ en entrée à partir des probabilités locale et directionnelle**

Il existe plusieurs voies possibles pour générer des stimuli à partir de ces indices : locale, directionnelle ou hybride. Elles sont évaluées comparativement dans la section 4. La première repose sur les indices locaux. Le stimulus en entrée $s(x_i, t)$ est directement calculé à partir de la probabilité locale :

$$s(x_i, t) = {}^{loc}P_i^m(t) \tag{4.8}$$

Dans le deuxième cas, le stimulus en entrée correspond à la probabilité directionnelle :

$$s(x_i, t) = {}^{dir}P_i^m(t) \tag{4.9}$$

La troisième méthode repose sur une somme pondérée de ces deux probabilités. D'après la loi d'isochronie du mouvement, la vitesse de déplacement de la main est proportionnelle à la distance à la cible. Ce travail propose ainsi d'attribuer une pondération plus importante aux indices directionnels lorsque la vitesse augmente, et une pondération plus forte à basse vitesse pour les indices locaux :

$$s(x_i, t) = \omega_{loc} \cdot {}^{loc}P_i^m(t) + \omega_{dir} \cdot {}^{dir}P_i^m(t) \tag{4.10}$$

$$\omega_{loc} = 1 - \omega_{dir} \tag{4.11}$$

$$\omega_{dir} = a \cdot v(t) \tag{4.12}$$

avec ω_{loc} le facteur de pondération de la probabilité locale, ω_{loc} le facteur de pondération de la probabilité directionnelle, et $v(t)$ la dérivée de la distance entre la main et le neurone, a un coefficient qui correspond à l'inverse de la vitesse maximum de la main. Cette dernière correspond au maximum de l'isochronie (voir chapitre 3, section 2.1).
Ces trois méthodes sont illustrées sur la figure 4.6.

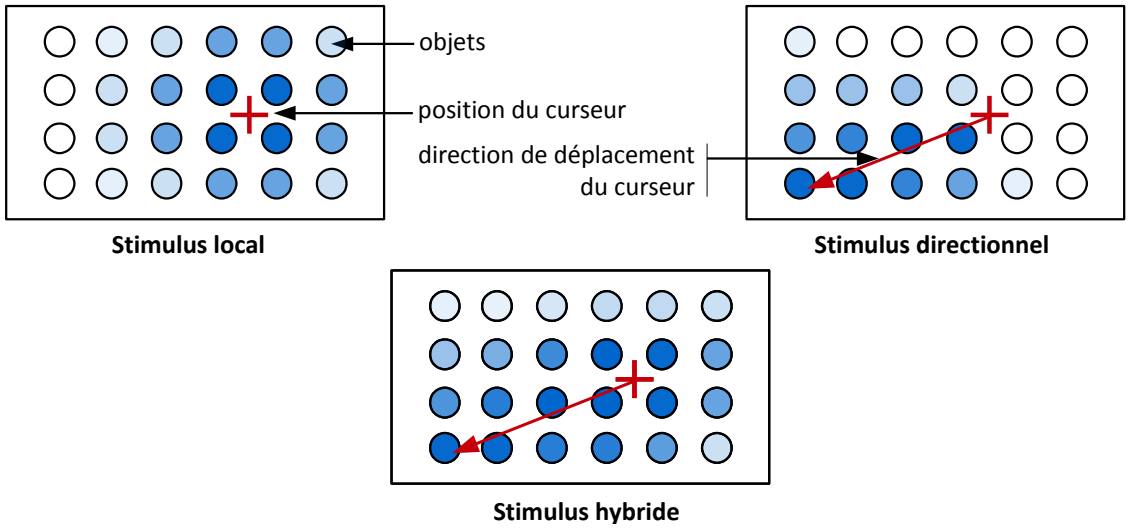


Figure 4.6 - *Trois méthodes d'activation du champ de neurones. L'intensité du bleu représente la valeur du stimulus en entrée du champ.*

3.4.2 L'influence des neurones voisins

Pour donner au champ neuronal la capacité de faire émerger une attention sélective, il est nécessaire que des comportements de collaboration et de compétition entre les neurones puissent exister. Dans ce but, l'influence $I(x_i, t)$ des neurones voisins k d'abscisse x_k sur le neurone i d'abscisse x_i est implémentée selon l'équation suivante :

$$I(x_i, t) = \sum_{k=1}^{N} w(x_i, x_k) \cdot f[u(x_k, t)] \tag{4.13}$$

L'influence du neurone y dépend de la fonction de pondération $w(x_i, x_k)$). Dans le système proposé par Amari [Amari 77], le champ neuronal est de type inhibition latérale :

- Les connexions excitatrices dominent pour les neurones proches
- Les connexions inhibitrices dominent à plus grande distance

Ce noyau d'interaction est modélisé par une différence de gaussiennes. Pour réaliser ce comportement, une convolution spatiale entre la différence de gaussiennes w et l'activité neuronale du champ u est mise en place (fig.4.7 à gauche). Dans l'équation d'Amari, l'activité u est seuillée par une fonction $f(u)$: en dessous d'un seuil d'activation, le neurone ne transmet rien au voisin ; au dessus de ce seuil, il transmet la valeur 1.

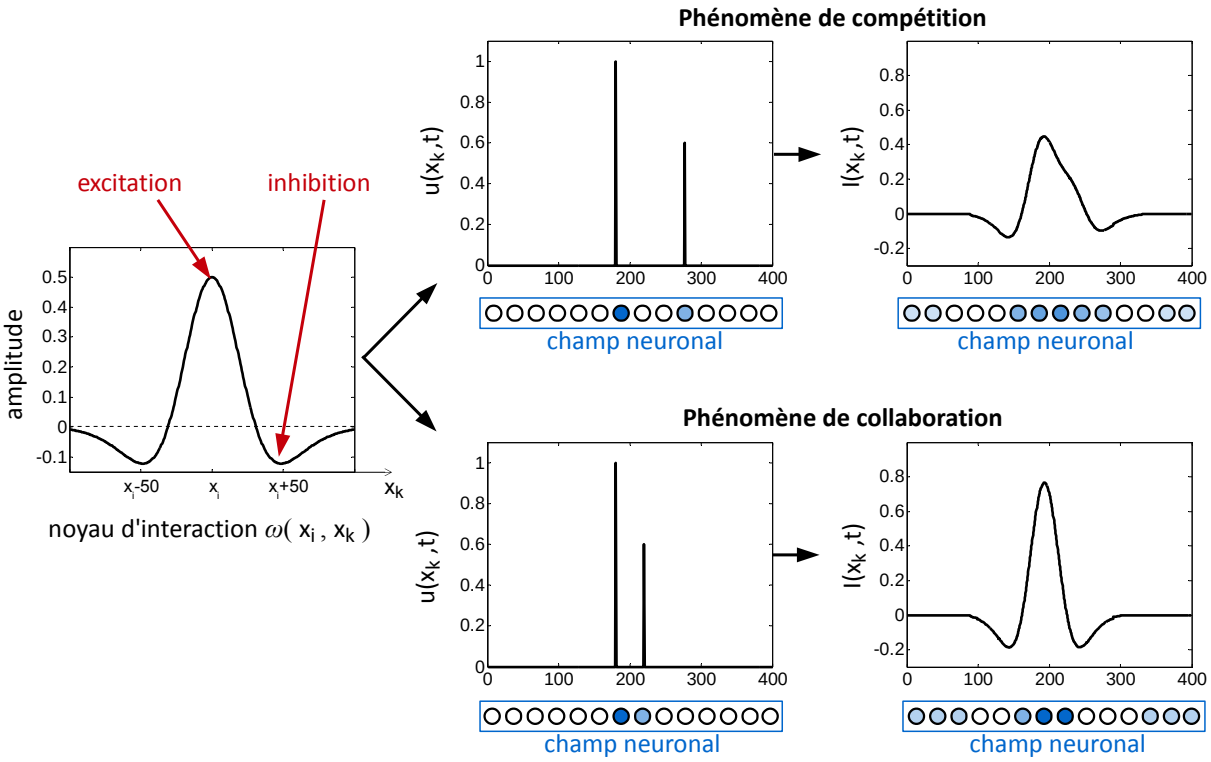


Figure 4.7 - *Une convolution spatiale est réalisée entre la différence de gaussiennes et l'activité neuronale pour modéliser l'influence des neurones voisins.*

Ce mécanisme rend possible l'émergence de comportements de compétition entre des entrées contradictoires. Ainsi, les stimuli liés au regard et à la main créent une bulle d'activité située à la position moyenne lorsqu'ils sont suffisamment proches (fig. 4.7 en bas). D'autre part, si ces deux stimuli sont trop éloignés, les neurones entrent en compétition en inhibant réciproquement leur activité. L'amplitude résultante est faible (fig. 4.7 en haut).

3.4.3 Le mécanisme d'hystérésis

Pour conserver une influence des instants précédents, une hystérésis est appliquée à l'activité neuronale. L'influence de l'activité à un instant donné décroit selon la fonction $\frac{(\tau-1)^{n-1}}{\tau^n}$ (voir équation (4.4)). Pour cela, une convolution temporelle est réalisée entre la somme des stimulus s_{tot} et cette fonction. À partir de cette hystérésis, une activation limitée dans le temps a une influence moindre sur l'activité neuronale en sortie. Ce mécanisme limite l'influence des fausses détections de la main et du regard par la Kinect et/ou le TrackIR. Il réalise une moyenne pondérée temporelle du signal, et réduit ainsi le bruit sur le stimulus. À l'inverse, une activation prolongée sur un neurone donné renforce son activité en sortie (fig. 4.8).
Cette dernière propriété est intéressante dans le cas du stimulus directionnel. La position visée par l'opérateur est activée par le stimulus directionnel pendant toute la durée du

geste d'atteinte. La bulle d'activité créée à la position cible est ainsi renforcée par ce mécanisme d'hystérésis.

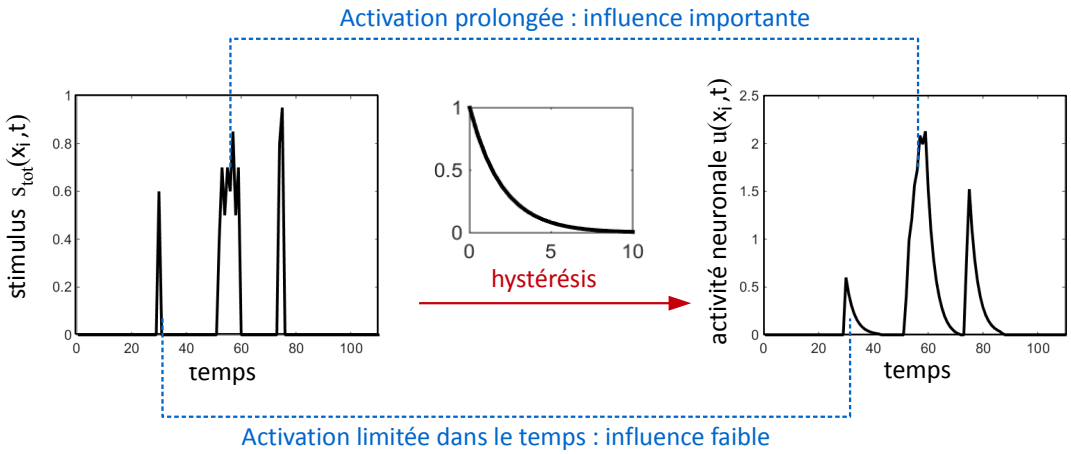


Figure 4.8 - *Exemple de l'influence de l'hystérésis sur l'activité d'un neurone.*

L'ensemble des mécanismes décrits dans cette section a pour objectif de construire une carte de saillance de la scène. La partie suivante donne des exemples de cartes de saillance obtenues selon les différents types de stimulus fournis en entrée.

3.4.4 Carte de saillance estimée

Selon la méthode d'activation du champ de neurones, différents types de cartes de saillance sont observés (fig. 4.11). Le stimulus local crée une bulle d'activité centrée sur le curseur. Le stimulus directionnel crée une forte activité neuronale dans la direction du geste. La somme pondérée des deux types de stimulus provoque la formation d'une bulle d'activité dont la position varie selon la vitesse de déplacement du curseur. À partir d'une carte de saillance donnée, la partie suivante décrit la méthode de sélection du focus d'attention proposée pour le cas discret et le cas continu.

Pour établir l'influence des paramètres d'hystérésis et d'excitation-inhibition, un exemple est proposé. Pour un même stimulus en entrée, différentes cartes de saillance sont estimées en faisant varier ces paramètres. Le signal en entrée correspond à un déplacement du curseur de la main du coin supérieur gauche vers le coin inférieur droit de l'image. Dans cet exemple, l'activation est réalisée avec le stimulus directionnel (voir partie 3.4.1).

3.4.4.1 Choix de l'écart-type

Le mécanisme de coopération/compétition des champs neuronaux repose sur une différence de gaussiennes excitatrice et inhibitrice. Ce fonctionnement est détaillé dans la

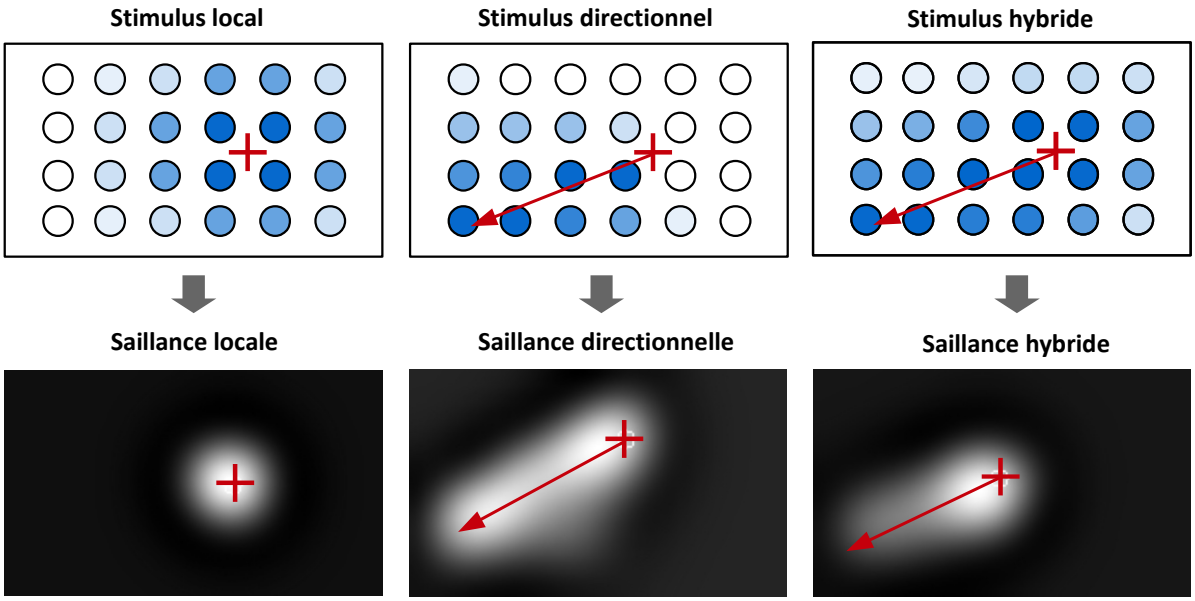


Figure 4.9 - *Cartes de saillance obtenues pour les trois méthodes d'activation du champ neuronal : locale, prédictive et hybride*

partie 3.4.2. La figure 4.10 montre l'influence des écarts-types de ces gaussiennes sur la carte de saillance estimée.

- Lorsque l'écart-type est inférieur à la moitié de la distance entre deux neurones voisins, aucun comportement de coopération ou compétition n'est possible (fig. 4.10 : a, b, c). Ainsi, la carte de saillance obtenue ne présente pas de focus d'attention. De nombreuses zones de la carte possèdent la même valeur de saillance.
- Des écarts-types de l'ordre de la distance entre deux neurones voisins rendent possible l'émergence de phénomènes de collaboration et de compétition locaux (fig. 4.10 : d, e, f). Cependant, ces phénomènes restent très localisés. Toute la direction du geste est activée. Il n'est pas possible de discriminer un focus d'attention sur cette direction.
- Les écarts-types de l'ordre de 1.5 fois la distance entre deux neurones font émerger une zone principale d'activité sur la carte de saillance (fig. 4.10 : g, h, i). Cette zone est appelée **bulle d'activité**. Ce comportement est rendu possible par une compétition de l'ensemble des neurones présents dans la direction du geste.
- Lorsque la valeur des écarts-types est de l'ordre de trois fois la distance entre deux neurones, une bulle d'activité importante est observée (fig. 4.10 : j, k, l). Cependant, cette bulle est diffuse et peu précise. L'ensemble des neurones collaborent et il n'existe plus de compétition pour le focus d'attention.

Pour faire émerger une bulle d'activité précise, l'écart-type de 1.5 fois la distance entre deux neurones voisins est adopté. Avec cet écart-type, les propriétés de collaboration et de compétition réalisent la sélection d'une zone d'activité principale, assimilable au focus d'attention.

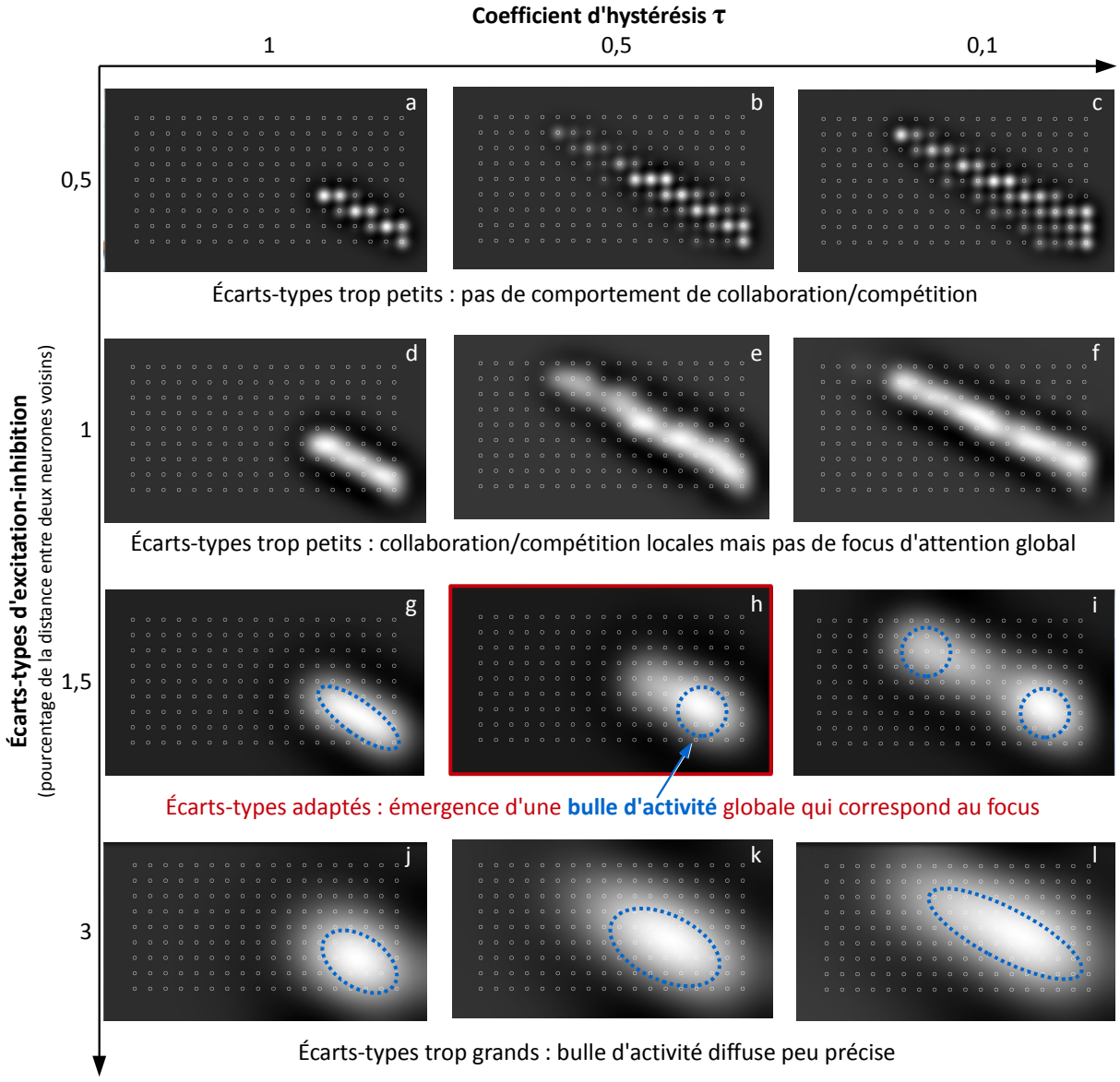


Figure 4.10 - *Cartes de saillance obtenues pour un même stimulus en entrée à partir de différents paramètres d'hystérésis et d'excitation-inhibition. L'écart-type de la gaussienne excitatrice est représenté en pourcentage de la distance entre deux neurones voisins. L'écart-type de la gaussienne inhibitrice est de* 1.5 *fois celui de la gaussienne excitatrice pour obtenir une fonction d'excitation proximale et d'inhibition distale comme proposé par Amari [Amari 77]. La constante de temps* τ *est un indicateur de l'effet mémoire dû à l'influence des instants précédents.*

3.4.4.2 Choix de la constante de temps pour l'hystérésis

L'hystérésis du système crée un effet mémoire. L'activité aux instants précédents a une influence sur la carte de saillance estimée. Ce fonctionnement est détaillée dans la partie 3.4.3. L'influence des instants précédents est régulée par une constante de temps. L'influence de la valeur de cette dernière sur la carte de saillance est évaluée.

- Lorsque la valeur de la constante de temps est proche de 0.1, l'influence des instants précédents est importante. La figure 4.10 i montre l'émergence de deux bulles d'activité distinctes. Une valeur de τ trop faible ne discrimine pas la bulle dûe aux instants passés de la bulle d'activité présente. Il existe donc un risque de fausses détections du focus d'attention. D'autre part, lorsque l'écart-type est de l'ordre de 3 (fig. 4.10 : l), la bulle d'activité est étendue sur l'ensemble de la direction du geste. Cette dernière est peu précise.
- Une valeur de τ de 0.5 fait émerger une bulle d'activité localisée (fig. 4.10 : h). La prise en compte des instants précédents augmente la saillance au niveau de la zone stimulée pendant une durée importante. Les zones activées pendant une durée réduite ont un impact moindre.
- Lorsque τ est proche de 1, l'influence des instants précédents est réduite. La bulle d'activité observée (fig. 4.10 : g) est étendue. Elle est donc peu précise. Cette propriété montre la perte d'information lorsque les instants précédents ne sont pas pris en compte.

Pour obtenir une estimation précise du focus d'attention sans perte d'information, la valeur de τ est fixée à 0.5.

3.5 Sélection du focus d'attention à partir de la carte de saillance

Pour estimer le focus d'attention à partir de la carte de saillance obtenue, les cas discret et continu sont distingués selon la tâche disponible. Le cas continu correspond à la dépose sur un point donné sur le substrat. Le point **maximum de la carte de saillance** fournit une estimation directe du focus d'attention de l'utilisateur (fig. 4.11 à droite). Il est cependant possible de limiter la recherche à une zone donnée du substrat en appliquant un masque qui dépend de la tâche.

Le cas discret correspond à la tâche de saisie. Les objets présents dans la scène virtuelle doivent être pris en compte. Dans ce but, un masque est appliqué sur la carte de saillance. Celui-ci conserve uniquement les zones qui correspondent à des objets. La valeur de la **saillance moyenne est évaluée au niveau de chaque objet**. L'objet dont la saillance moyenne est la plus grande est considéré comme le focus d'attention de l'opérateur (fig. 4.11 à gauche).

Pour éviter l'influence de la taille des objets sur la reconnaissance du focus, la moyenne est estimée sur une zone de taille constante autour du centre des objets. Cette méthode

facilite la discrimination lorsque les objets sont de grande taille et séparés par des distances courtes.

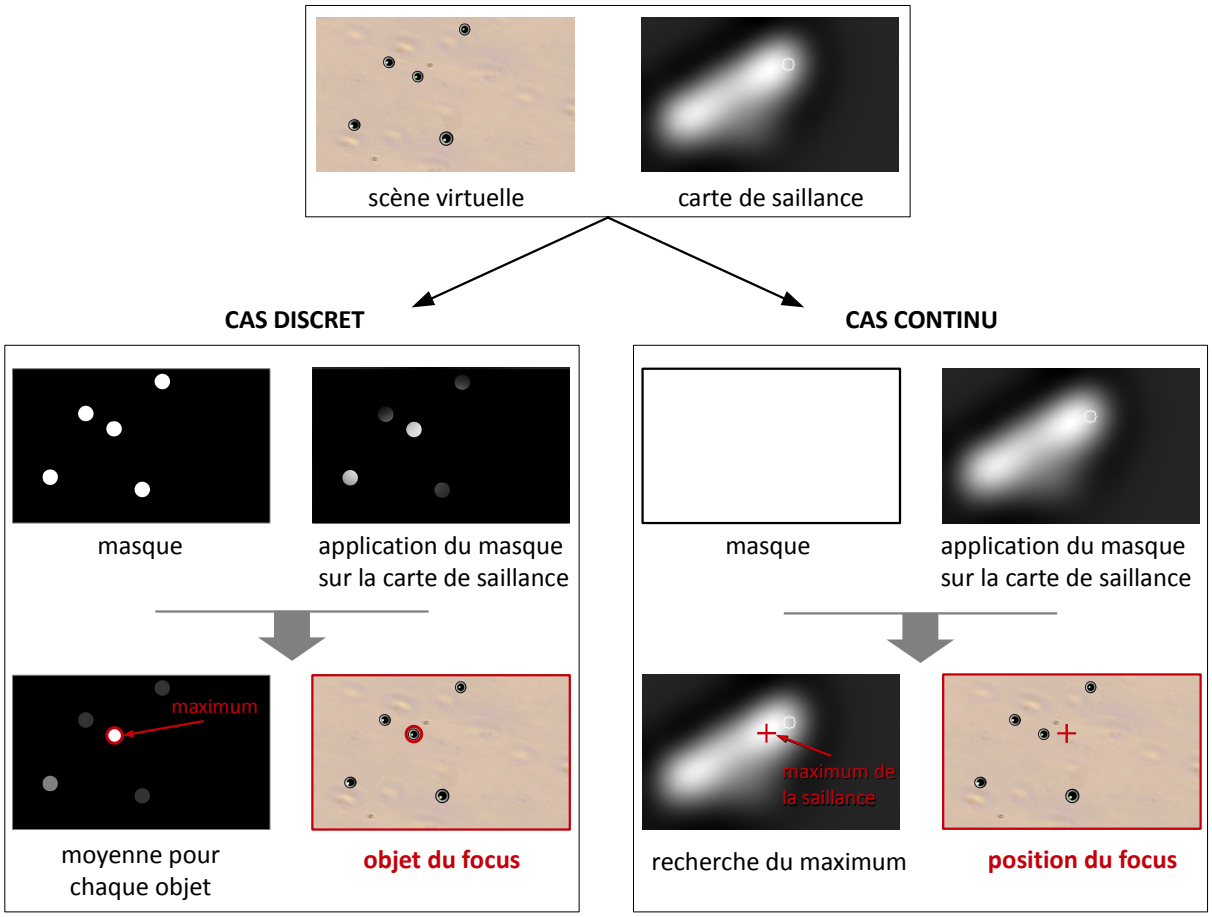


Figure 4.11 - *Méthode d'estimation du focus d'attention dans le cas discret et le cas continu à partir de la scène virtuelle et de la carte de saillance.*

La partie suivante se concentre sur la validation expérimentale du modèle proposé dans le cadre d'une expérience utilisateurs.

4 Évaluation du modèle d'estimation

Une expérience utilisateurs est mise en place pour tester le modèle proposé. En particulier, ce protocole a pour objectif :

- d'évaluer l'apport de chacune des modalités (geste, regard) pour déterminer le focus d'attention,
- de quantifier la prédictivité du système,
- de sélectionner les stimuli les plus pertinents pour activer le champs de neurones (local, prédictif ou hybride).

4.1 Protocole expérimental

La tâche consiste à sélectionner un objet donné sur une grille d'objets en 2 dimensions. Les objets sont représentés par des cercles. L'objet cible est un disque blanc. La position de la main de l'utilisateur est visualisée par un disque rouge. Il est demandé au sujet de sélectionner la cible avec ce curseur. Aucune consigne n'est donnée sur la direction du regard. L'objet est sélectionné lorsque la distance entre le curseur de la main et le centre de l'objet est inférieure à son rayon. Les grilles proposées à l'utilisateur varient selon deux caractéristiques :

- **le nombre d'objets présents** : l'objectif de cette étape est de quantifier la précision du système à partir du taux de reconnaissance du focus d'attention. Cette propriété est estimée lorsque le nombre d'objets de la scène augmente.
- **la difficulté de la tâche** : il est montré que la difficulté de la tâche influence les paramètres cinématiques du gestes [MacKenzie 92]. Il s'agit dans notre cas de valider la généricité du système lorsque la difficulté de la tâche augmente. Pour faire varier celle-ci, la taille des objets est modifiée. Plus l'objet est petit, plus la tâche est difficile. Il est important de rappeler que la taille de l'objet n'a pas d'influence sur le traitement réalisé par le modèle proposé (voir section 3.5). La taille de l'objet influence donc uniquement le comportement de l'utilisateur.

Les différentes grilles présentées aux utilisateurs sont montrées sur la figure 4.12.

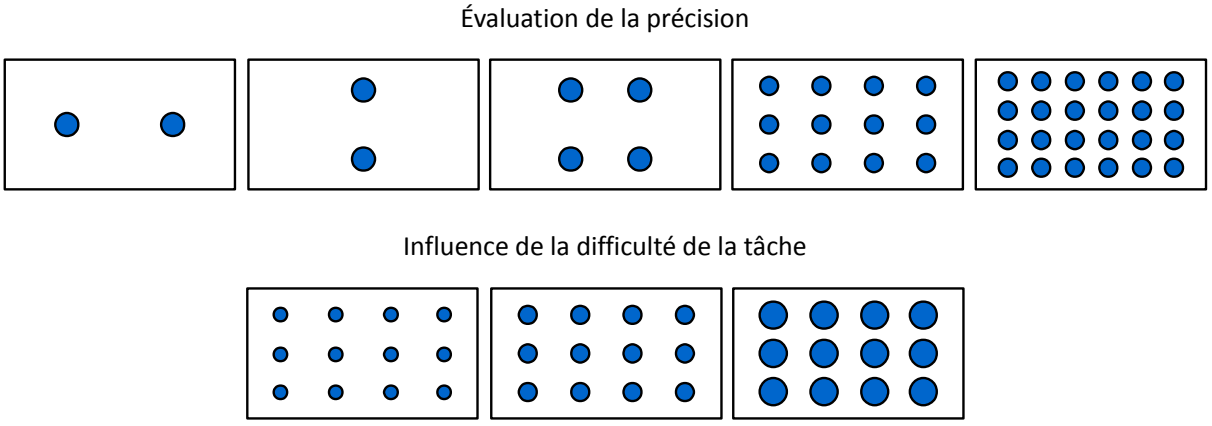


Figure 4.12 - *Test utilisateur réalisé pour plusieurs configurations : la disposition, le nombre d'objets et leur taille varient.*

Lorsque l'objet blanc est sélectionné par l'utilisateur, une nouvelle cible est désignée. La grille est modifiée à chaque fois que 15 objets sont sélectionnés. Chacune des trois méthodes d'activation est ainsi évaluée 5 fois pour chaque grille. Le nombre total de grilles différentes est 8. L'expérience est répétée deux fois pour évaluer la modalité du geste seul et celle de la fusion entre le regard et le geste. Cette expérience est réalisée par 10 utilisateurs adultes naïfs. Une base de données de 2400 gestes est ainsi constituée.

Pour évaluer à la fois le cas discret et le cas continu, les deux méthodes présentées dans la partie précédentes sont employées parallèlement pour déterminer le focus d'attention. Dans le cas discret, l'objet du focus d'attention coïncide à la saillance la plus importante. Dans le cas continu, le focus d'attention estimé correspond au maximum de la carte de saillance. Dans ce cas, l'objet à sélectionner est considéré comme la vérité terrain de la position visée par l'utilisateur. Les parties suivantes détaillent les résultats obtenus pour ces deux cas.

4.2 Cas discret

4.2.1 Évaluation de l'influence du nombre d'objets

L'objet du focus d'attention est estimé à chaque pas de temps à partir du début de l'interaction. Deux critères sont évalués. Le **taux de réussite** est le taux de vrais positifs sur l'ensemble de l'interaction avant l'atteinte de l'objet. Il est important de noter que dans les approches de la littérature, le taux de réussite est estimé uniquement à la fin de l'interaction, lorsque l'objet est atteint.
Un second critère d'évaluation est la **prédictivité** du système. Dans ce but, un critère adapté à son évaluation est proposé. Ce dernier évalue à partir de quel instant de l'interaction le taux de réussite est de 100%. La figure 4.13 illustre ce critère.

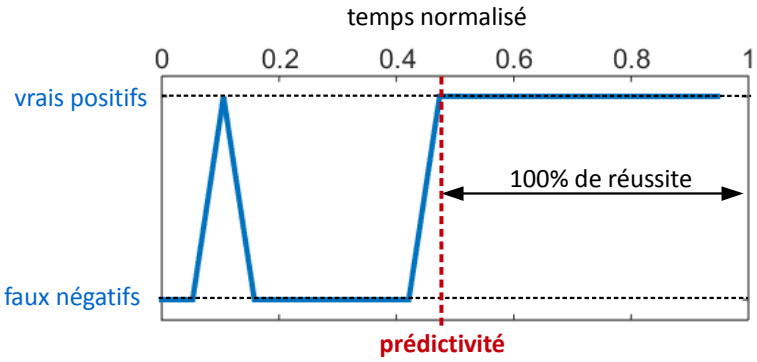


Figure 4.13 - *Définition du critère de prédictivité proposé dans ce travail. La courbe bleue représente l'estimation du focus d'attention pour un objet cible donné en fonction du temps lors d'une tâche.*

Ainsi, une prédictivité de 0.5 indique qu'à partir de la moitié du geste, le focus d'attention est toujours bien reconnu. Plus la valeur de la prédictivité est basse, plus le système est performant. Les courbes suivantes montrent l'influence du nombre d'objets présents dans la scène sur la réussite et la prédictivité du système (fig.4.14).

De manière générale, l'augmentation du nombre d'objets diminue le taux de réussite et rend le système moins prédictif.

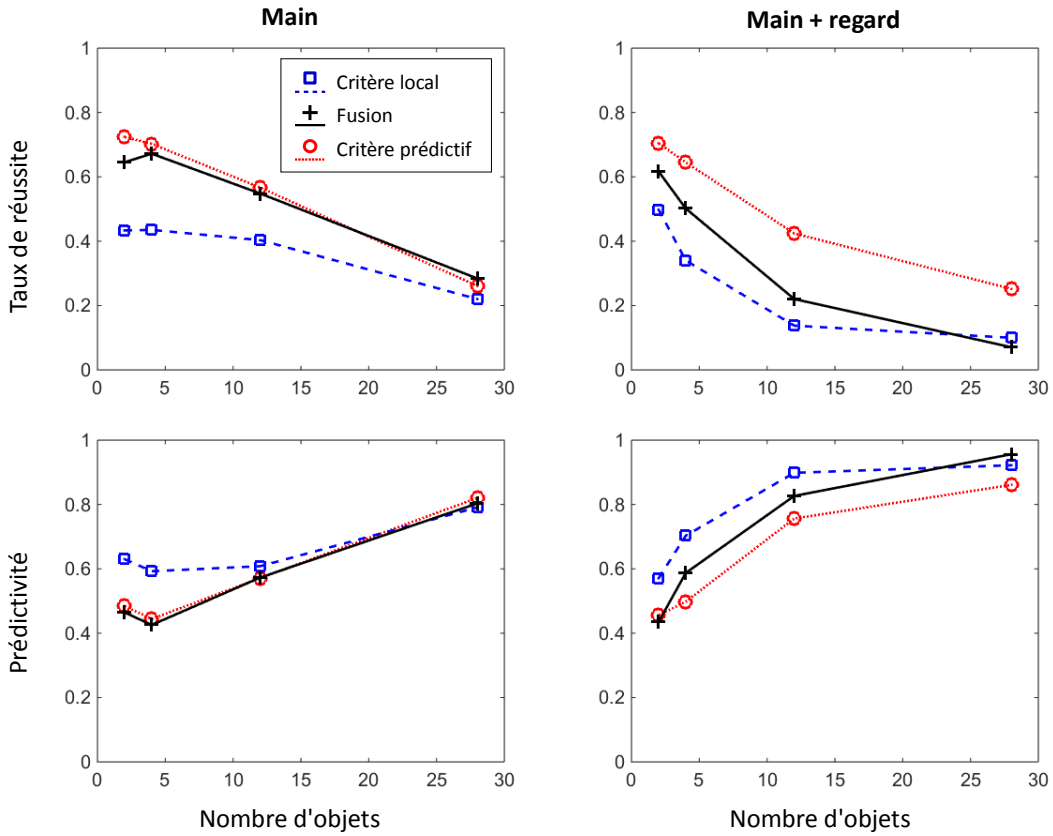


Figure 4.14 - *Taux de réussite en fonction du nombre d'objets présents évalué sur l'ensemble de la durée de la tâche. La prédictivité est évaluée en pourcentage de la durée totale de la tâche.*

Les résultats indiquent que l'activation par un stimulus local obtient dans tous les cas les plus mauvaises performances. Le critère prédictif et la fusion des critères ont des performances similaires lorsque seule la main est considérée. Lorsque les deux modalités sont présentes (main et regard), le critère prédictif fournit les meilleures performances.

La modalité de la main réalise les meilleurs résultats, à la fois en terme de taux de réussite et de prédictivité. Inclure le regard détériore les performances. Pour expliquer cette observation, deux hypothèse peuvent être envisagées. La taille réduite de l'écran implique des déplacements de la tête d'une amplitude faible. Ainsi, il est possible que les utilisateurs exploitent principalement le mouvement des yeux pour suivre les cibles dans la scène. D'autre part, la main est représentée par un curseur dans la scène. L'opérateur dispose ainsi d'un retour visuel de cette modalité. Il s'agit d'une boucle sensori-motrice entre l'utilisateur et l'interface. Cette boucle fermée améliore la précision du geste. Cependant, l'utilisateur ne dispose par de retour sur la position du regard. Ce contrôle en boucle ouverte offre une moins grande précision.

4.2.2 Évaluation de l'influence de la difficulté de la tâche

La figure 4.15 montre l'influence de la difficulté de la tâche dans le cas discret sur le taux de réussite et la prédictivité.

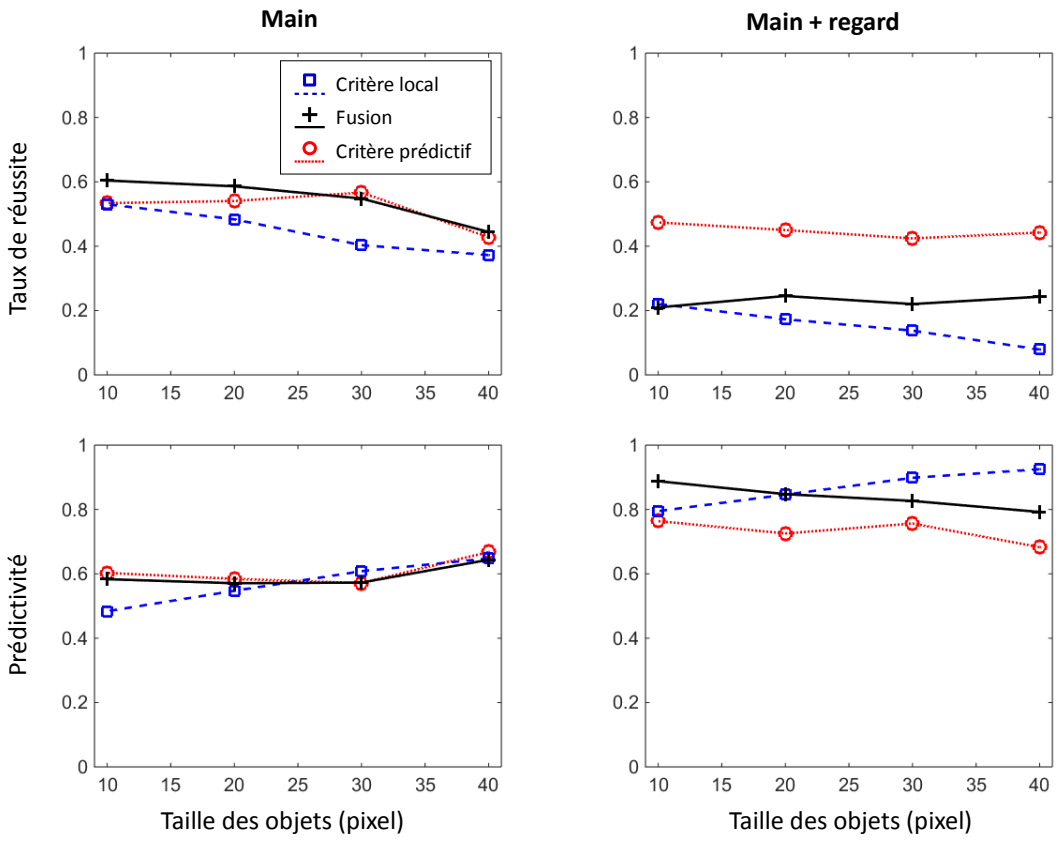


Figure 4.15 - *Taux de réussite en fonction de la difficulté de la tâche évalué sur l'ensemble de la durée du geste. La prédictivité est évaluée en pourcentage de la durée totale de la tâche.*

Contrairement à l'augmentation du nombre d'objets, la difficulté de la tâche implique une influence moindre sur la réussite et la prédictivité. Le modèle proposé semble donc généralisable à des tâches de difficulté variables.

Dans le cas de la modalité de la main, les trois méthodes d'activation obtiennent des résultats similaires. Lorsque les deux modalités sont présentes, le critère prédictif réalise les meilleurs performances. Comme observé précédemment, le regard diminue globalement le taux de réussite et limite la prédictivité.

Il est intéressant de noter que les performances décroissent légèrement en fonction de l'augmentation de la taille des objets de la scène. Ainsi, le système réalise un taux de réussite plus important lorsque la tâche est plus difficile pour l'utilisateur. Une hypothèse est que les gestes de l'utilisateur sont plus lents à l'approche de l'objet. Cette propriété laisse le temps à une bulle d'activité précise de se créer.

4.3 Cas continu

La figure 4.16 montre l'évolution de la distance entre le focus d'attention estimé et la cible réelle dans le cas continu en fonction du temps normalisé de la tâche.

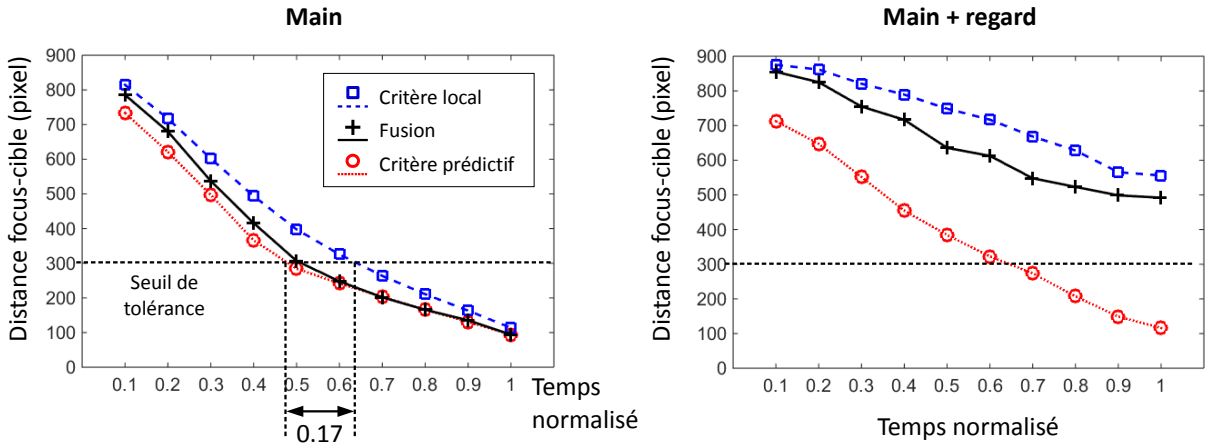


Figure 4.16 - *Évolution de la distance entre le focus d'attention estimé et la cible lors du geste d'atteinte.*

Le stimulus local donne les moins bonnes performances comparé aux deux autres critères. Pour un seuil de tolérance en précision donnée, il est possible d'évaluer la différence de prédictivité entre deux critères. Le seuil est fixé par exemple à une tolérance de 300 pixels. Lorsque seule la modalité de la main est exploitée, il est atteint par le critère prédictif à 48% de la durée totale de la tâche, et par le critère local à 65% de celle-ci. Ainsi, le critère local est plus prédictif de 17% pour ce seuil de précision. Dans le cas de la fusion entre la main et le regard, seul le critère prédictif atteint le seuil de tolérance, à 65% de la durée totale du geste. Le regard fait donc aussi baisser les performances dans le cas continu. Cependant, le critère prédictif limite cette diminution des performances. Lors de mouvements de faible amplitude de la tête, le regard complète la trajectoire de celle-ci pour s'orienter vers la cible visuelle. Le critère prédictif est directionnel. Il complète le mouvement partiellement réalisé par la tête. Ce dernier donne ainsi une meilleure estimation du point regardé à partir de la pose de la tête comparé au critère local.

Les résultats obtenus désignent **la main** comme la modalité la plus adaptée pour déterminer le focus d'attention dans le cas discret comme dans le cas continu. La précision des capteurs de la pose de la tête semble insuffisante pour exploiter cette dernière. Il semble ainsi probable que l'être humain exploite principalement le mouvement des yeux pour déterminer le focus d'attention lorsque l'amplitude des mouvements de la tête est insuffisante.

D'autre part, les résultats montrent l'apport d'un **stimulus prédictif** pour activer le champ de neurones dynamiques. Les approches classiques se concentrent sur l'évaluation locale du focus d'attention à un instant donné. Les résultats obtenus montrent que le

critère proposé, basé sur la direction du mouvement, est à la fois plus précis et plus prédictif que le critère classique. En particulier, il est intéressant de noter que celui-ci est plus précis que le critère hybride dans tous les cas et indépendamment de la modalité. La bulle d'activité qui émerge du modèle semble ainsi être capable de discriminer le point précis du focus dans la direction du geste.

Le modèle proposé offre une méthode prédictive pour estimer le focus d'attention avant l'atteinte de l'objet. Il existe un compromis entre la prédictivité et la précision du système. Ce compromis est mis en avant dans le cas continu. Plus le seuil de tolérance en précision est grand, plus le système est prédictif. D'autre part, il est montré que la difficulté de la tâche influence peu les performances du système. Le modèle proposé est ainsi généralisable à des tâches de difficulté variable.

Ce système est adaptable à la tâche de micromanipulation en assimilant le plan du champ neuronal au plan du substrat. Le cas discret rend possible l'estimation de la sphère cible lors de la saisie de manière prédictive. Plus le nombre de sphères est important, moins le système sera prédictif. Lors de la tâche de dépose, le modèle proposé donne une estimation de la zone attentionnelle sur le substrat à partir du cas continu. Il est ainsi possible de prédire la zone cible de l'opérateur lors de la dépose d'un objet.
Cette sélection est basée uniquement sur le comportement de l'opérateur, sans qu'il lui soit nécessaire d'expliciter une commande symbolique spécifique. Cette méthode constitue donc une perspective intéressante pour réaliser les tâches de sélection dans le cadre d'une interface naturelle.

5 Conclusion

Le chapitre III propose une approche de prédiction de l'intention de l'opérateur. Elle est capable de déterminer une intention de saisie ou de dépose mais nécessite de connaître à l'avance la position de l'objet ou la zone de dépose. Les applications sont donc limitées à des scènes où un seul objet est présent et où une cible préalablement placée sur le substrat indique la zone de dépose. Cependant, les scènes réalistes de micromanipulation comprennent souvent plusieurs objets à manipuler. De plus, la cible n'est généralement pas connue a priori. Pour généraliser le système à ces environnements réalistes, ce chapitre propose une méthode de sélection prédictive de la cible basée sur le comportement de l'opérateur. Le système proposé s'inspire des mécanismes cognitifs attentionnels pour déterminer le focus d'attention de l'opérateur dans la scène. Il exploite un modèle de la dynamique des champs neuronaux. Le champ correspond au plan du substrat. Ce travail propose une méthode d'activation prédictive de ce dernier. Les stimuli d'activation du champ sont basés sur le regard et le geste de l'opérateur. En plus de considérer le voisinage local, le modèle prend en compte les caractéristiques cinématiques de ces stimuli comme la direction du geste. La dynamique obtenue fait émerger un maximum dans le champ neuronal qui est assimilée au focus d'attention.

Une expérience utilisateurs est mise en place pour évaluer la pertinence de différents stimuli basés sur le regard et le geste et leur somme pondérée. Parmi les stimuli testés, le geste produit les meilleures performances par rapport au regard ou à une pondération dynamique pour déterminer le focus d'attention. Néanmoins, cette différence peut être en partie expliquée par la meilleure qualité des données capteurs brutes relatives à la main par rapport au regard. Les résultats montrent qu'exploiter la direction du geste en entrée du système plutôt qu'une position locale de la main classiquement utilisée dans la littérature renforce la prédictivité et le taux de réussite du système.

Le modèle proposé ouvre des perspectives pour une sélection d'objets plus naturelle dans le cadre d'une interface de micromanipulation. Ce dernier réalise la sélection prédictive de la cible lors de tâches discrètes d'atteinte d'un objet. Cette tâche correspond à la phase de saisie lors de la manipulation. Lors de la phase de dépose, le modèle sélectionne la zone cible sur le substrat. Ce modèle est donc pertinent pour la sélection prédictive de la cible de l'opérateur lors de ces tâches. Cette sélection est basée sur le comportement naturel de l'opérateur. Elle ne nécessite pas de commande symbolique spécifique en entrée. L'approche proposée est une solution naturelle pour réaliser la sélection lors d'une tâche de micromanipulation.

Conclusions et Perspectives

Les techniques de micromanipulation sont prometteuses pour le développement de nouveaux produits, grâce aux possibilités d'accès aux micro-objets individuels. Les capacités de manipulation et de caractérisation électromécanique sur des composants inorganiques (électronique, photonique, MEMS...) ou organiques (cellules, bactéries, biopolymères...) seraient un moteur d'innovation dans les PME et la recherche. L'adoption actuelle des systèmes de micromanipulation reste cependant limitée par la complexité de leur utilisation.

La problématique de ce travail se concentre donc sur la synthèse d'une interface naturelle et intuitive pour interagir avec le micromonde par l'intermédiaire de la réalité virtuelle. Dans ce mémoire, une interface est considérée comme naturelle si l'opérateur interagit sans avoir à apprendre un langage symbolique pour communiquer l'action qu'il souhaite réaliser. En particulier, l'étude de la littérature montre un manque d'outils adaptés à la détection des décisions de l'utilisateur. Dans le cadre de la micromanipulation, ces décisions correspondent à des tâches unitaires typiques : la saisie, le déplacement, la dépose et la sélection d'objets. Pour dépasser ces limites, ce travail propose des modèles de haut niveau capables d'interpréter le comportement naturel humain lors de tâches de manipulation. Ces modèles reposent sur l'interprétation de signaux comportementaux de bas niveau extractibles par les capteurs.

- Pour réaliser l'évaluation utilisateurs des stratégies d'interaction proposées dans cette thèse, un simulateur intuitif de micromanipulation par forces d'adhésion est mis en place. Cette interface masque la complexité du système réel. Elle inclut une main virtuelle afin de faciliter le lien entre la main de l'utilisateur et l'effecteur dans la scène de manipulation. Des méthodes de couplage utilisateur-main virtuelle et main virtuelle-scène de micro manipulation sont proposées. Les tâches unitaires typiques de la micromanipulation sont la saisie, le déplacement et la dépose de

micro-objets sur un substrat. Pour détecter les décisions de l'utilisateur liées à ces opérations, une méthode par reconnaissance de gestes est mise en place.

- Les performances de l'approche par reconnaissance de gestes sont limitées dans le cadre d'une interface naturelle. Pour dépasser cette limite, une approche non symbolique de prédiction de l'intention est proposée. Elle est basée sur l'analyse du comportement naturel de l'opérateur lors de la manipulation. Ce travail montre qu'il est possible de reconnaître l'intention à partir de la cinématique du geste ciblé. Cette observation conduit à une reformulation de la loi d'isochronie du mouvement et du profil gaussien de la vitesse pour intégrer ce paramètre. Pour détecter les décisions, un modèle haut niveau basé sur la prédiction de l'intention par un être humain est adopté. Ce dernier propose un contrôle en position. Pour distinguer un geste ciblé d'un geste aléatoire lorsque ceux-ci ont la même trajectoire, ce travail propose de reformuler ce modèle en vitesse. Il repose sur la construction d'un prédicteur à partir des invariants en vitesse du geste ciblé. Sans consigne donnée à l'utilisateur, les performances obtenues sont significativement améliorées en termes de succès et de durée de la tâche comparées à l'approche classique par reconnaissance de gestes. De plus, un questionnaire utilisateurs montre une préférence significative des sujets par rapport à l'approche classique.
- Le modèle de l'intention proposé nécessite de connaître l'objet cible pour la saisie, et la zone cible du substrat pour la dépose. Pour généraliser l'interface à des environnements de micromanipulation plus réalistes où la cible n'est pas connue a priori, une méthode de sélection basée sur le comportement de l'opérateur est proposée. Ce système s'inspire des mécanismes cognitifs attentionnels pour déterminer le focus d'attention de l'opérateur dans la scène. Il exploite un modèle de la dynamique des champs neuronaux. Ce travail propose une méthode d'activation prédictive de ce dernier. La dynamique obtenue fait émerger un maximum dans le champ neuronal, assimilé au focus d'attention. Une expérience utilisateurs est mise en place pour évaluer l'influence de différents stimuli basés sur le regard et le geste de l'utilisateur. Les résultats obtenus montrent que le stimulus basé sur la direction du geste donne une estimation plus robuste et prédictive du focus d'attention comparée à une activation locale, généralement exploitée dans la littérature. Ce modèle distingue un cas discret de sélection d'un objet et un cas continu qui réalise la sélection d'un point quelconque du plan. Ces cas correspondent aux tâches unitaires de saisie et dépose lors de la micromanipulation. Le modèle proposé constitue ainsi une perspective pour sélectionner une cible de manière naturelle lors d'une tâche de manipulation.

Les résultats obtenus dépendent fortement des capteurs de vision disponibles. En particulier, la Kinect fournit des données bruitées. Cependant, les modèles proposés dans ce travail ont vocation à être généralisables à de nombreux types de capteurs. Les mesures de bas niveau exploitées en entrée sont généralement des vitesses et des positions, extractibles à partir de nombreux dispositifs. Ainsi, la position de la main exploitée dans les modèles de l'intention et du focus d'attention est extractible par exemple à partir d'un bras haptique. Il serait intéressant d'évaluer la sensibilité du système à la qualité des données à partir de différents dispositifs.

Ces modèles de haut niveau sont généralisables à d'autres modalités comportementales humaines. Le chapitre 4 exploite par exemple la main et le regard comme deux entrées possibles d'un même modèle. L'émergence de nouveaux capteurs plus précis comme la Kinect 2 et d'algorithmes de suivi plus robustes ouvrent des perspectives pour valider l'indépendance à la modalité de nos modèles. Le modèle actuel de l'intention exploite la modalité de la main. Le regard est un autre indice essentiel de celle-ci. Une détection précise du point regardé prédirait certainement l'intention à partir du regard. De plus, la Kinect 2 réalise le suivi des doigts. Le modèle actuel ne considère que la position du poignet. Il serait possible de l'appliquer à la fermeture/ouverture des doigts lors de la saisie/dépose pour une estimation plus précise de l'intention.

Ce travail de thèse propose des méthodes naturelles pour détecter les décisions de l'utilisateur. Le système actuel se concentre sur la manipulation par adhésion. Les opérations de saisie, dépose et sélection sont traitées. Ces dernières constituent un ensemble de tâches unitaires typiques de micromanipulation. L'approche proposée ouvre ainsi des perspectives d'application vastes dans le champ des interfaces de micromanipulation d'objets virtuels au sens large. Ces tâches typiques sont rencontrées dans de nombreuses autres champs applicatifs qui impliquent une interaction avec la réalité virtuelle. En particulier, les modèles proposés ne nécessitent pas de consigne donnée à l'utilisateur. Ils sont donc une solution prometteuse pour des dispositifs d'interaction dans les lieux publics, par exemple des panneaux publicitaires interactifs, lorsqu'il n'est pas possible de fournir un mode d'emploi à l'utilisateur.

Les travaux réalisés mettent en exergue le problème des symboles dans les interfaces. Les interfaces classiques reposent sur des symboles gestuels pour déclencher les actions du système. Ces symboles donnent une indication sur la forme du geste mais pas sur son sens. Cette limite est appelée "problème de l'ancrage symbolique" dans le domaine de l'intelligence artificielle. L'objectif central de cette thèse est de proposer une alternative non symbolique à ces systèmes. Cependant, éviter tout symbole défini a priori reste un verrou scientifique. Ainsi, les modèles proposés exploitent par exemple un a priori sur la forme gaussienne des vitesses. Un nouveau domaine émerge pour dépasser ces limites avec la robotique développementale. Cette dernière propose d'exploiter uniquement les interactions sensori-motrices entre le système et son environnement pour faire émerger ses capacités. Une perspective de ce travail est de faire émerger les modèles proposés par ce type d'apprentissage. Une telle interface éviterait la nécessité de fournir des aprioris au système.

L'approche non symbolique, basée sur l'interprétation du comportement humain, constitue ainsi une perspective prometteuse pour de nouvelles interfaces homme-machine plus naturelles et intuitives.

Annexe A

Discrétisation de l'équation des champs neuronaux dynamiques

L'équation des champs neuronaux dynamiques proposée par Amari [Amari 77] est la suivante :

$$\tau \cdot \frac{\partial u(x_i,t)}{\partial t} = -u(x_i,t) + \int_{-\infty}^{+\infty} w(x_i - x_k) \cdot f[u(x_k,t)]\mathrm{d}x_k + h + s(x_i,t) \qquad \text{(A.1)}$$

avec τ une constante de temps, $u(x_i,t)$ l'activité du neurone i d'abscisse x_i à l'instant t, $w(x_i - x_k)$ le noyau d'une fonction qui caractérise l'influence du neurone voisin k d'abscisse x_k sur le neurone i d'abscisse x_i, f une fonction de seuillage, h le potentiel de repos du champs neuronal et $s(x_i,t)$ le stimulus en entrée sur le neurone i.

Elle peut être formulée de manière discrète sur N neurones suivant l'équation :

$$\tau \cdot \frac{\Delta u(x_i,t)}{\Delta t} = -u(x_i,t) + \sum_{k=1}^{N} w(x_i - x_k) \cdot f[u(x_k,t)] + h + s(x_i,t) \qquad \text{(A.2)}$$

On pose $\Delta t = 1$ (pas de temps constant). T est le nombre de pas. Le potentiel u au pas suivant $t+1$ peut s'écrire ainsi :

$$\tau \cdot (u(x_i,t+1) - u(x_i,t)) = -u(x_i,t) + \sum_{k=1}^{N} w(x_i - x_k) \cdot f[u(x_k,t)] + h + s(x_i,t)$$

$$\Rightarrow \quad u(x_i, t+1) = \frac{1}{\tau}[(\tau - 1) \cdot u(x_i, t) + \sum_{k=1}^{N} w(x_i - x_k) \cdot f[u(x_k, t)] + h + s(x_i, t)] \tag{A.3}$$

On pose :

$$s_{tot}(x_i, t) = s(x_i, t) + \sum_{k=1}^{N} w(x_i - x_k) \cdot f[u(x_k, t)] \tag{A.4}$$

avec $s_{tot}(x_i, t)$ la somme du stimulus reçu et de l'influence sommée des voisins sur le neurone à la position x_i à l'instant t. De plus, on considère le potentiel de repos h comme nul.

$$u(x_i, t+1) = \frac{\tau - 1}{\tau} \cdot u(x_i, t) + \frac{1}{\tau} \cdot s_{tot}(x_i, t) \tag{A.5}$$

$$\begin{aligned} u(x_i, t+2) &= \frac{\tau - 1}{\tau} \cdot u(x_i, t+1) + \frac{1}{\tau} \cdot s_{tot}(x_i, t+1) \\ &= \frac{(\tau - 1)^2}{\tau^2} \cdot u(x_i, t) + \frac{\tau - 1}{\tau^2} s_{tot}(x_i, t) + \frac{1}{\tau} \cdot s_{tot}(x_i, t+1) \end{aligned}$$

$$\begin{aligned} u(x_i, t+3) &= \frac{(\tau - 1)^2}{\tau^2} \cdot [u(x_i, t+1)] + \frac{\tau - 1}{\tau^2} \cdot s_{tot}(x_i, t+1)] + \frac{1}{\tau} \cdot s_{tot}(x_i, t+2) \\ &= \frac{(\tau - 1)^3}{\tau^3} \cdot u(x_i, t) + \frac{(\tau - 1)^2}{\tau^3} \cdot s_{tot}(x_i, t) + \frac{\tau - 1}{\tau^2} \cdot s_{tot}(x_i, t+1) + \frac{1}{\tau} \cdot s_{tot}(x_i, t+2) \end{aligned}$$

$$\Rightarrow \quad u(x_i, t+T) = \left(\frac{\tau - 1}{\tau}\right)^T \cdot u(x_i, 0) + \sum_{n=1}^{T} \frac{(\tau - 1)^{n-1}}{\tau^n} \cdot s_{tot}(x_i, t+T-n) \tag{A.6}$$

Si on pose comme condition initiale $u(x_i, 0) = 0$, et pour $t = t_0 = 0$:

$$\boxed{u(x_i, T) = \sum_{n=1}^{T} \frac{(\tau - 1)^{n-1}}{\tau^n} \cdot s_{tot}(x_i, T-n)} \tag{A.7}$$

Bibliographie

[Agnus 13] **Joël Agnus**, **Nicolas Chaillet**, **Cédric Clévy**, **Sounkalo Dembélé**, **Michaël Gauthier**, **Yassine Haddab**, **Guillaume Laurent**, **Philippe Lutz**, **Nadine Piat**, **Kanty Rabenorosoa et al.** *Robotic microassembly and micromanipulation at FEMTO-ST.* Journal of Micro-Bio Robotics, vol. 8, n° 2, pages 91–106, 2013. 6

[Ai 98] **Zhuming Ai et Torsten Fröhlich**. *Molecular dynamics simulation in virtual environments.* In Computer Graphics Forum, volume 17, pages 267–273. Wiley Online Library, 1998. 15

[Aigner 12] **Roland Aigner**, **Daniel Wigdor**, **Hrvoje Benko**, **Michael Haller**, **David Lindbauer**, **Alexandra Ion**, **Shengdong Zhao et Jeffrey Tzu Kwan Valino Koh**. *Understanding mid-air hand gestures : A study of human preferences in usage of gesture types for hci.* Microsoft Research TechReport, 2012. 21, 41

[Amano 04] **Sachiko Amano**, **Emiko Kezuka et Atsuko Yamamoto**. *Infant shifting attention from an adult's face to an adult's hand : A precursor of joint attention.* Infant Behavior and Development, vol. 27, n° 1, pages 64–80, 2004. 78

[Amari 77] **Shun'ichi Amari**. *Dynamics of pattern formation in lateral-inhibition type neural fields.* Biological cybernetics, vol. 27, n° 2, pages 77–87, 1977. viii, 84, 87, 91, 105

[Ammi 06] **Mehdi Ammi**, **Hamid Ladjal et Antoine Ferreira**. *Evaluation of 3D pseudo-haptic rendering using vision for cell micromanipula-*

tion. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 2115–2120. IEEE, 2006. 29

[Ansuini 08] **Caterina Ansuini**, **Livia Giosa**, **Luca Turella**, **Gianmarco Altoè et Umberto Castiello**. *An object for an action, the same object for other actions : effects on hand shaping*. Experimental Brain Research, vol. 185, n° 1, pages 111–119, 2008. 52, 53

[Ardito 14] **Carmelo Ardito**, **Maria Francesca Costabile et Hans-Christian Jetter**. *Gestures that people can understand and use*. Journal of Visual Languages & Computing, vol. 25, n° 5, pages 572–576, 2014. 21

[Becchio 10] **Cristina Becchio**, **Luisa Sartori et Umberto Castiello**. *Toward you the social side of actions*. Current Directions in Psychological Science, vol. 19, n° 3, pages 183–188, 2010. vi, 52

[Becchio 12] **Cristina Becchio**, **Valeria Manera**, **Luisa Sartori**, **Andrea Cavallo et Umberto Castiello**. *Grasping intentions : from thought experiments to empirical evidence*. Frontiers in human neuroscience, vol. 6, 2012. 51, 53, 59

[Benveniste 39] **Emile Benveniste**. *Nature du signe linguistique*. Acta linguistica, vol. 1, n° 1, pages 23–29, 1939. 10

[Binnig 86] **Gerd Binnig**, **Calvin F. Quate et Christoph Gerber**. *Atomic force microscope*. Physical review letters, vol. 56, n° 9, page 930, 1986. 28

[Bivall 11] **Petter Bivall**, **Shaaron Ainsworth et Lena A. E. Tibell**. *Do haptic representations help complex molecular learning ?* Science Education, vol. 95, n° 4, pages 700–719, 2011. 14

[Bolopion 10a] **Aude Bolopion**, **Barthelemy Cagneau**, **Stéphane Redon et Stéphane Régnier**. *Comparing position and force control for interactive molecular simulators with haptic feedback*. Journal of Molecular Graphics and Modelling, vol. 29, n° 2, pages 280–289, 2010. v, 9

[Bolopion 10b] **Aude Bolopion**, **Hui Xie**, **Sinan Haliyo et Stéphane Régnier**. *3D haptic handling of microspheres*. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 6131–6136, 2010. 17

[Bolopion 11] **Aude Bolopion**, **Christian Stolle**, **Robert Tunnell**, **Sinan Haliyo**, **Stéphane Régnier et Sergej Fatikow**. *Remote microscale teleoperation through virtual reality and haptic feedback*. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 894–900, 2011. 17, 18, 29

[Bolopion 12] **Aude Bolopion**, **Hui Xie**, **Sinan Haliyo et Stéphane Régnier**. *Haptic teleoperation for 3-D microassembly of spherical objects*. IEEE/ASME Transactions on Mechatronics, vol. 17, n° 1, pages 116–127, 2012. 8, 13

[Bolopion 13a] **Aude Bolopion**, **Guillaume Millet**, **Cécile Pacoret et Stéphane Régnier**. *Haptic Feedback in Teleoperation in Micro-and Nanoworlds*. Reviews of Human Factors and Ergonomics, vol. 9, n° 1, pages 57–93, 2013. 7

[Bolopion 13b] **Aude Bolopion et Stéphane Régnier**. *A review of haptic feedback teleoperation systems for micromanipulation and microassembly*. IEEE Transactions on Automation Science and Engineering, vol. 10, n° 3, pages 496–502, 2013. 29

[Borji 13] **Ali Borji et Laurent Itti**. *State-of-the-art in visual attention modeling*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, n° 1, pages 185–207, 2013. 24, 75

[Boukhnifer 06] **Moussa Boukhnifer et Antoine Ferreira**. *Stability and transparency for scaled teleoperation system*. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 4217–4222. IEEE, 2006. 29

[Bowman 97] **Doug A. Bowman et Larry F. Hodges**. *An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments*. In ACM Symposium on Interactive 3D graphics, pages 35–ff, 1997. 10

[Bowman 11] **Richard W. Bowman**, **Graham M. Gibson**, **David Carberry**, **Loren Picco**, **M. Miles et Miles J. Padgett**. *iTweezers : optical micromanipulation controlled by an Apple iPad*. Journal of Optics, vol. 13, n° 4, page 044002, 2011. 14, 15, 18

[Bowman 12] **Doug A. Bowman**, **Ryan P. McMahan et Eric D. Ragan**. *Questioning naturalism in 3D user interfaces*. Communications of the ACM, vol. 55, n° 9, pages 78–88, 2012. 10

[Brooke 96] **John Brooke**. *SUS-A quick and dirty usability scale*. Usability evaluation in industry, vol. 189, page 194, 1996. 44, 68

[Buryanov 10] **Alexander Buryanov et Viktor Kotiuk**. *Proportions of Hand Segments*. International Journal of Morphology, vol. 28, n° 3, pages 755–758, 2010. 32

[Cail 12] **François Cail**. Le travail sur écran en 50 questions. INRS, Institut national de recherche et de sécurité, 2012. 39

[Campanella 11] **Francesco Campanella**, **Giulio Sandini et Maria Concetta Morrone**. *Visual information gleaned by observing grasping mo-*

vement in allocentric and egocentric perspectives. Proceedings of the Royal Society B : Biological Sciences, vol. 278, n° 1715, pages 2142–2149, 2011. 54, 55

[Carrasco 10] **Miguel Carrasco et Xavier Clady**. *Prediction of user's grasping intentions based on eye-hand coordination.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 4631–4637, 2010. 22, 23

[Choumane 10] **Ali Choumane**, **Géry Casiez et Laurent Grisoni**. *Buttonless clicking : Intuitive select and pick-release through gesture analysis.* In IEEE Virtual Reality Conference (VR), pages 67–70, 2010. 22, 23

[Dimension] **Dimension**. *Bras Omega, http ://www.forcedimension.com.* v, 13

[Duncan 84] **John Duncan**. *Selective attention and the organization of visual information.* Journal of Experimental Psychology : General, vol. 113, n° 4, page 501, 1984. 74

[Fanelli 11] **Gabriele Fanelli**, **Thibaut Weise**, **Juergen Gall et Luc Van Gool**. *Real time head pose estimation from consumer depth cameras.* In Pattern Recognition, pages 101–110. Springer, 2011. 79

[Fatikow 07] **Sergej Fatikow**. Automated nanohandling by microrobots. Springer Science & Business Media, 2007. 6

[Ferreira 04] **Antoine Ferreira**, **Claude Cassier et Shigeoki Hirai**. *Automatic microassembly system assisted by vision servoing and virtual reality.* IEEE/ASME Transactions on Mechatronics, vol. 9, n° 2, pages 321–333, 2004. 29

[Ferreira 06] **Antoine Ferreira et Constantinos Mavroidis**. *Virtual reality and haptics for nanorobotics.* Robotics & Automation Magazine, IEEE, vol. 13, n° 3, page 78–92, 2006. 13

[Frischen 07] **Alexandra Frischen**, **Andrew P. Bayliss et Steven P. Tipper**. *Gaze cueing of attention : visual attention, social cognition, and individual differences.* Psychological bulletin, vol. 133, n° 4, page 694, 2007. 77

[Gao 07] **Dashan Gao et Nuno Vasconcelos**. *Bottom-up saliency is a discriminant process.* In IEEE International Conference on Computer Vision (ICCV), pages 1–6, 2007. vii, 75

[Gillian 14a] **Nicholas Gillian et Joseph A Paradiso**. *The gesture recognition toolkit.* The Journal of Machine Learning Research, vol. 15, n° 1, pages 3483–3487, 2014. 20

[Gillian 14b] **Nicholas Gillian**, **Sara Pfenninger**, **Spencer Russell et Joseph A. Paradiso**. *Gestures everywhere : a multimodal sensor*

fusion and analysis framework for pervasive displays. In ACM International Symposium on Pervasive Displays, page 98, 2014. v, 19, 20

[Grieve 09] **James A. Grieve**, **Arturas Ulcinas**, **Sriram Subramanian**, **Graham M. Gibson**, **Miles J. Padgett**, **David M. Carberry et Mervyn J. Miles**. *Hands-on with optical tweezers : a multitouch interface for holographic optical trapping.* Optics express, vol. 17, n° 5, pages 3595–3602, 2009. v, 14, 15, 18

[Haag 14] **Moritz P. Haag**, **Alain C. Vaucher**, **Maël Bosson**, **Stéphane Redon et Markus Reiher**. *Interactive chemical reactivity exploration.* Chemphyschem : a European journal of chemical physics and physical chemistry, vol. 15, n° 15, pages 3301–3319, 2014. 8

[Haliyo 04] **Sinan Haliyo**, **Fabien Dionnet et Stéphane Régnier**. *Controlled rolling of microobjects for autonomous manipulation.* Journal of Micromechatronics, vol. 3, n° 2, pages 75–102, 2004. 28

[Hannaford 89] **Blake Hannaford**. *A design framework for teleoperators with kinesthetic feedback.* IEEE Transactions on Robotics and Automation, vol. 5, n° 4, pages 426–434, 1989. 34

[Haption] **Haption**. *Bras Virtuose, http ://www.haption.com.* 13

[Harnad 90] **Stevan Harnad**. *The symbol grounding problem.* Physica D : Nonlinear Phenomena, vol. 42, n° 1, pages 335–346, 1990. 11

[Henderson 99] **John M. Henderson et Andrew Hollingworth**. *High-level scene perception.* Annual review of psychology, vol. 50, n° 1, pages 243–271, 1999. 75

[Hilliges 12] **Otmar Hilliges**, **David Kim**, **Shahram Izadi**, **Malte Weiss et Andrew Wilson**. *HoloDesk : direct 3d interactions with a situated see-through display.* In ACM SIGCHI Conference on Human Factors in Computing Systems, pages 2421–2430, 2012. v, 21

[Hoffman 06] **Matthew W. Hoffman**, **David B. Grimes**, **Aaron P. Shon et Rajesh P.N. Rao**. *A probabilistic model of gaze imitation and shared attention.* Neural Networks, vol. 19, n° 3, pages 299–310, 2006. 82

[Horvitz 03] **Eric Horvitz**, **Carl Kadie**, **Tim Paek et David Hovel**. *Models of attention in computing and communication : from principles to applications.* Communications of the ACM, vol. 46, n° 3, pages 52–59, 2003. 22, 23, 24

[Immersion] **Immersion**. *Bras Phantom, http ://www.immersion.fr.* v, 13

[Itti 01] **Laurent Itti et Christof Koch**. *Computational modelling of visual attention.* Nature reviews neuroscience, vol. 2, n° 3, pages 194–203, 2001. 76

[Kastner 04] **Sabine Kastner et Mark A. Pinsk**. *Visual attention as a multilevel selection process.* Cognitive, Affective, & Behavioral Neuroscience, vol. 4, n° 4, pages 483–500, 2004. 24

[Keysers 06] **Christian Keysers et Valeria Gazzola**. *Towards a unifying neural theory of social cognition.* Progress in brain research, vol. 156, pages 379–401, 2006. 62

[Koch 06] **Kristin Koch**, **Judith McLean**, **Ronen Segev**, **Michael A Freed**, **Michael J Berry**, **Vijay Balasubramanian et Peter Sterling**. *How much the eye tells the brain.* Current Biology, vol. 16, n° 14, pages 1428–1434, 2006. 74

[Langton 00] **Stephen R.H. Langton et Vicki Bruce**. *You must see the point : Automatic processing of cues to the direction of social attention.* Journal of Experimental Psychology : Human Perception and Performance, vol. 26, n° 2, page 747, 2000. 78

[LaViola 11] **Joseph J. LaViola et Daniel F. Keefe**. *3D spatial interaction : applications for art, design, and science.* In ACM SIGGRAPH Courses, page 1, 2011. 16

[Lv 13] **Zhihan Lv**, **Alex Tek**, **Franck Da Silva**, **Charly Empereur-Mot**, **Matthieu Chavent et Marc Baaden**. *Game on, science-how video game technology may help biologists tackle visualization challenges.* PloS one, vol. 8, n° 3, page e57990, 2013. 16

[MacKenzie 92] **I. Scott MacKenzie**. *Fitts' law as a research and design tool in human-computer interaction.* Human-computer interaction, vol. 7, n° 1, pages 91–139, 1992. 94

[Malizia 12] **Alessio Malizia et Andrea Bellucci**. *The artificiality of natural user interfaces.* Communications of the ACM, vol. 55, n° 3, pages 36–38, 2012. 10, 20

[Manera 10] **Valeria Manera**, **Ben Schouten**, **Cristina Becchio**, **Bruno G. Bara et Karl Verfaillie**. *Inferring intentions from biological motion : A stimulus set of point-light communicative interactions.* Behavior research methods, vol. 42, n° 1, pages 168–178, 2010. vi, 54, 55

[Manresa 05] **Cristina Manresa**, **Francisco J. Perales**, **Ramon Mas et Javier Varona**. *Hand tracking and gesture recognition for human-computer interaction.* In Electronic letters on computer vision and image analysis (ELCVIA), volume 5, pages 096–104, 2005. 19, 20

[McDonald 13] **Craig McDonald**, **Matthew McPherson**, **Craig McDougall et David McGloin**. *HoloHands : games console interface for controlling holographic optical manipulation.* Journal of Optics, vol. 15, n° 3, page 035708, 2013. 16, 17, 18

[Melnyk 14] **Artem Melnyk**. *Perfectionnement des algorithmes de contrôle-commande des robots manipulateur électriques en interaction physique avec leur environnement par une approche bio-inspirée.* Thèse de doctorat, Université de Cergy-Pontoise, Université nationale technique de Donetsk, 2014. 22

[Millet 08] **Guillaume Millet**, **Anatole Lécuyer**, **Jean-Marie Burkhardt**, **Sinan Haliyo et Stéphane Régnier**. *Improving perception and understanding of nanoscale phenomena using haptics and visual analogy.* In Haptics : Perception, Devices and Scenarios, pages 847–856. Springer, 2008. 18, 30

[Millet 13] **Guillaume Millet**, **Anatole Lécuyer**, **Jean-Marie Burkhardt**, **Sinan Haliyo et Stéphane Régnier**. *Haptics and graphic analogies for the understanding of atomic force microscopy.* International Journal of Human-Computer Studies, vol. 71, n° 5, pages 608–626, 2013. 12, 14, 29

[Mitra 07] **Sushmita Mitra et Tinku Acharya**. *Gesture recognition : A survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C : Applications and Reviews, vol. 37, n° 3, pages 311–324, 2007. 39

[Mohand-Ousaid 12] **Abdenbi Mohand-Ousaid**, **Guillaume Millet**, **Stéphane Régnier**, **Sinan Haliyo et Vincent Hayward**. *Haptic interface transparency achieved through viscous coupling.* The International Journal of Robotics Research, vol. 31, n° 3, pages 319–329, 2012. v, 13

[Mohand-Ousaid 14] **Abdenbi Mohand-Ousaid**, **Aude Bolopion**, **Sinan Haliyo**, **Stéphane Régnier et Vincent Hayward**. *Stability and transparency analysis of a teleoperation chain for microscale interaction.* In IEEE International Conference on Robotics and Automation (ICRA), pages 5946–5951, 2014. 6

[Nagasaki 89] **H. Nagasaki**. *Asymmetric velocity and acceleration profiles of human arm movements.* Experimental Brain Research, vol. 74, n° 2, pages 319–326, 1989. 55

[Nimbarte 08] **Ashish D. Nimbarte**, **Rodrigo Kaz et Zong-Ming Li**. *Finger joint motion generated by individual extrinsic muscles : A cadaveric study.* Journal of orthopaedic surgery and research, vol. 3, page 27, 2008. 33

[Norman 10] **Donald A. Norman**. *Natural user interfaces are not natural.* interactions, vol. 17, n° 3, pages 6–10, 2010. 20

[Nuku 08] **Pines Nuku et Harold Bekkering**. *Joint attention : Inferring what others perceive (and don't perceive).* Consciousness and cognition, vol. 17, n° 1, pages 339–349, 2008. 78

[Nummenmaa 09] **Lauri Nummenmaa et Andrew J. Calder**. *Neural mechanisms of social attention.* Trends in cognitive sciences, vol. 13, n° 3, pages 135–143, 2009. 78

[Oztop 05] **Erhan Oztop**, **Daniel Wolpert et Mitsuo Kawato**. *Mental state inference using visual control parameters.* Cognitive Brain Research, vol. 22, n° 2, pages 129–151, 2005. 3, 50, 62, 70

[Pacoret 13] **Cécile Pacoret et Stéphane Régnier**. *A review of haptic optical tweezers for an interactive microworld exploration.* Review of Scientific Instruments, vol. 84, n° 8, page 081301, 2013. 6

[Palacios 13] **José Manuel Palacios**, **Carlos Sagüés**, **Eduardo Montijano et Sergio Llorente**. *Human-Computer interaction based on hand gestures using RGB-D sensors.* Sensors, vol. 13, n° 9, pages 11842–11860, 2013. 19, 20

[Park 07] **In-Yong Park**, **Seung-Yong Sung**, **Jong-Hyun Lee et Yong-Gu Lee**. *Manufacturing micro-scale structures by an optical tweezers system controlled by five finger tips.* Journal of Micromechanics and Microengineering, vol. 17, n° 10, page N82, 2007. v, 15, 18

[Parks 14] **Daniel Parks**, **Ali Borji et Laurent Itti**. *Augmented saliency model using automatic 3D head pose detection and learned gaze following in natural scenes.* Vision research, 2014. 82

[Pavlovic 96] **Vladimir I. Pavlovic**, **Rajeev Sharma et Thomas S. Huang**. *Gestural interface to a visual computing environment for molecular biologists.* In IEEE International Conference on Automatic Face and Gesture Recognition, pages 30–35. IEEE, 1996. 16, 17, 18

[Poupyrev 96] **Ivan Poupyrev**, **Mark Billinghurst**, **Suzanne Weghorst et Tadao Ichikawa**. *The go-go interaction technique : non-linear mapping for direct manipulation in VR.* In ACM symposium on user interface software and technology, pages 79–80, 1996. 10

[Poupyrev 98] **Ivan Poupyrev**, **T. Ichikawa**, **S. Weghorst et M. Billinghurst**. *Egocentric object manipulation in virtual environments : empirical evaluation of interaction techniques.* In Computer Graphics Forum, volume 17, pages 41–52. Wiley Online Library, 1998. 22

[Probst 07] **Martin Probst**, **Christoph Hürzeler**, **Ruedi Borer et Bradley J. Nelson**. *Virtual reality for microassembly.* In Internatio-

nal Symposium on Optomechatronic Technologies, pages 67180D–67180D. International Society for Optics and Photonics, 2007. 12

[Ren 13] **Gang Ren et Eamonn O'Neill**. *3D selection with freehand gesture*. Computers & Graphics, vol. 37, n° 3, pages 101–120, 2013. 10, 40

[Rizzolatti 96] **Giacomo Rizzolatti**, **Luciano Fadiga**, **Vittorio Gallese et Leonardo Fogassi**. *Premotor cortex and the recognition of motor actions*. Cognitive brain research, vol. 3, n° 2, pages 131–141, 1996. 62

[Rougier 06] **Nicolas P. Rougier et Julien Vitay**. *Emergence of attention within a neural population*. Neural Networks, vol. 19, n° 5, pages 573–581, 2006. 83

[Sartori 09] **Luisa Sartori**, **Cristina Becchio**, **Bruno G. Bara et Umberto Castiello**. *Does the intention to communicate affect action kinematics ?* Consciousness and cognition, vol. 18, n° 3, pages 766–772, 2009. 53

[Sartori 11] **Luisa Sartori**, **Cristina Becchio et Umberto Castiello**. *Cues to intention : the role of movement information*. Cognition, vol. 119, n° 2, pages 242–252, 2011. 53, 55, 59

[Sauvet 12] **Bruno Sauvet**, **Nizar Ouarti**, **Sinan Haliyo et Stéphane Régnier**. *Virtual reality backend for operator controlled nanomanipulation*. In IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO), pages 121–127, 2012. v, 8, 18, 29

[Schauerte 14] **Boris Schauerte et Rainer Stiefelhagen**. *"Look at this !" learning to guide visual saliency in human-robot interaction*. In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 995–1002, 2014. 82

[Schiavo 13] **Gianluca Schiavo**, **Eleonora Mencarini**, **Kevin Vovard et Massimo Zancanaro**. *Sensing and reacting to users' interest : an adaptive public display*. In ACM CHI'13 Extended Abstracts on Human Factors in Computing Systems, pages 1545–1550, 2013. 24

[Searle 82] **John R. Searle**. *The Chinese room revisited*. Behavioral and Brain Sciences, vol. 5, n° 02, pages 345–348, 1982. 11

[Searle 83] **John R. Searle**. Intentionality : An essay in the philosophy of mind. Cambridge University Press, 1983. 50

[Shotton 13] **Jamie Shotton**, **Toby Sharp**, **Alex Kipman**, **Andrew Fitzgibbon**, **Mark Finocchio**, **Andrew Blake**, **Mat Cook et Richard Moore**. *Real-time human pose recognition in parts from single depth*

images. Communications of the ACM, vol. 56, n° 1, pages 116–124, 2013. 42

[Sitti 98] **Metin Sitti**, **Satoshi Horiguchi et Hideki Hashimoto**. *Nano tele-manipulation using virtual reality interface*. In IEEE International Symposium on Industrial Electronics (ISIE), volume 1, pages 171–176, 1998. 12

[Stapel 12] **Janny C. Stapel**, **Sabine Hunnius et Harold Bekkering**. *Online prediction of others' actions : the contribution of the target object, action context and movement kinematics*. Psychological research, vol. 76, n° 4, pages 434–445, 2012. 54, 55, 64

[Stefanov 10] **Nikolay Stefanov**, **Angelika Peer et Martin Buss**. *Online intention recognition in computer-assisted teleoperation systems*. In Haptics : Generating and Perceiving Tangible Sensations, pages 233–239. Springer, 2010. 22, 23

[Sulzmann 95] **Armin Sulzmann et Jacques Jacot**. *3D computer graphics based interface to real microscopic worlds for microrobot telemanipulation and position control*. In IEEE International Conference on Systems, Man and Cybernetics, volume 1, page 286–291, 1995. 12

[Treisman 80] **Anne M. Treisman et Garry Gelade**. *A feature-integration theory of attention*. Cognitive psychology, vol. 12, n° 1, pages 97–136, 1980. 75

[Van den Bergh 11] **Michael Van den Bergh et Luc Van Gool**. *Combining RGB and ToF cameras for real-time 3D hand gesture interaction*. In IEEE Workshop on Applications of Computer Vision (WACV), pages 66–72. IEEE, 2011. v, 19, 20, 21

[Venture 06] **Gentiane Venture**, **Sinan Haliyo**, **Alain Micaelli et Stéphane Régnier**. *Force-feedback coupling for micro-handling applications*. International Journal of Micromechatronics, Special Issue on Microhandling, vol. 3, page 3–4, 2006. 29

[Viviani 95] **Paolo Viviani et Tamar Flash**. *Minimum-jerk, two-thirds power law, and isochrony : converging approaches to movement planning*. Journal of Experimental Psychology : Human Perception and Performance, vol. 21, n° 1, page 32, 1995. 56

[Vogel 05] **Daniel Vogel et Ravin Balakrishnan**. *Distant freehand pointing and clicking on very large, high resolution displays*. In ACM Symposium on User interface software and technology, pages 33–42, 2005. 22

[Whyte 06] **Graeme Whyte**, **Graham Gibson**, **Jonathan Leach**, **Miles Padgett**, **Daniel Robert et Mervyn Miles**. *An optical trapped*

microhand for manipulating micron-sized objects. Optics express, vol. 14, n° 25, pages 12497–12502, 2006. 16, 18

[Yarbus 67] **Basil Yarbus Alfred L.and Haigh et Lorrin A. Rigss**. Eye movements and vision, volume 2. Plenum press New York, 1967. vii, 76

[Yucel 13] **Zeynep Yucel**, **Albert Ali Salah**, **Çetin Meriçli**, **Tekin Meriçli**, **Roberto Valenti et Theo Gevers**. *Joint attention by gaze interpolation and saliency*. IEEE Transactions on Cybernetics, vol. 43, n° 3, pages 829–842, 2013. 82

[Zafrulla 11] **Zahoor Zafrulla**, **Helene Brashear**, **Thad Starner**, **Harley Hamilton et Peter Presti**. *American sign language recognition with the kinect*. In ACM International conference on multimodal interfaces, pages 279–286, 2011. 40

[Zeller 97] **Michael Zeller**, **James C. Phillips**, **Andrew Dalke**, **William Humphrey**, **Klaus Schulten**, **Rajeev Sharma**, **T.S. Huang**, **VI Pavlovic**, **Y. Zhao et Z. Lo**. *A visual computing environment for very large scale biomolecular modeling*. In IEEE International Conference on Application-Specific Systems, Architectures and Processors, pages 3–12, 1997. vi, 16, 40

Liste des publications

Communication avec actes (conférence internationale) :

1. **L. Cohen**, S. Haliyo, M. Chetouani and S. Régnier
   Intention prediction approach to interact naturally with the microworld.
   IEEE/ASME AIM, 2014. Pages 396-401.

2. **L. Cohen**, S. Haliyo, M. Chetouani and S. Régnier
   Metaphor-free interaction for micromanipulation.
   IEEE/ASME AIM workshop : Merging micro and macro manipulation and manufacturing technologies and methods, 2014. À paraître.

3. **L. Cohen**, W. Abbassi, M. Chetouani and S. Boucenna
   Intention inference development in a robot through the interaction with a caregiver.
   IEEE ICDL-EPIROB, 2014. À paraître.