HAL
open science

Interaction à distance en environnement physique augmenté

William Delamare

To cite this version:

William Delamare. Interaction à distance en environnement physique augmenté. Interface homme-machine [cs.HC]. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAM032 . tel-01270557

HAL Id: tel-01270557

https://theses.hal.science/tel-01270557v1

Submitted on 8 Feb 2016

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques, Sciences et technologies de l'information, Informatique**

Arrêté ministériel : 7 août 2006

Présentée par

William Delamare

Thèse dirigée par **Laurence NIGAY** et
codirigée par **Céline COUTRIX**

préparée au sein du **Laboratoire d'Informatique de Grenoble (LIG), équipe Ingénierie de l'Interaction Homme-Machine (IIHM)**
dans **l'École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique (ED-MSTII)**

Interaction à distance en environnement physique augmenté

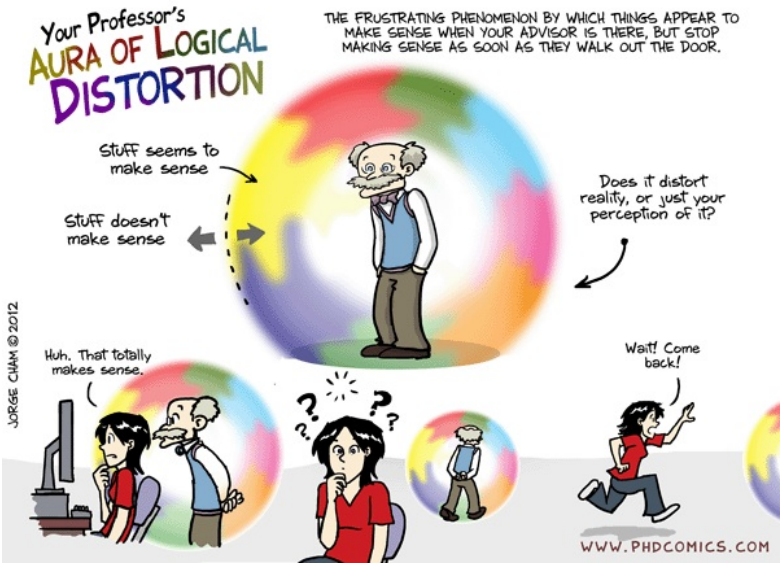
Thèse soutenue publiquement le **2 novembre 2015**,
devant le jury composé de :

M. Michel BEAUDOUIN-LAFON
Professeur Université Paris-Sud, Président
M. Géry CASIEZ
Professeur Université Lille 1, Rapporteur
M. Martin HACHET
Chargé de Recherche Inria Bordeaux, Rapporteur
M. Kris LUYTEN
Professor Hasselt University, Examinateur
M. Radu-Daniel VATAVU
Associate Professor University Stefan cel Mare of Suceava, Examinateur
Mme. Laurence NIGAY
Professeur Université Joseph Fourier, Directrice de thèse
Mme. Céline COUTRIX
Chargée de Recherche CNRS, Co-directrice de thèse

Remerciements

Je remercie avant tout les membres du jury qui ont accepté d'évaluer mes travaux de thèse et qui se sont déplacés parfois de loin pour ma soutenance.

Ensuite, je remercie l'équipe IIHM ainsi que l'équipe enseignante de l'IUT2 de Grenoble. Ces deux expériences m'ont permis de découvrir ce qu'est la vie d'un enseignant-chercheur en France.
Je remercie plus particulièrement mes directrices de thèse : Laurence NIGAY et Céline COUTRIX. Elle ont accepté de me prendre au sein de l'équipe alors que je n'avais aucune connaissance quant à la nature d'un travail de recherche. Travailler auprès de Laurence et Céline a réellement été enrichissant. Je pense avoir désormais dépassé le stade dépeint par l'image ci-dessous, et ceci grâce à leur encadrement, leur manière de travailler ainsi que leurs conseils durant ces quatre années de thèse.
Encore merci, et j'espère que nous aurons l'opportunité de retravailler ensemble dans le futur. Après tout, c'est grâce à vous que je veux continuer dans cette voie !



Je remercie aussi mes (*faux !*) potes de Grenoble ! Je ne peux évidemment pas détailler toutes les raisons qui me poussent à les inclure dans ces remerciements, mais il va de soi qu'un bon groupe d'amis ne peut qu'être une bonne chose ! Ils m'ont permis de m'évader pour passer de bons moments. Même si le monde du travail s'impose souvent dans notre vie personnelle et que l'ordinateur du laboratoire a autant voyagé que moi, les bons moments passés sur Grenoble (e.g., au Fams) et ailleurs ont indubitablement contribué à améliorer cette expérience de vie qu'a été la thèse.

Pour finir, je remercie ma famille, et plus particulièrement mon fils, Shin, et sa patience, puisqu'il a entendu des centaines de fois le couplet "il faut que je travaille", l'amenant ainsi à connaître mon discours de soutenance quasiment aussi bien que moi !

Table des matières

Liste des figures **vii**

Liste des tableaux **xv**

1 Introduction **1**

2 Interaction gestuelle pour l'intelligence ambiante **5**
2.1 Intelligence ambiante : contexte des travaux 5
2.1.1 Domaine applicatif : sources lumineuses 5
2.1.2 Généralisation à l'intelligence ambiante 6
2.2 Interaction gestuelle : sujet des travaux 8
2.2.1 Caractérisation articulatoire du geste 9
2.2.1.1 Exécution d'un geste avec une partie du corps 9
2.2.1.2 Exécution d'un geste avec ou sans contact physique . . . 10
2.2.2 Caractérisation fonctionnelle du geste 11

I Sélection d'un objet physique
Pointage par des gestes 3D : Nécessité de désambigüiser **15**

3 Pointer dans le monde physique **19**
3.1 Différentes méthodes de sélection dans le monde physique 19
3.2 Limites du pointage dans le monde physique 20
3.3 Le pointage et la désambigüisation 22

4 Espace de Conception **27**
4.1 Introduction 27
4.2 Espace de Conception 29
4.2.1 Groupe *Interaction* 29
4.2.1.1 Espace de représentation 29
4.2.1.2 Espace de contrôle 31
4.2.2 Groupe *Système de Désambigüisation* 32
4.2.2.1 *Obtention des connaissances* 32
4.2.2.2 Sous-groupe *Quoi* 33

4.2.2.3 Sous-groupe *Où* 34
4.2.3 Liens entre les groupes *Interaction* et *Système de Désambigüisation* 35
4.2.4 Positionnement : espaces de conception existants 36
4.2.4.1 Désambigüisation en environnements numériques 36
4.2.4.2 Interaction dans le monde physique 38
4.2.4.3 Relation entre les mondes physique et numérique 38
4.3 Evaluation de l'espace de conception 39
4.4 Conclusion 45

5 Techniques de désambigüisation proposées : P2Roll et P2Slide 47
5.1 Introduction 48
5.1.1 Rechercher des chemins inexplorés dans notre espace de conception 48
5.1.2 Propriétés des actions 48
5.1.3 Désambigüisation et loi de Fitts 50
5.2 Conception de P2Roll et P2Slide 51
5.2.1 Quasi-mode 51
5.2.2 Physical Pointing Roll : P2Roll 52
5.2.3 Physical Pointing Slide : P2Slide 54
5.3 Evaluation expérimentale 56
5.3.1 Technique de référence : liste 56
5.3.2 Hypothèses 57
5.3.3 Equipement 58
5.3.4 Participants 59
5.3.5 Procédure 59
5.3.6 Protocole expérimental 60
5.3.7 Résultats 61
5.3.7.1 H1a : Garder l'attention de l'utilisateur dans le monde physique est plus efficace que de forcer l'utilisateur à porter son attention sur une représentation numérique . . 62
5.3.7.2 H1b : Les utilisateurs préfèrent garder leur attention dans le monde physique plutôt que de porter leur attention sur une représentation numérique 66
5.3.7.3 H2 : Les petits mouvements ne sont pas forcément plus efficaces 66
5.4 Conclusion sur l'évaluation expérimentale 67
5.5 Travaux futurs à partir de l'évaluation expérimentale 68

6 Conclusion de la partie sélection 71
6.1 Conclusion sur la désambigüisation 71
6.2 Travaux futurs sur la désambigüisation 72

II Contrôle d'un objet physique
Spécification de commandes par des gestes 3D : nécessité de guider 75

7 Introduction 79

8 Espace de Conception 83
8.1 Espace de Conception 85
8.1.1 Structure générale 85
8.1.1.1 Catégories d'axes : retours d'information rétroactif et prédictif 85
8.1.1.2 Groupes d'axes *Quand*, *Quoi*, *Comment* et *Où* 86
8.1.2 Axes 87
8.1.2.1 Groupe d'axes : *Quand* 87
8.1.2.1.1 Déclenchement du guidage 88
8.1.2.1.2 Exécution du geste 92
8.1.2.1.3 Mode de guidage 93
8.1.2.2 Groupe d'axes : *Quoi* 94
8.1.2.2.1 Geste(s) affiché(s) 94
8.1.2.2.2 Valeur de performance pour le retour d'information rétroactif 98
8.1.2.2.3 Impact du retour d'information rétroactif sur le retour d'information prédictif 98
8.1.2.3 Groupe d'axes : *Comment* 99
8.1.2.4 Groupe d'axes : *Où* 102
8.1.3 Positionnement : espaces de conception 103
8.1.3.1 Retour d'information rétroactif 104
8.1.3.2 Retour d'information prédictif 105
8.1.4 Positionnement : techniques de guidage 106
8.2 Evaluation 108
8.2.1 Outil interactif pour faciliter l'utilisation des pouvoirs de notre espace de conception 109
8.2.1.1 Fonction : description détaillée d'une technique de guidage 109
8.2.1.2 Fonction : comparaison entre les techniques de guidage . 111
8.2.1.3 Fonction : rechercher des zones inexplorées dans l'espace de conception 113
8.2.1.4 Fonction : concevoir une nouvelle technique de guidage . 114
8.2.1.5 Fonctions secondaires liées à la base de connaissance de l'outil 115
8.2.2 Scénarios d'utilisation 115
8.2.2.1 Trouver une solution d'interaction 115
8.2.2.2 Trouver un problème de recherche 117
8.3 Conclusion 118

9 Système de Guidage : OctoPocus3D **121**
9.1 Introduction 122
9.2 Choix de conception 124
9.2.1 Recherche d'une solution grâce à notre espace de conception 124
9.2.2 Description d'OctoPocus 124
9.2.2.1 Retour d'information rétroactif d'OctoPocus et d'OctoPocus3D 125
9.2.2.1.1 Groupe *Quand* 125
9.2.2.1.2 Groupe *Quoi* 125
9.2.2.1.3 Groupe *Comment* 126
9.2.2.1.4 Groupe *Où* 126
9.2.2.2 Retour d'information prédictif 127
9.2.2.2.1 Groupe *Quand* 127
9.2.2.2.2 Groupe *Quoi* 127
9.2.2.2.3 Groupe *Comment* 128
9.2.2.2.4 Groupe *Où* 128
9.2.2.3 Vue d'ensemble 129
9.2.3 Alternatives de conception : exemples 129
9.2.3.1 Fréquence de mise à jour du retour d'information rétroactif 130
9.2.3.2 Position du guide définie par le retour d'information prédictif 132
9.2.3.3 Portion de gestes affichée par le retour d'information prédictif 134
9.3 Conception d'OctoPocus3D 135
9.3.1 Affichage des gestes 3D 135
9.3.1.1 Solutions retenues pour la version finale d'OctoPocus3D . 140
9.3.1.1.1 Occultation et transparence 140
9.3.1.1.2 Perspective linéaire 141
9.3.1.1.3 Éclairage et nuances d'ombres 141
9.3.1.1.4 Profondeur cinétique et couplage tête - affichage 141
9.3.1.2 Solutions non retenues pour la version finale d'OctoPocus3D 142
9.3.1.2.1 Ajouts d'éléments visuels 142
9.3.1.2.2 Exploitation des propriétés graphiques des tubes 143
9.3.2 Retour d'information rétroactif : terminal et/ou continu 144
9.3.3 Retour d'information prédictif : origine du guide et portions des gestes affichées 145
9.4 Évaluations expérimentales 148
9.4.1 Ressources 149
9.4.1.1 Conception des jeux de gestes pour les évaluations 149
9.4.1.2 Équipement et configuration 149
9.4.1.3 Etude préliminaire : OctoPocus3D *vs* vidéos de démonstration 150
9.4.2 Expérience 1 : retour d'information rétroactif 151
9.4.2.1 Participants et plan d'expérience 151

9.4.2.2 Résultats . . . 153
9.4.2.2.1 Taux de reconnaissance . . . 153
9.4.2.2.2 Temps d'exécution . . . 155
9.4.2.2.3 Discussion . . . 155
9.4.3 Expérience 2 : retour d'information prédictif . . . 156
9.4.3.1 Participants et plan d'expérience . . . 156
9.4.3.2 Résultats . . . 158
9.4.3.2.1 Taux de reconnaissance . . . 158
9.4.3.2.2 Temps d'exécution . . . 158
9.4.3.2.3 Discussion . . . 159
9.5 Déploiement *in situ* . . . 160
9.6 Conclusion . . . 161
9.7 Travaux futurs . . . 162

10 Conclusion de la partie contrôle **167**
10.1 Conclusion sur le guidage . . . 167
10.2 Travaux futurs sur le guidage . . . 168

11 Conclusion **171**
11.1 Résumé des contributions . . . 171
11.1.1 Sélection d'objets physiques augmentés . . . 172
11.1.2 Contrôle d'objets physiques augmentés . . . 173
11.2 Perspectives envisagées des travaux de thèse . . . 174

Bibliographie **179**

Liste des figures

1.1 Méthode de travail et contributions. Nous avons considéré l'interaction en deux étapes (ou sous-tâches) : la sélection et le contrôle des objets physiques augmentés. Pour chacune de ces étapes, nous apportons des résultats conceptuels (établissement d'un espace de conception) ainsi que des résultats pratiques (conception, développement et évaluation de techniques). 2
1.2 Structure du manuscrit. 3

2.1 Exemple d'application du projet DELight dans le bureau : les matériaux lumineux (innovants et traditionnels) sont contrôlés à distance par la BOX dans un environnement de bureaux [1]. La BOX est un point de convergence des différents équipements tels que les capteurs, les actionneurs ou bien les interfaces homme-machine (IHM). 7
2.2 Les différentes étapes de l'interaction gestuelle : l'initialisation du geste, sa continuation et sa terminaison. 10
2.3 Vue d'ensemble des types de gestes à partir de trois classifications existantes [Cadoz, 1994] [Karam et Schraefel, 2005] [Aigner *et al.*, 2012] et positionnement des deux parties principales du manuscrit. 12

3.1 Pointage basé sur une sélection volumique. Un ensemble d'objets physiques inclus dans le volume (c'est-à-dire le cône) est sélectionné. 22
3.2 En commençant par un geste de pointage dans le monde physique, trois chemins sont possibles pour sélectionner un objet. ($t1^*.t2.t3^*.t4$) : geste de pointage suivi par un processus de désambigüisation interactif. ($t1^*.t5$) : geste de pointage seulement. ($t1^*.t6.t7 = t1^*.t6$) : geste de pointage accompagné d'une heuristique afin de sélectionner automatiquement un objet. 23

4.1 Espace de conception avec 11 axes organisés selon deux groupes : *Interaction* et *Système de Désambigüisation*. Ce dernier est lui-même composé de deux sous-groupes reliés à deux questions : *Quoi* et *Où*. 30
4.2 Axe *Etendue des Connaissances sur le Monde* de [Milgram et Kishino, 1994]. 39
4.3 Pouvoir descriptif de notre espace de conception. Les techniques de désambigüisation existantes sont représentées par des chemins. 41

5.1 Exploration de notre espace de conception. Axe 1 : création de technique de désambigüisation laissant l'attention de l'utilisateur dans le monde physique. Axe 2 : exploration de deux options non évaluées. 49
5.2 Automates de nos techniques illustrant le comportement attendu d'utilisateurs novices. 51
5.3 Illustration de P2Roll. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. L'élément courant est l'objet 2 puisque l'angle du roulis est dans sa zone. (c) L'utilisateur effectue un roulis vers la gauche. L'objet 3 devient alors l'élément courant. (d) L'utilisateur effectue un roulis vers l'extrémité droite de sorte que l'élément courant passe de 3 à 1 en passant par 2. 53
5.4 Automates de nos techniques illustrant le comportement attendu d'utilisateurs experts. 54
5.5 Illustration de P2Slide. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. L'élément courant est l'objet 2 puisque le pouce est dans sa zone. (c) L'utilisateur effectue un glissement du pouce vers la gauche. L'objet 3 devient alors l'élément courant. (d) L'utilisateur effectue un glissement du pouce vers l'extrémité droite de sorte que l'élément courant passe de 3 à 1 en passant par 2. . . 55
5.6 Illustration de la technique de référence Liste. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. A ce stade, le pointage est figé. (c) L'apparition de la liste sur l'écran du téléphone. 56
5.7 Bouton physique *retour* par défaut du téléphone pour annuler l'étape de pointage avec la liste. 57
5.8 Plateau avec 16 objets physiques (balles de ping-pong) intégrant des LEDs. La numéro 21 (rouge) représente la cible de l'essai en cours, la numéro 8 (bleue) représente l'élément sélectionné courant. 58
5.9 Configuration de l'expérience. 60
5.10 Temps de sélection (en secondes). Les barres d'erreur représentent les intervalles de confiance à 95%. 62
5.11 Taux d'erreurs (en pourcentage) pour chaque technique en fonction de la densité. Les barres d'erreurs indiquent les intervalles de confiance à 95%. . 63
5.12 Taux d'erreurs (en pourcentage) pour P2Roll en fonction de la densité et du type d'erreur. 64
5.13 Taux d'erreurs (en pourcentage) pour P2Slide en fonction de la densité et du type d'erreur. 64
5.14 Illustration de rapMenu de [Ni *et al.*, 2011]. Les quatre gestes de pincement (de l'index au petit doigt) sélectionnent les éléments 1-4, 5-8 et 9-12 lorsque l'angle du roulis du poignet se trouve dans une de ces trois zones. 65

7.1 Différents niveaux d'abstraction pour étudier l'interaction gestuelle. 80

8.1 Distinction entre retour d'information rétroactif et retour d'information prédictif pour une technique de guidage de gestes : affichage d'un retour d'information rétroactif (i.e. gestes effectué/prévu/reconnu dans le passé) et d'un retour d'information prédictif (i.e. gestes disponibles dans le futur) pendant que l'utilisateur exécute un geste. 85
8.2 Exemple illustrant les concepts d'affordances perçues et de retour d'information prédictif selon [Vermeulen *et al.*, 2013]. 87
8.3 Aperçu de l'espace de conception avec 35 axes organisés selon deux catégories - retour d'information rétroactif et retour d'information prédictif. Chaque catégorie est composée de quatre groupes - Quand (caractéristiques temporelles), Quoi (caractéristiques du contenu), Comment (caractéristiques de la modalité) et Où (caractéristiques spatiales). Chaque axe possède des options de conception ainsi qu'une option "aucun". 88
8.4 Catégorie : retour d'information rétroactif. 89
8.5 Catégorie : retour d'information prédictif. 90
8.6 Groupes *Quand* du retour d'information rétroactif (haut) et du retour d'information prédictif (bas). 91
8.7 Distinction entre les étapes de l'interaction gestuelle (initialisation, continuation et terminaison d'un geste) et les étapes du guidage (enclenchement du guide, comportement pendant l'éxecution et sortie du mode de guidage). 92
8.8 Groupes *Quoi* du retour d'information rétroactif (haut) et du retour d'information prédictif (bas). Le retour d'information rétroactif fournit de l'information quant aux actions précédemment effectuées (A) ainsi qu'au traitement de ces actions par le système (B). Le retour d'information prédictif fournit de l'information sur les actions futures (C). Ces sous-groupes ont trois axes similaires concernant les gestes affichés (1), les caractéristiques impliquées (2) et les portions affichées (3). 95
8.9 Illustration de l'axe *Portion perçue*. Ligne du haut : le retour d'information rétroactif représente la portion de geste effectué complète (gauche) ou une portion partielle (droite). Ligne du bas : le retour d'information prédictif représente la portion restante du geste à effectuer (gauche) ou seulement la partie imminente (droite). 98
8.10 Illustration de l'axe *Présentation - Correction*. (a) Etat initial. (b) Projection : le geste représenté comprend un segment supplémentaire pour ramener l'utilisateur sur le point le plus proche du geste à effectuer. (c) Correction lisse : le geste représenté comprend un segment supplémentaire pour progressivement ramener l'utilisateur sur le geste initial. (d) Re-localisation : le geste est représenté selon la position actuelle de l'utilisateur. (e) Maximisation du score : la combinaison du geste représenté et de la partie du geste déjà effectuée maximise le score de reconnaissance du geste. 100

8.11 Groupe *Comment* commun au retour d'information rétroactif et au retour d'information prédictif. 100
8.12 Illustrations [2]de l'axe *Perspective* : parallèle avec un jeu vidéo de courses. (a) Perspective égocentrique. (b) Perspective couplée. (c) Perspective exocentrique. Ces perspectives sont aussi présentes lorsqu'une technique de guidage affiche des informations visuelles. 101
8.13 Groupe *Où* commun au retour d'information rétroactif et au retour d'information prédictif. 103
8.14 Onglets de notre outil en ligne. Ces onglets mènent à différentes fonctionnalités proposées afin de faciliter l'usage de notre espace de conception : une présentation de ce dernier (Overview), une description des techniques de guidage existantes (One by One), une comparaison de ces techniques (Compare), une exploration interactive des options utilisées (Interactive Design Space) et la création de nouveaux guides (Create), une bibliographie ainsi qu'une page de contact. 109
8.15 Description de toutes les techniques de guidage une par une (A). Tout d'abord, l'utilisateur choisit une technique dans une liste de tous les guides inclus dans notre outil (B). L'outil affiche ensuite la référence de la technique choisie (C1) ainsi que des liens vers la description de chaque étape composant le guide (C2). Enfin, l'outil affiche pour chaque étape les options utilisées en les colorant en vert dans l'espace de conception (A et D). Certaines options indiquent via un astérisque la possibilité d'accéder à une info-bulle affichant les raisons menant au choix de cette option selon notre analyse (D). 110
8.16 Comparaison globale (A). Sélection des techniques à comparer (B). Tableau comparatif présentant dans chaque cellule trois scores de similarités - retour d'information rétroactif, retour d'information prédictif et global (C). Une info-bulle rappelle les techniques correspondant à la cellule survolée par le curseur (C). 111
8.17 Comparaison deux à deux (A). Résumé de la comparaison (B). Coloration de l'espace de conception en fonction des deux techniques impliquées (C). Les options partagées par les deux techniques sont colorées avec un dégradé entre les deux couleurs (C). 112
8.18 Espace de conception interactif (A). Dans cet exemple, l'état courant contient un filtre (B). L'outil n'affiche donc que les quatre guides implémentant cette option (C). L'utilisateur peut alors visualiser combien de techniques implémentent les options composant l'espace de conception (D). 113

9.1 Utilisateur effectuant un geste avec notre prototype à l'entrée d'un restaurant. Le guide affiche les gestes à effectuer sous forme de tubes 3D sur un écran distant placé en face de l'utilisateur. Coin haut-droit : zoom de l'écran. 123
9.2 Groupe d'axes *Quand* du retour d'information rétroactif d'OctoPocus. . . 125

9.3 Groupe d'axes *Quoi* du retour d'information rétroactif d'OctoPocus. . . . 126
9.4 Groupe d'axes *Comment* du retour d'information rétroactif d'OctoPocus. . 126
9.5 Groupe d'axes *Où* du retour d'information rétroactif d'OctoPocus. 127
9.6 Groupe d'axes *Où* du retour d'information prédictif d'OctoPocus. 127
9.7 Groupe d'axes *Quoi* du retour d'information prédictif d'OctoPocus. 128
9.8 Groupe d'axes *Comment* du retour d'information prédictif d'OctoPocus. . 129
9.9 Groupe d'axes *Où* du retour d'information prédictif d'OctoPocus. 129
9.10 Résumé des options de conception explorées lors de l'adaptation d'OctoPocus aux gestes 3D. Pour le retour d'information rétroactif, comparaison entre une stratégie d'apprentissage consistant à utiliser une mise à jour discrète (A) ou continue (B). Pour le retour d'information prédictif, exploration d'options impactant la complexité visuelle de la scène 3D selon la portion de geste affichée (C et D) ou la stabilité de la représentation visuelle selon la correction de la présentation (E et F). 130
9.11 Illustrations de différents retours d'information rétroactifs terminaux. (a) Flèche sur une échelle de couleur. (b) Rediffusion des mouvements de l'enseignant et de l'utilisateur. Les deux techniques viennent de [Kamal *et al.*, 2014]. 131
9.12 Version finale d'OctoPocus3D représentant les gestes avec des tubes 3D. . 141
9.13 Ajout de cubes répartis sur les tubes afin d'ajouter des indices de perspective, de l'occultation locale, des gradients de taille, et des nuances d'ombres. 143
9.14 Ajout de plans de coupe aux tubes afin de visualiser des gradients de taille et ajouter un indice visuel quant à l'orientation du geste. 144
9.15 Représentation du jeu de geste (a) avec des tubes droits entourés de spirales, (b) avec seulement des spirales. 145
9.16 Illustration du retour d'information rétroactif concurrent. Le mécanisme de retour d'information prédictif affiche la portion restante complète des gestes à effectuer à partir de la représentation graphique de la main. (a) Scène d'origine. (b) L'utilisateur commence à suivre un geste vers le haut-gauche de l'écran. Le rayon des gestes sur la droite de la scène ayant peu de probabilité d'être reconnus a diminué (retour d'information rétroactif concurrent). (c) Les gestes avec un score de reconnaissance sous un seuil limite ont disparu (retour d'information rétroactif concurrent). 146
9.17 Illustration du retour d'information rétroactif terminal et de la profondeur cinétique. Le geste à effectuer est le geste bleu. Le geste exécuté par l'utilisateur est en vert. (a) A partir du point de vue d'origine. (b) A partir d'une perspective inclinée sur la droite. Nous rappelons que ce retour d'information rétroactif terminal n'est possible que si le système connait le geste que l'utilisateur souhaite réaliser, comme en situation d'entraînement. 146
9.18 Représentation du jeu de geste numéro 3. (a) Portions restantes complètes à effectuer des gestes. (b) Portions imminentes à venir des gestes. 147

9.19 Résumé des deux expériences au regard des options de conception évaluées. 148
9.20 Les quatre jeux de gestes utilisés dans nos expériences. Afin de présenter l'ensemble des gestes, les jeux de gestes sont ici illustrés avec un retour d'information prédictif affichant les portions complètes des gestes. 150
9.21 Dispositif expérimental situé en contexte à l'entrée d'un restaurant. 151
9.22 Protocole d'expérimentation pour évaluer les options de retour d'information rétroactif. Quatre blocs (un pour chaque technique avec un jeu de gestes dédié), chacun composé de 6 répétitions elles-mêmes composées de 8 exécutions de gestes. 152
9.23 Taux de reconnaissance (en pourcentage) pour chaque combinaison de retour d'information rétroactif durant les 6 répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%. 153
9.24 Taux de reconnaissance (en pourcentage) pour chaque combinaison de retour d'information rétroactif agrégés sur les deux premières répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%. 154
9.25 Temps d'exécution (en secondes) pour chaque combinaison de retours d'information rétroactif sur les 6 répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%. 155
9.26 Classement qualitatif de chaque combinaison de retours d'information rétroactif (la technique préférée est classée première). Les barres d'erreur indiquent les intervalles de confiance à 95%. 156
9.27 Protocole d'expérimentation pour évaluer les options de retour d'information prédictif. 157
9.28 Taux de reconnaissance (en pourcentage) obtenus en fonction des portions affichées et de la correction de l'origine du guide. Les barres d'erreur représentent les intervalles de confiance à 95%. 158
9.29 Temps d'exécution (en secondes) obtenus en fonction des portions affichées et de la correction de l'origine du guide. Les barres d'erreur représentent les intervalles de confiance à 95%. 159
9.30 Déploiement *in situ* à l'hôtel Royal de Lyon (a) vue de l'extérieur, (b) vue de l'intérieur. Un utilisateur exécute un geste pour contrôler la lumière de la chambre. La main sur la hanche indique que le geste est commencé. . . 160
9.31 Exploration de différents dispositifs de sortie afin d'explorer les axes *perspective* et *localisation spatiale* de notre espace de conception. 163

11.1 Les différents niveaux d'extension à nos travaux de thèse. 1) Niveau de chaque contribution pratique. 2) Niveau de chaque sous-tâche – sélection et contrôle. 3) Niveau pour faire le lien des contributions pratiques vers celles conceptuelles. 4) Niveau global pour relier les sous-tâches de sélection et de contrôle. 175

11.2 Détail du niveau 3 avec le lien pouvant mener à l'enrichissement de nos espaces de conception (*affine*) et le lien permettant l'établissement du *pouvoir évaluatif* de nos espaces grâce aux résultats expérimentaux. Le schéma montre aussi le niveau 1 permettant les améliorations de nos techniques à partir des évaluations expérimentales. 176
11.3 Continuum de l'interaction dans un environnement physique. 177

Liste des tableaux

8.1 Exemples montrant l'orthogonalité des axe *Perspective* et *Référentiel*. . . . 102

9.1 Indices graphiques [Cipiloglu *et al.*, 2010], aussi appelés monoculaires [Cutting et Vishton, 1995] ou monoculaires statiques [Ware, 2004] 136
9.2 Indices oculomoteurs [Cipiloglu *et al.*, 2010] [Cutting et Vishton, 1995] aussi appelés binoculaires [Ware, 2004] 138
9.3 Indices binoculaires [Cipiloglu *et al.*, 2010] [Cutting et Vishton, 1995] [Ware, 2004] 139
9.4 Indices reliés à la notion de mouvement [Cipiloglu *et al.*, 2010], aussi appelés monoculaires dynamiques [Ware, 2004] 139
9.5 Indices artificiels [Ware, 2004] 139

1 Introduction

Ce manuscrit présente les travaux de thèse réalisés dans l'équipe Ingénierie de l'Interaction Homme-Machine (IIHM) du Laboratoire d'Informatique de Grenoble (LIG). Cette thèse a été encadrée par Laurence Nigay et Céline Coutrix.

Nos travaux concernent l'interaction avec des objets physiques augmentés. Nous nous plaçons dans un contexte d'environnement intelligent dans lequel l'utilisateur peut interagir avec les objets environnants sans être obligé de se rapprocher pour manipuler directement l'interface physique de l'objet ou bien d'avoir une télécommande pour chaque objet interactif. Par exemple, nous aimerions que lorsqu'un utilisateur s'assied sur son canapé avec un livre, il puisse régler l'ambiance lumineuse et sonore sans avoir à se lever. Il s'avère en effet que si l'objet est à plus d'1,1m de distance de l'utilisateur, celui-ci préfèrera interagir à distance plutôt que de se lever pour toucher l'objet [Välkkynen *et al.*, 2006]. Pour interagir avec les objets physiques augmentés dans un contexte d'environnement intelligent, nous explorons l'interaction gestuelle 3D. L'interaction gestuelle permet en effet d'interagir à distance, et ce de manière dite familière donc intuitive [Raskin, 1994] puisque déjà utilisée au quotidien [Karam et Schraefel, 2005]. Pour étudier l'interaction gestuelle pour interagir avec des objets physiques augmentés distants, notre démarche consiste à distinguer deux sous-tâches de l'interaction : la sélection et le contrôle. Notre objectif est donc la conception de techniques d'interaction efficaces et plaisantes pour ces deux sous-tâches.

Cette démarche de travail a mené à quatre contributions distinctes (Figure 1.1) :

— **Sélection** des objets physiques augmentés

1) Nous contribuons au domaine du pointage en environnement intelligent par un geste en l'air de pointer vers l'objet physique avec la création d'un espace de conception. Cet espace de conception est dédié aux techniques de désambigüisation nécessaire lorsque plusieurs objets physiques sont sélectionnés par l'étape de pointage (chapitre 4). Cet espace de conception considère deux points de vue

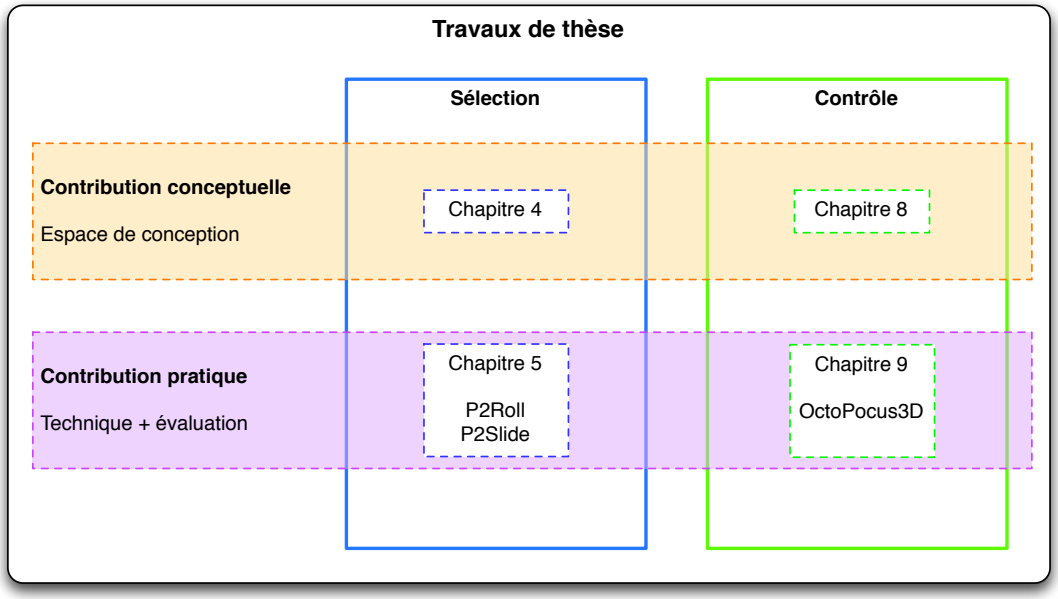


FIGURE 1.1 – Méthode de travail et contributions. Nous avons considéré l'interaction en deux étapes (ou sous-tâches) : la sélection et le contrôle des objets physiques augmentés. Pour chacune de ces étapes, nous apportons des résultats conceptuels (établissement d'un espace de conception) ainsi que des résultats pratiques (conception, développement et évaluation de techniques).

complémentaires : l'interaction et le système de désambigüisation.

2) Nous présentons ensuite la conception et l'évaluation expérimentale de deux techniques de désambigüisation, P2Roll (Physical Pointing Roll) et P2Slide (Physical Pointing Slide) (chapitre 5) en nous concentrant sur la partie *interaction* de notre espace. P2Roll utilise un roulis du poignet afin de naviguer dans l'ensemble des objets sélectionnés lors du pointage pour désigner l'objet cible. P2Slide quant à elle utilise un mouvement horizontal du pouce sur l'écran tactile d'un téléphone.

— **Contrôle** des objets physiques augmentés

3) Nous contribuons au domaine du contrôle gestuel avec la création d'un second espace de conception dédié aux techniques de guidage d'un ensemble de gestes 3D (chapitre 8) correspondant à des commandes. Cet espace considère deux catégories d'information fournies durant le guidage : le retour d'information rétroactif (feedback) et le retour d'information prédictif (feedforward). Ces deux catégories sont elles-mêmes composées de quatre groupes selon quatre questions : Quand, Quoi, Comment et Où. L'espace de conception étant conséquent en terme de nombre d'axes et de techniques existantes, nous fournissons un outil en ligne permettant

de faciliter son usage pour décrire, comparer et créer de nouvelles techniques de guidage.

4) Nous explorons ensuite plusieurs axes et options de conception pour la création d'une technique de guidage, OctoPocus3D (chapitre 9). Ce guide adapte un guide existant, OctoPocus [Bau et Mackay, 2008] pour des gestes 2D, à des ensembles de gestes 3D. L'ensemble des gestes 3D est représenté par des tubes 3D affichés sur un écran distant de l'utilisateur : ce dernier réalise des gestes 3D en l'air tout en étant guidé visuellement.

Relatant notre démarche de travail, le manuscrit comprend deux parties principales, l'une dédiée à la sélection par pointage d'objets et l'autre au contrôle d'un objet par spécification d'une commande via un geste (Figure 1.2).

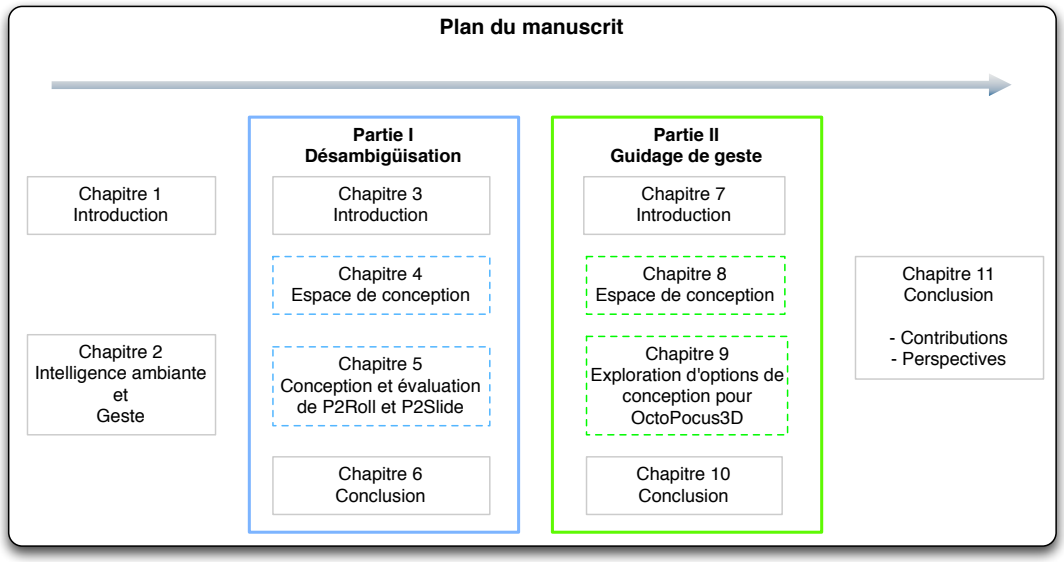


FIGURE 1.2 – Structure du manuscrit.

Nous présentons en premier lieu un chapitre introductif des deux parties du manuscrit pour définir le contexte et le sujet des travaux réalisés (chapitre 2). Ce chapitre décrit d'abord le contexte des travaux qui s'inscrivent dans le vaste domaine de l'intelligence ambiante (paragraphe 2.1). Nous abordons ensuite le concept de geste (paragraphe 2.2) qui est central aux travaux décrits dans les deux parties suivantes du manuscrit.

Nous présentons ensuite nos travaux sur la désambigüisation du pointage (partie I). Cette partie I comprend quatre chapitres. Le premier (chapitre 3) introduit le concept de la désambigüisation pour la sélection d'un objet dans le monde physique. Le second chapitre (chapitre 4) décrit notre espace de conception dédié aux techniques de désambigüisation.

Le troisième chapitre (chapitre 5) présente la conception et l'évaluation expérimentale de deux techniques de désambigüisation issues de l'exploration de notre espace de conception : P2Roll et P2Slide. Le dernier chapitre (chapitre 6) conclut sur le concept de la désambigüisation.

Nous présentons ensuite nos contributions sur le guidage de geste (partie II). Cette partie comprend quatre chapitres. Le premier chapitre (chapitre 7) motive notre intérêt pour le guidage de geste. Le second chapitre (chapitre 8) décrit un espace de conception dédié au comportement de technique de guidage. Le troisième chapitre (chapitre 9) présente l'exploration d'options de conception pour une technique de guidage de gestes 3D : OctoPocus3D. Le dernier chapitre conclut nos travaux sur le guidage de gestes (chapitre 10).

Nous concluons l'exposé de nos travaux en rappelant les contributions (paragraphe 11.1) et en définissant des pistes de recherche pour étendre ces travaux de thèse (paragraphe 11.2).

2 Interaction gestuelle pour l'intelligence ambiante

Sommaire

2.1 Intelligence ambiante : contexte des travaux 5

2.1.1 Domaine applicatif : sources lumineuses 5

2.1.2 Généralisation à l'intelligence ambiante 6

2.2 Interaction gestuelle : sujet des travaux 8

2.2.1 Caractérisation articulatoire du geste 9

2.2.1.1 Exécution d'un geste avec une partie du corps 9

2.2.1.2 Exécution d'un geste avec ou sans contact physique . . . 10

2.2.2 Caractérisation fonctionnelle du geste 11

2.1 Intelligence ambiante : contexte des travaux

Nous présentons le domaine applicatif de nos travaux lié à l'interaction avec des sources lumineuses et généralisons nos travaux au contexte de l'intelligence ambiante.

2.1.1 Domaine applicatif : sources lumineuses

Les dernières avancées technologiques concernant les LEDs (Light Emitting Diode) et OLEDs (Organic Light Emitting Diode) permettent d'envisager leur intégration dans l'éclairage de tous les jours. Ces sources lumineuses pourront être insérées directement dans les éléments environnants (murs, objets, tissus). De plus, l'utilisateur pourra contrôler plus de propriétés qu'avec les luminaires habituels comme la couleur, l'intensité, la température de couleur et la direction. Les interrupteurs traditionnels ne sont donc plus adaptés pour contrôler à la fois le grand nombre de sources lumineuses et leurs propriétés. Ces nouvelles fonctionnalités ouvrent de nouvelles pistes de recherche en terme d'interaction, discutées au

cours d'ateliers de travail [1] [Aliakseyeu *et al.*, 2012]. Cette thèse a pour domaine applicatif le contrôle interactif des lumières LEDs et OLEDs, avec comme objectif pratique la conception et le développement logiciel de techniques d'interaction pour ce contrôle de lumière.

Pour cela, ces travaux de thèse s'inscrivent dans le projet DELight de type FUI [2]. DELight est un projet coordonné par Schneider Electric et dédié aux sources lumineuses LEDs et OLEDs. Le but final est la création d'une LEDbox pour alimenter mais aussi contrôler des luminaires traditionnels, à LED mais aussi OLED. La LEDbox permet ainsi d'envisager des scénarios tels que le contrôle à distance des luminaires un par un ou bien la création d'ambiances lumineuses pré-enregistrées : par exemple une ambiance *lecture* lorsque l'on s'assied sur le canapé avec un livre. Le consortium est composé de onze partenaires, à savoir Saint Gobain pour la réalisation des produits de construction, Schneider Electric pour l'architecture électrique, Texas Instrument pour le développement électronique, Sonepar pour l'aménagement d'un site expérimental, le Centre de Recherche de l'Institut de Paul Bocuse pour l'aménagement de deux sites expérimentaux, Motwin pour la conception logicielle embarquée, EFS pour l'intégration technologique, Ingelux pour la conception des éclairages, ENTPE/DGCB pour la modélisation de l'éclairage, Université Paul Sabatier/Laplace pour la fabrication d'OLEDs, et le LIG pour l'interaction homme-machine (IHM).
Dans ce projet, nos travaux concernent l'interaction avec les sources lumineuses d'un environnement physique (Figure 2.1). Les techniques d'interaction conçues et développées communiquent donc avec les sources lumineuses en envoyant des requêtes à la LEDbox.

Trois démonstrateurs ont été définis dans le cadre de ce projet : une chambre d'hôtel, un bureau et une salle de restaurant. Nos travaux ont concerné les deux premiers démonstrateurs : la chambre d'hôtel et le bureau. Ces deux démonstrateurs nous ont permis de déployer nos techniques d'interaction conçues en contexte réel (OctoPocus3D paragraphe 9.5).

Bien que notre domaine applicatif soit le contrôle de sources lumineuses, nous généralisons nos travaux aux objets physiques augmentés en général (télévision, volets, chauffage, etc.), et pas seulement les éclairages à LED ou OLED. Cette généralisation à tous les objets potentiellement augmentés et interactifs nous place dans le contexte de l'intelligence ambiante décrite ci-après.

2.1.2 Généralisation à l'intelligence ambiante

L'intelligence ambiante, aussi notée informatique ubiquitaire ou pervasive, représente une vision de notre environnement physique quotidien augmenté [Coutaz et Crowley, 2008]

1. http://interactivelight.id.tue.nl/ (lien valide en septembre 2015) et https://lightingworkshop.wordpress.com/ (lien valide en septembre 2015) par exemple
2. Fonds unique interministériel

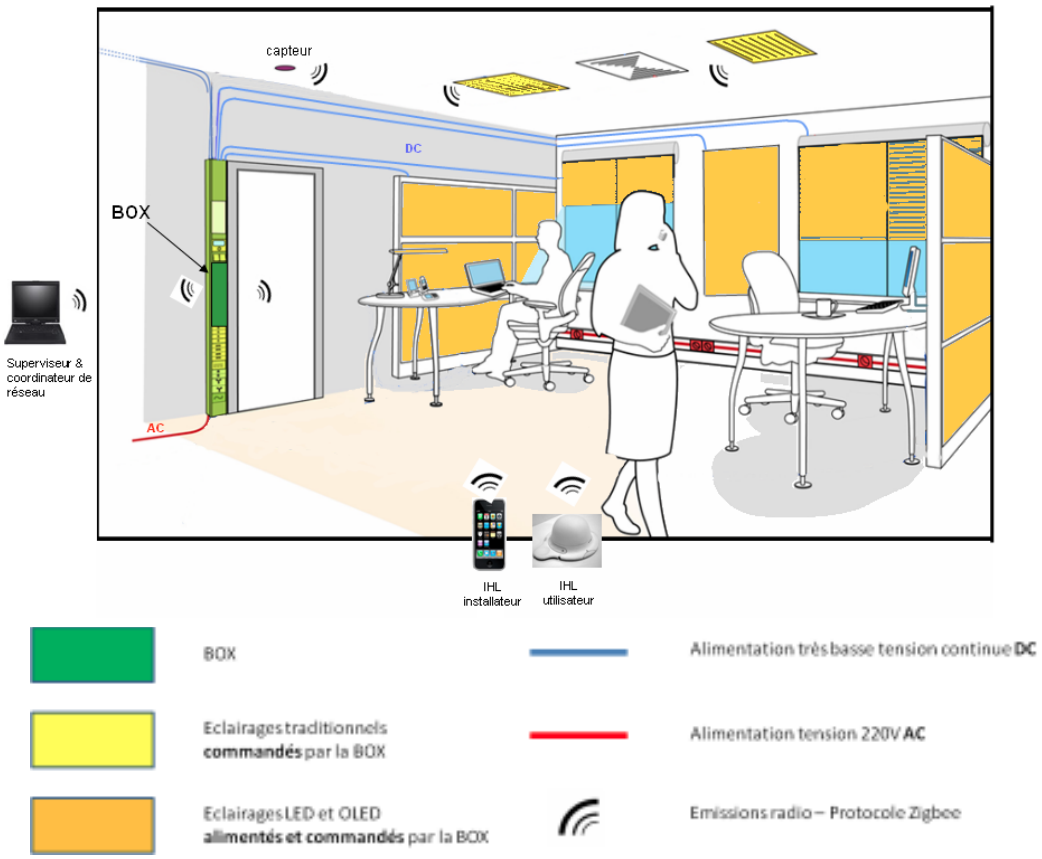


FIGURE 2.1 – Exemple d'application du projet DELight dans le bureau : les matériaux lumineux (innovants et traditionnels) sont contrôlés à distance par la BOX dans un environnement de bureaux[3]. La BOX est un point de convergence des différents équipements tels que les capteurs, les actionneurs ou bien les interfaces homme-machine (IHM).

[Sadri, 2011]. Cette augmentation permet aux utilisateurs d'interagir avec des informations ou services numériques confondus avec les environnements et activités du quotidien. Dans ce contexte, les techniques d'interaction en entrée et en sortie traditionnelles disparaissent (clavier et écran par exemple), laissant place à des capteurs et processeurs dans les objets physiques. Ces objets sont alors en mesure de faciliter les tâches de l'utilisateur de manière transparente.

Plusieurs qualificatifs sont utilisés pour caractériser l'intelligence ambiante [Sadri, 2011]. Celle-ci doit être :

— intégrée et discrète : l'aspect numérique n'est pas explicite car intégré dans le monde physique,

3. http://www.minalogic.org/PAR_TPL_IDENTIFIANT/2379/TPL_CODE/TPL_PROJET/30-search.htm#.Vb91RkWycyB (lien valide en septembre 2015)

— inter-connectée : les objets peuvent communiquer entre-eux, mais aussi les services, les utilisateurs,
— intelligente, adaptable et dynamique : les services s'adaptent au contexte et à l'expertise de l'utilisateur par exemple.

Les deux caractéristiques clés sont donc :

— *Intelligence* : l'environnement peut s'adapter à différentes situations et apprendre du comportement des utilisateurs,
— *Ambiante* : l'environnement intègre l'augmentation numérique de manière diffuse, de façon à ce que l'aspect informatique soit en arrière-plan des activités et de l'environnement de l'utilisateur.

Les travaux présentés dans ce manuscrit se rapportent à l'aspect *ambiant*. Nous cherchons à répondre aux besoins des utilisateurs lorsqu'ils interagissent avec les objets physiques augmentés. En effet, les services proposés par cette intelligence ambiante reposent avant tout sur des objets physiques augmentés. Cette augmentation offre de nouvelles possibilités quant aux manières d'interagir avec les objets physiques, notamment celle d'interagir à distance. Prenons le scénario d'un utilisateur s'installant dans son canapé avec un livre. L'habitat intelligent est configuré pour déclencher une ambiance tamisée avec un fond musical. Si l'utilisateur désire plus de luminosité, ou bien éteindre la musique, nous voulons lui offrir la possibilité de le faire sans qu'il n'ait besoin de se déplacer pour atteindre un interrupteur ou une télécommande. Nous étudions donc l'interaction à distance avec les objets physiques augmentés qui composent l'environnement ambiant. Pour cela, différentes modalités peuvent être utilisées telles que l'utilisation de commandes vocales par exemple. Nous pensons d'ailleurs que les habitats intelligents seront un jour en mesure de prendre en compte plusieurs modalités en entrée pour interagir avec les objets du quotidien. Chacune de ces modalités d'entrée doit être étudiées par la communauté afin d'être utilisable prochainement de manière multimodale. Notre contribution concerne l'une de ces modalités : la modalité gestuelle. Nous considérons en effet la modalité gestuelle, que tout le monde utilise déjà quotidiennement afin d'interagir avec des objets ou d'autres personnes, mais qui nécessite encore des améliorations lorsque l'on souhaite s'en servir pour interagir avec un système informatique. Afin de cerner les gestes concernés par nos travaux, la partie suivante présente une vue d'ensemble des catégories de gestes en fonction de leurs caractérisations articulatoire et fonctionnelle.

2.2 Interaction gestuelle : sujet des travaux

Le geste peut être caractérisé selon deux aspects complémentaires : sa réalisation, c'est-à-dire son aspect articulatoire, et sa sémantique, c'est-à-dire son aspect fonctionnel.

2.2.1 Caractérisation articulatoire du geste

Deux gestes peuvent transmettre le même message tout en ayant des aspects articulatoires différents quant à leur réalisation. Par exemple, l'agacement peut être transmis en tapant le devant de la paume du pied sur le sol, en tapant du doigt sur une table ou bien en faisant un mouvement de tête. Nous distinguons les parties du corps engagées dans la réalisation du geste ainsi que les gestes impliquant un contact physique ou pouvant être exécutés en l'air.

2.2.1.1 Exécution d'un geste avec une partie du corps

L'interaction gestuelle concerne la totalité du corps humain, rendant ainsi cette modalité d'entrée très riche en terme d'expressivité [Cadoz, 1994]. Il existe un large ensemble de possibilités pour utiliser l'interaction gestuelle. Celle-ci peut en effet demander l'utilisation du corps entier ou bien simplement des micro-mouvements effectués avec un seul membre du corps.

— *Corps entier* : au-delà d'une utilisation du corps pour des mouvements artistiques, certains travaux proposent d'utiliser l'ensemble du corps afin d'interagir avec un système.
Par exemple, BodyAvatar permet à l'utilisateur d'utiliser son propre corps afin de modéliser un avatar 3D [Zhang *et al.*, 2013].

— *Plusieurs membres* : certains travaux proposent d'utiliser un sous-ensemble des parties du corps pour interagir avec un système.
Par exemple, [Cohé et Hachet, 2012] ont étudié les gestes multipoints (multitouch) permettant d'interagir avec des environnements 3D sur des surfaces tactiles 2D. L'utilisateur peut ainsi manipuler le contenu numérique grâce à des gestes effectués avec les deux mains. Gestu-Wan est une technique de guidage qui propose de guider des gestes effectués avec toute la partie haute du corps [Rovelo *et al.*,], pas seulement les deux mains.

— *Un seul membre* : certains travaux proposent d'utiliser un seul membre pour l'interaction gestuelle. Les systèmes utilisant un dispositif tel qu'une télécommande wiimote pour capturer les gestes rentrent dans cette catégorie [Vatavu, 2013].
Par exemple, un Marking Menu permet d'effectuer des gestes avec seulement une main [Kurtenbach *et al.*, 1993].
Le membre choisi peut être n'importe quel membre du corps humain, allant de la tête [Morency et Darrell, 2006] au pied [Alexander *et al.*, 2012] en passant par la main [Ren et O'Neill, 2012].

Dans la partie I du manuscrit sur la sélection d'un objet physique, nous étudions les gestes effectués avec le bras pour désigner un objet. Dans la partie II sur le contrôle et guidage de gestes, nous considérons dans un premier temps tous les gestes, sans prendre en

compte les parties du corps avec lesquelles ils sont effectués. Cela nous permet d'examiner un large spectre d'options de conception existantes. Dans un second temps, nous nous intéressons aux gestes 3D effectués avec la main pour effectuer des gestes symboliques dynamiques et interagir avec des objets physiques augmentés.

2.2.1.2 Exécution d'un geste avec ou sans contact physique

La distinction entre des gestes exécutés avec ou sans contact est proche de la distinction faite entre des gestes 2D ou 3D. Toutefois, des gestes exécutés en l'air et sans contact peuvent être 2D, c'est-à-dire dans un plan, ou bien 3D. Nous considérons la première distinction car elle permet de traiter un aspect important de l'interaction gestuelle : la segmentation d'un geste.

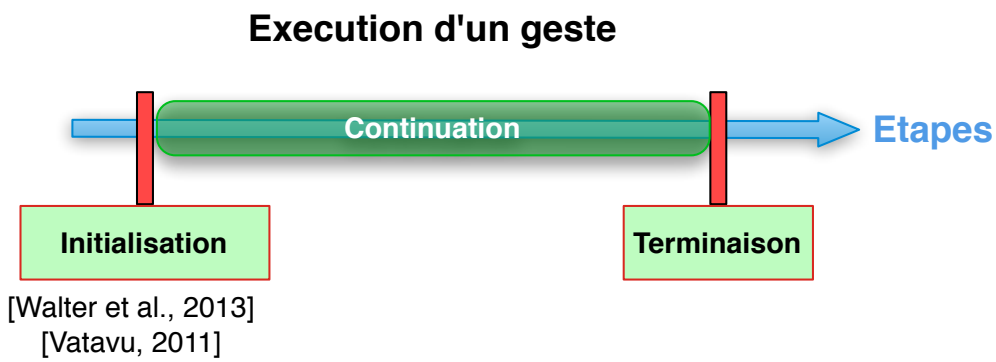


FIGURE 2.2 – Les différentes étapes de l'interaction gestuelle : l'initialisation du geste, sa continuation et sa terminaison.

Trois étapes sont présentes dans l'interaction gestuelle [Walter *et al.*, 2013] (Figure 2.2) :

1. l'*initialisation* du geste, i.e. le moment à partir duquel l'utilisateur et le système savent que le geste est commencé,
2. la *continuation* du geste, i.e. l'exécution du geste et
3. la *terminaison* du geste, i.e. le moment à partir duquel l'utilisateur et le système savent que le geste est fini.

Pour des gestes 2D sur une surface tactile par exemple [Freeman *et al.*, 2009], l'initialisation (respectivement la terminaison) correspond souvent au moment où le doigt se pose sur (respectivement quitte) la surface. Il est plus difficile de définir ces étapes d'initialisation et de terminaison avec des gestes 3D, d'autant plus s'ils sont effectués sans instrument pour laisser les mains libres, et donc sans contact. Certains travaux ont traité le problème de l'initialisation des gestes 3D. Par exemple, [Walter *et al.*, 2013] se sont intéressés à la révélation d'un geste initial pouvant servir à détecter le début d'un geste afin d'interagir avec des dispositifs d'affichage publiques. Les résultats montrent que les utilisateurs ont tendance à exécuter le geste initial (forme d'une théière avec les bras et le haut du corps

en mettant une main sur la hanche) lorsque ce geste est révélé, plus particulièrement lorsqu'il est affiché en permanence sur une partie dédiée de l'écran. Au lieu de chercher à reconnaître un geste ou une posture spécifique pour la phase d'initialisation, d'autres approches visent à caractériser les différentes manières de détecter des évènements tels que le début d'un geste par traitement vidéo [Vatavu, 2011].

Dans la partie I sur la sélection d'objets augmentés, nous ne considérons que des gestes 3D. Au contraire, dans la partie II dédiée au contrôle et guidage de gestes, nous considérons dans un premier temps toutes les techniques de guidage existantes, que ce soit pour des gestes avec ou sans contact. Comme précédemment, ce choix nous permet de considérer le plus grand nombre d'options de conception existantes. Dans un second temps, nous nous intéressons au guidage de gestes 3D de la main en l'air, donc sans contact, afin d'interagir avec les objets physiques augmentés.

2.2.2 Caractérisation fonctionnelle du geste

Plusieurs travaux proposent de classer les gestes que nous effectuons dans la vie de tous les jours [Cadoz, 1994] [Karam et Schraefel, 2005] [Aigner *et al.*, 2012]. Ces classifications sont généralement dépendantes du domaine de recherche dans lequel elles s'inscrivent : elles peuvent être liées à la caractérisation mathématique des gestes [Vatavu et Pentiuc, 2008] ou bien être liées à la linguistique lorsque le geste accompagne un dialogue ou un discours [McNeil, 1992] par exemple. Nous présentons des classifications se basant sur la fonction des gestes, c'est-à-dire des classifications sémantiques des gestes [Sorel, 2012] (Figure 2.3).

[Cadoz, 1994] définit trois types de gestes :

- *Epistémique* : ce type de gestes permet d'obtenir des informations sur l'environnement grâce au sens du toucher. Ce sens s'appuie sur les propriétés tactiles, haptiques et proprioceptives de la main. Cela permet d'appréhender la température d'un objet, l'état de sa surface ou encore sa forme.
- *Ergotique (ou manipulatif dans les environnements numériques [Karam et Schraefel, 2005] [Aigner* et al.*, 2012])* : ce type de gestes vise la manipulation, la modification ou la création des artéfacts du monde physique (ou du monde numérique). Cela permet de manipuler directement un objet en le tournant par exemple.
- *Sémiotique* : ce type de gestes permet de transmettre de l'information. Cela permet de spécifier des commandes par exemple.

Nous nous intéressons aux gestes sémiotiques permettant de transmettre de l'information. Ces gestes sémiotiques peuvent être déclinés à leur tour en plusieurs types en fonction du message informationnel transmis [Karam et Schraefel, 2005] [Aigner *et al.*, 2012] (Figure 2.3) :

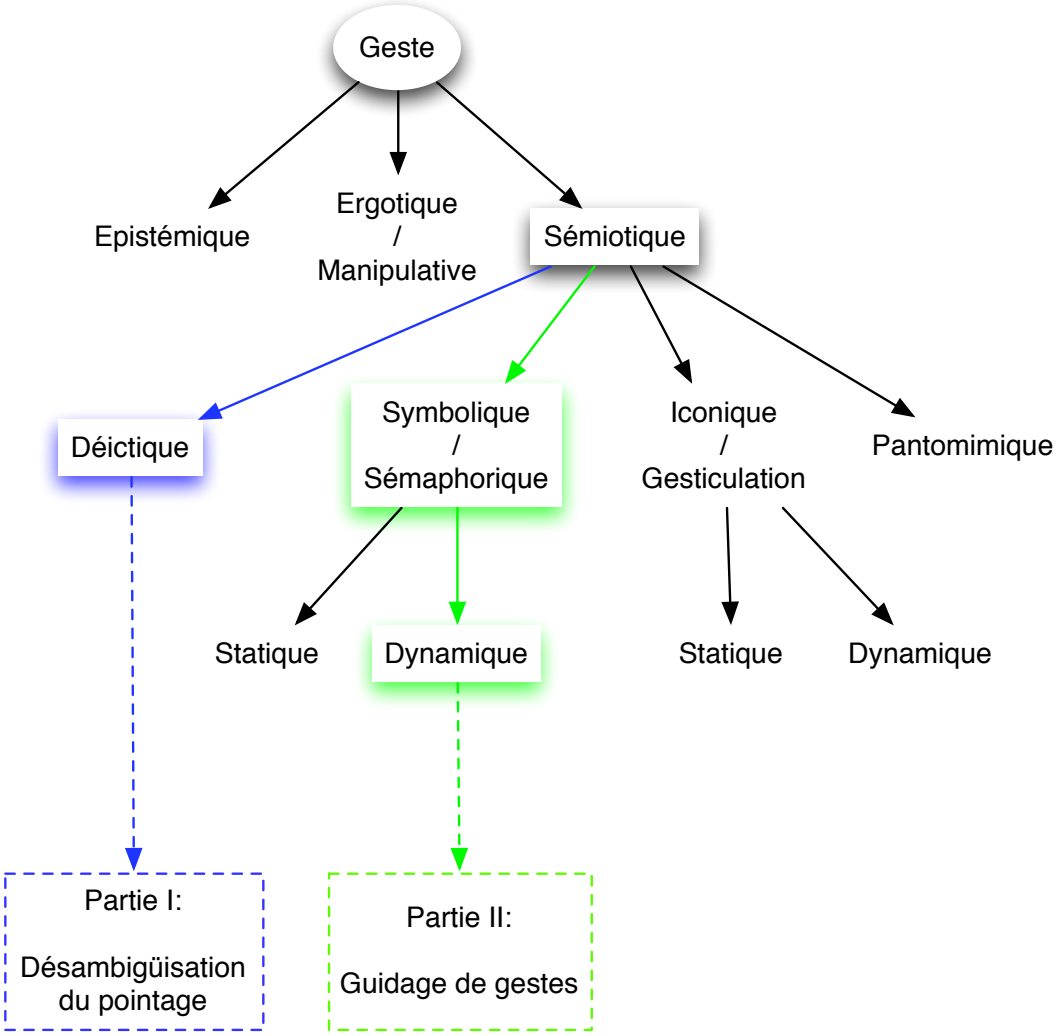


FIGURE 2.3 – Vue d'ensemble des types de gestes à partir de trois classifications existantes [Cadoz, 1994] [Karam et Schraefel, 2005] [Aigner *et al.*, 2012] et positionnement des deux parties principales du manuscrit.

— *Déictique* : ce type concerne les gestes servant à désigner un objet ou une localisation spatiale (pointer vers une source lumineuse par exemple).
— *Symbolique (ou sémaphorique)* : ce type inclut les gestes auxquels sont associées des significations, l'association pouvant être arbitraire ou non. Ces associations définissent un dictionnaire et ainsi un vocabulaire permettant de transmettre une information via un geste. Ces gestes peuvent être statiques (par exemple faire le signe "OK") ou bien dynamique (par exemple lever la main pour allumer la cafetière).
— *Iconique (ou gesticulation ou coverbal)* : ce type concerne les gestes permettant d'ajouter de l'information au discours (par exemple représenter la forme, la taille

ou encore l'orientation d'un objet dont on est en train de parler). A nouveau, ces gestes peuvent être statiques (par exemple représenter la forme d'un triangle avec ses doigts) ou dynamiques (par exemple représenter la forme d'un triangle en la traçant en l'air).

— *Pantomimique* : ce type concerne les gestes permettant de montrer un mouvement (par exemple tourner un volant imaginaire).

Ces classifications de gestes permettent de situer nos travaux de thèse vis-à-vis des gestes étudiés dans le contexte de l'intelligence ambiante et en particulier le contrôle de sources lumineuses. Comme expliqué dans le chapitre introductif (chapitre 1), notre démarche de travail consiste à distinguer deux sous-tâches dans l'interaction avec des objets physiques augmentés : la *sélection* de l'objet puis son *contrôle*. Pour chacune de ces sous-tâches, nous considérons un type de geste différent.

La première sous-tâche, la *sélection* d'un objet physique, traite de gestes déictiques afin de sélectionner un objet via un geste de pointage (partie I). Plus précisément, nous nous intéressons à la désambigüisation nécessaire après un geste de pointage non précis. Ce geste de pointage non précis facilite la sélection d'objets petits et/ou distants, mais peut néanmoins impliquer la sélection de plusieurs objets dans le cas d'environnements denses : une phase de désambigüisation est alors nécessaire.

La seconde, la sous-tâche de *contrôle*, traite des gestes symboliques dynamiques qui permettent de spécifier des commandes (partie II). Ce type de gestes est communément utilisé pour spécifier des commandes [Aigner *et al.*, 2012] [Karam et Schraefel, 2005] [Walter *et al.*, 2013]. Plus précisément, nous nous intéressons au guidage de ce type de gestes qui nécessite d'informer l'utilisateur du vocabulaire gestuel proposé. En effet, un utilisateur novice doit pouvoir découvrir les commandes disponibles ainsi que les gestes associés afin de contrôler son environnement augmenté. Dans un premier temps, pour considérer le plus grand nombre d'options de conception possibles pour une technique de guidage de gestes, nous considérons toutes les techniques de guidage existantes, sans distinction quant aux types de gestes guidés par chacun. Dans un second temps, nous étudions le guidage de gestes *symboliques dynamiques* afin d'interagir avec des objets physiques augmentés.

Première partie

Sélection d'un objet physique

Pointage par des gestes 3D : Nécessité de désambigüiser

Table des matières

3 Pointer dans le monde physique **19**

3.1 Différentes méthodes de sélection dans le monde physique 19

3.2 Limites du pointage dans le monde physique 20

3.3 Le pointage et la désambigüisation 22

4 Espace de Conception **27**

4.1 Introduction 27

4.2 Espace de Conception 29

4.2.1 Groupe *Interaction* 29

4.2.2 Groupe *Système de Désambigüisation* 32

4.2.3 Liens entre les groupes *Interaction* et *Système de Désambigüisation* 35

4.2.4 Positionnement : espaces de conception existants 36

4.3 Evaluation de l'espace de conception 39

4.4 Conclusion 45

5 Techniques de désambigüisation proposées : P2Roll et P2Slide **47**

5.1 Introduction 48

5.1.1 Rechercher des chemins inexplorés dans notre espace de conception 48

5.1.2 Propriétés des actions 48

5.1.3 Désambigüisation et loi de Fitts 50

5.2 Conception de P2Roll et P2Slide 51

5.2.1 Quasi-mode 51

5.2.2 Physical Pointing Roll : P2Roll 52

5.2.3 Physical Pointing Slide : P2Slide 54

5.3 Evaluation expérimentale 56

5.3.1 Technique de référence : liste 56

5.3.2 Hypothèses 57

5.3.3 Equipement 58
5.3.4 Participants 59
5.3.5 Procédure 59
5.3.6 Protocole expérimental 60
5.3.7 Résultats 61
5.4 Conclusion sur l'évaluation expérimentale 67
5.5 Travaux futurs à partir de l'évaluation expérimentale 68

6 Conclusion de la partie sélection **71**
6.1 Conclusion sur la désambigüisation 71
6.2 Travaux futurs sur la désambigüisation 72

3 Pointer dans le monde physique

Sommaire

3.1 Différentes méthodes de sélection dans le monde physique . . . 19
3.2 Limites du pointage dans le monde physique 20
3.3 Le pointage et la désambigüisation 22

Nous présentons dans ce chapitre notre étude concernant la *sélection* d'un objet physique augmenté via un geste de pointage, c'est-à-dire un geste *déictique* (paragraphe 2.2). Pour cela, nous exposons tout d'abord les avantages et inconvénients de différentes manières de sélectionner un objet physique, nous ayant mené à l'étude des gestes déictiques. Nous listons ensuite les contraintes du pointage dans le monde physique et terminons par l'introduction du concept de désambigüisation.

3.1 Différentes méthodes de sélection dans le monde physique

Afin de sélectionner des objets physiques augmentés, trois paradigmes d'interaction ont été définis [Rukzio *et al.*, 2007] [Ailisto *et al.*, 2006] :

— *Le balayage* est une méthode de sélection en deux étapes. La première permet à l'utilisateur d'obtenir la liste des objets interactifs présents autour de lui. L'utilisateur choisit ensuite (dans une liste affichée sur un dispositif mobile par exemple) l'objet qu'il désire sélectionner. Bien que cette solution présente l'inconvénient d'être en deux étapes, elle permet à l'utilisateur de connaître les objets interactifs environnants sans qu'il ne soit obligé d'avoir une connaissance *a priori* des objets de l'environnement dans lequel il se trouve.

— *Le pointage* permet une sélection à distance des objets interactifs via un geste déictique (paragraphe 2.2). Ce type de geste est souvent qualifié d'intuitif car utilisé dans des situations quotidiennes [Karam et Schraefel, 2005]. Cette méthode

de sélection nécessite cependant que l'utilisateur puisse (1) visuellement identifier et (2) viser les objets interactifs, c'est-à-dire avoir une ligne de visée dégagée.
— *Le toucher* permet une sélection en touchant directement l'objet interactif. Bien que cette méthode de sélection ne demande pas de précision de la part de l'utilisateur, elle implique que les objets interactifs soient à portée de main de celui-ci.

Des travaux ont comparé ces trois méthodes de sélection. Il s'avère que les utilisateurs préfèrent pointer vers un objet plutôt que de le sélectionner dans une liste affichée sur un écran [Swindells *et al.*, 2002]. De plus, la distance séparant l'utilisateur de l'objet qu'il souhaite sélectionner influence la méthode de sélection choisie [Rukzio *et al.*, 2007] [Välkkynen *et al.*, 2006] : dès que les objets physiques augmentés sont à plus de 1,1m de distance, les utilisateurs préfèrent pointer les objets pour les sélectionner plutôt que de se déplacer et les toucher. Ces résultats montrent que le pointage est une solution prometteuse pour la sélection d'objets physiques à distance.

Résumé

Nous nous intéressons à la sélection d'un objet physique grâce à un geste de pointage. En effet, les utilisateurs ont tendance à préférer pointer vers un objet afin de le sélectionner si cet objet n'est pas à portée de main.

3.2 Limites du pointage dans le monde physique

De nombreuses études sur le pointage 3D en environnement virtuel existent et sont référencées dans [Argelaguet et Andujar, 2013]. Les premiers systèmes numériques visaient à imiter le monde physique [Poupyrev et Ichikawa, 1999]. Des améliorations ont permis ensuite de palier aux difficultés dues à la taille des cibles, à leur densité dans la scène 3D ou bien à l'occultation d'objets. Aussi le monde physique et ses spécificités rendent difficile l'extension de techniques de pointage 3D existantes en environnement numérique :

— L'emplacement des cibles physiques est difficilement modifiable par le système. De fait, une reconfiguration spatiale pour faciliter la sélection comme le fait le Flower Ray en réarrangeant les objets pointés sous forme de menu circulaire est difficile [Grossman et Balakrishnan, 2006].
— L'emplacement visé par l'utilisateur est directement couplé au dispositif de pointage. La modification de ce couplage afin d'appliquer du gain et ainsi faciliter la sélection est donc compliquée [König *et al.*, 2009].
— La représentation de l'outil de pointage est inexistante. Par exemple, en utilisant un laser, il n'y a pas de représentation du rayon, seulement du point extrême. Il est donc difficile d'adapter des techniques comme le Flexible Pointer au monde physique [Feiner, 2003]. En effet, le Flexible Pointer nécessite de visualiser la courbure du rayon qui permet de pointer des objets occultés.

Comme dans les environnements numériques 3D, il existe deux métaphores du pointage physique : le pointage effectué avec un rayon (comme avec un laser) et celui effectué avec un volume (comme avec une télécommande infrarouge).
Cependant, pointer simplement avec un laser mène à des limites quant à la précision obtenue [König *et al.*, 2009] [Myers *et al.*, 2002]. Les tremblements inhérents au pointage et les limites de la précision motrice humaine créent des déviations de $\pm 5\,^{\circ}$ [Forsberg *et al.*, 1996], soit 17cm lors d'une interaction à une distance de 2m. Ces tremblements peuvent apparaître lorsque l'utilisateur essaie de garder une position stable du bras en l'air. Notons aussi que le fait de valider la sélection en appuyant ou relâchant un bouton peut provoquer des mouvements non intentionnels au dernier moment et ainsi mener à une mauvaise sélection. Ce phénomène est connu sous le nom d'*effet Heisenberg* [Bowman *et al.*, 2001].

Les techniques existantes tentant de répondre à ce problème ont d'autres limitations :

1. La gestion par le système d'un modèle du monde physique [Chavez *et al.*, 2008] [Wilson et Pham, 2003].
   Par exemple, WorldCursor [Wilson et Pham, 2003] est un système permettant le pointage d'objets physiques composé de deux parties : un dispositif tenu dans la main de l'utilisateur, XWand [Wilson et Shafer, 2003], et un laser accroché à une plateforme motorisée au plafond. Tandis que l'utilisateur vise un objet avec XWand, le découplage entre le dispositif de pointage et le pointeur lui-même permet d'adapter des techniques de pointage conçues initialement pour les interfaces graphiques. Par exemple, WorldCursor peut appliquer un gain comme le Semantic Pointing [Blanch *et al.*, 2004] ou bien un filtre [Casiez *et al.*, 2012], pour profiter du fait que le mouvement de pointage (espace moteur) ne soit pas lié de manière absolue à l'objet (espace visuel). Le pointeur peut aussi *s'accrocher* à l'objet le plus proche de la direction visée, comme la technique proposée pour le Delphian Desktop [Asano *et al.*, 2005]. Pour cela, WorldCursor a besoin de connaître la position des objets physiques. Ces positions doivent être enregistrées dans le système pendant une phase d'installation effectuée par l'utilisateur.
   La gestion d'un modèle du monde physique par le système limite ainsi sa dynamicité (par exemple, bouger une lampe dans le monde physique doit être reporté au système afin de mettre à jour le modèle).

2. Des systèmes invasifs, tels que de grands capteurs [Patel et Abowd, 2003] ou bien une plate-forme fixée au plafond [Wilson et Pham, 2003].
   Par exemple, la solution proposée par [Patel et Abowd, 2003] demande d'installer de larges capteurs sur les objets physiques afin de faciliter la visée lors d'une sélection via pointage.
   Outre l'aspect esthétique, des systèmes invasifs sont en conflit sur l'aspect *diffus* de l'intelligence ambiante (paragraphe 2.1.2), qui veut que l'aspect numérique servant à augmenter le monde physique soit le plus discret possible.

Les techniques de pointage basées sur un volume de sélection sont de bonnes alternatives pour éviter des problèmes de précision motrice (Figure 3.1). Cependant, plusieurs objets peuvent se trouver dans ce volume de sélection, nécessitant ainsi une étape supplémentaire afin d'affiner la sélection faite par l'étape de pointage.

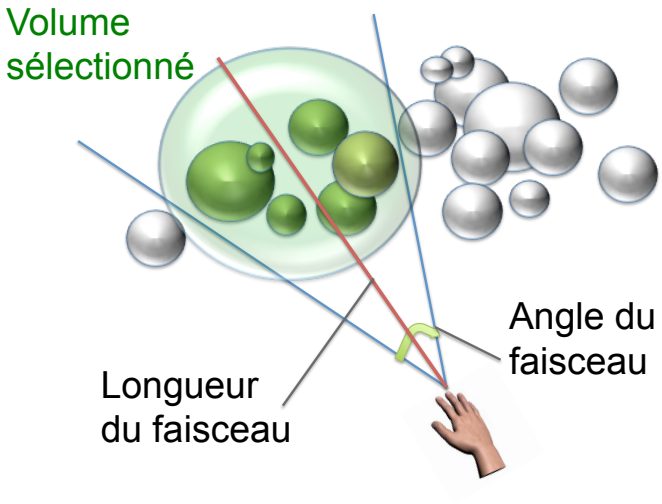


FIGURE 3.1 – Pointage basé sur une sélection volumique. Un ensemble d'objets physiques inclus dans le volume (c'est-à-dire le cône) est sélectionné.

Résumé

Un geste de pointage absolu basé sur un rayon demande une certaine précision de la part de l'utilisateur. De plus, le pointage dans le monde physique peut difficilement bénéficier des améliorations faites pour les techniques de pointage en environnements numériques 3D.

3.3 Le pointage et la désambigüisation

La machine à états finis de la Figure 3.2 modélise une tâche de pointage incluant une étape de désambigüisation. L'automate illustre trois stratégies de désambigüisation sans montrer chaque détail (comme le fait d'annuler une action pour retourner à l'état précédent) afin de ne pas surcharger le schéma. L'état initial correspond à l'état dans lequel aucun objet physique n'est sélectionné. Nous identifions trois chemins à partir de cet état initial :

— $(t1^*.t5)$ décrit un *pointage direct*, par exemple lors d'un pointage absolu avec un laser [Chavez *et al.*, 2008] [Patel et Abowd, 2003] [Wilson et Pham, 2003] [Wilson et Shafer, 2003]. Ce type de pointage est efficace lorsque l'on souhaite sélectionner de grands objets physiques (une télévision par exemple) dans un environnement peu dense.

— $(t1^*.t6.t7 = t1^*.t6)$ correspond à un mécanisme de désambigüisation *automatique*,

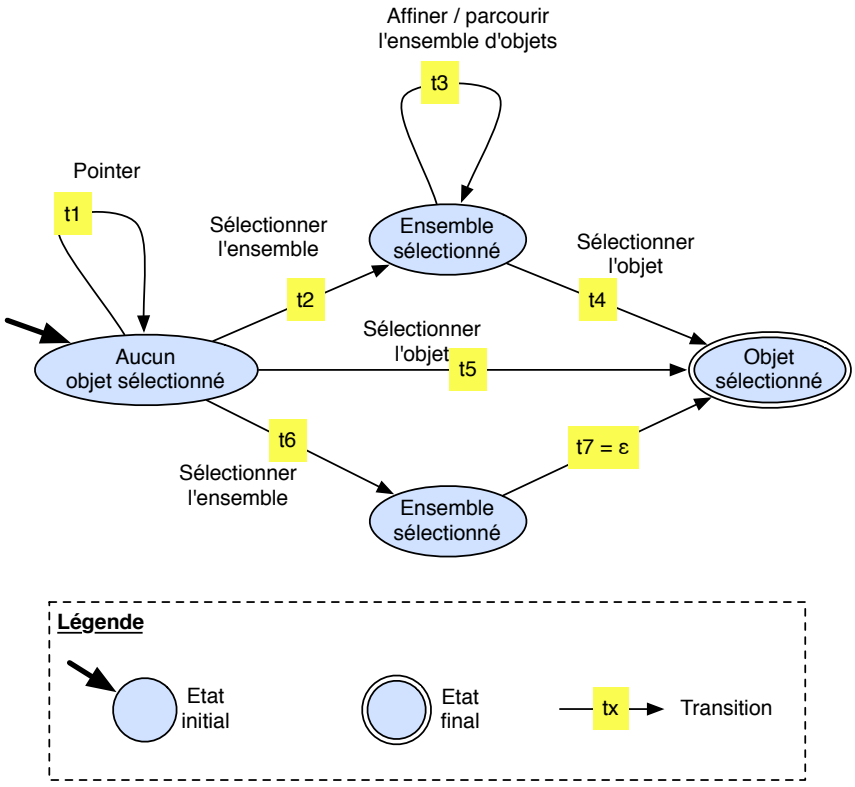


FIGURE 3.2 – En commençant par un geste de pointage dans le monde physique, trois chemins sont possibles pour sélectionner un objet. ($t1^*.t2.t3^*.t4$) : geste de pointage suivi par un processus de désambigüisation interactif. ($t1^*.t5$) : geste de pointage seulement. ($t1^*.t6.t7 = t1^*.t6$) : geste de pointage accompagné d'une heuristique afin de sélectionner automatiquement un objet.

c'est-à-dire effectué par le système. L'utilisateur n'effectue pas d'action supplémentaire pour atteindre l'état final. Deux types d'automatisme existent [Argelaguet et Andujar, 2013] :

- Certains automatismes définissent des heuristiques qui ne dépendent que de caractéristiques au moment de la validation de la sélection. Le système peut par exemple sélectionner l'objet le plus proche de la génératrice du cône (droite définissant le centre du cône) [Rahman *et al.*, 2010].
- Certains automatismes définissent des heuristiques qui dépendent du comportement de l'utilisateur pendant l'étape de pointage. Le système peut par exemple tenir compte de la taille de l'objet ainsi que de sa distance par rapport à la génératrice du cône (droite définissant le centre), mais aussi de la vitesse du geste de pointage pendant toute la durée de l'interaction [Steed, 2006].

— ($t1^*.t2.t3^*.t4$) correspond à un mécanisme de désambigüisation *interactif*. Trois stratégies existent pour interactivement sélectionner un objet parmi n :

- L'utilisateur *parcourt* l'ensemble d'objets pré-sélectionnés ($t3^*$) jusqu'à at-

teindre et valider l'objet désiré ($t4$). Par exemple, après un pointage dans le monde physique, le nom des objets peut être affiché sur un dispositif mobile. L'utilisateur doit alors parcourir cette liste ($t3^*$) afin de sélectionner l'objet voulu ($t4$).

- L'utilisateur *affine progressivement* sa sélection jusqu'à ce que cette sélection ne contienne plus que l'objet voulu. C'est le cas du Flower Ray [Grossman et Balakrishnan, 2006] ou de SQUAD [Kopper *et al.*, 2011], qui proposent d'afficher les objets numériques pré-sélectionnés par groupe dans un menu tout autour du centre de la sélection. Lorsque l'utilisateur choisit un groupe (c'est-à-dire une option du menu), le menu est mis à jour avec les éléments de ce groupe répartis en nouveaux sous-groupes ($t3^*$), et ce jusqu'à ce qu'il n'y ait qu'un objet par option du menu ($t4$).
- L'utilisateur effectue une autre désignation ($t3$) dans l'ensemble d'objets précédemment sélectionnés. Par exemple, RFIG Lamp [Raskar *et al.*, 2004] utilise ce type de désambigüisation interactive : Après un geste de pointage ($t1$), le système projette des balises sur les objets physiques sélectionnés. L'utilisateur peut ensuite utiliser un laser afin de désigner l'objet désiré ($t3$) en visant la balise correspondante et valider ($t4$).

Nous constatons que :

1. Un pointage direct ($t1^*.t4$) atteint ses limites lorsque l'on souhaite sélectionner des objets petits et/ou éloignés et/ou dans un environnement dense, et
2. Les méthodes de désambigüisation automatiques ($t1^*.t6.t7 = t1^*.t6$) peuvent empêcher l'utilisateur de corriger son erreur. En effet, l'heuristique utilisée peut passer inaperçue auprès de l'utilisateur, l'empêchant ainsi de comprendre pourquoi sa sélection est erronée, et donc de se corriger la fois suivante.

Aussi à partir de notre analyse de la Figure 3.2, nous nous intéressons aux mécanismes de désambigüisation interactifs ($t1^*.t2.t3^*.t4$).

Résumé

Afin de faciliter la sélection d'objets physiques augmentés, nous nous intéressons à l'étape de désambigüisation qui suit une action de pointage 3D utilisant un volume de sélection.
Les solutions basées sur une sélection volumique répondent au problème lié à la précision d'un pointage absolu. Il peut toutefois y avoir une ambiguïté si le volume de sélection contient plusieurs objets. Les mécanismes de désambigüisation automatiques pouvant empêcher l'utilisateur de corriger leurs erreurs, nous nous proposons d'étudier les mécanismes de désambigüisation interactifs, c'est-à-dire effectués par l'utilisateur.

Afin d'étudier les mécanismes de désambigüisation interactifs, nous proposons deux études complémentaires. La première étude concerne la conception de mécanismes de

désambigüisation avec la création d'un espace de conception dédié aux techniques de désambigüisation dans le monde physique (chapitre 4). Reposant sur notre espace de conception, la seconde étude est pratique avec la conception et l'évaluation expérimentale de deux techniques de désambigüisation afin d'explorer des options de conceptions de notre espace (chapitre 5).

4 Espace de Conception

Sommaire

4.1 Introduction **27**
4.2 Espace de Conception **29**
4.2.1 Groupe *Interaction* 29
4.2.1.1 Espace de représentation 29
4.2.1.2 Espace de contrôle 31
4.2.2 Groupe *Système de Désambigüisation* 32
4.2.2.1 *Obtention des connaissances* 32
4.2.2.2 Sous-groupe *Quoi* 33
4.2.2.3 Sous-groupe *Où* 34
4.2.3 Liens entre les groupes *Interaction* et *Système de Désambigüisation* 35
4.2.4 Positionnement : espaces de conception existants 36
4.2.4.1 Désambigüisation en environnements numériques 36
4.2.4.2 Interaction dans le monde physique 38
4.2.4.3 Relation entre les mondes physique et numérique 38
4.3 Evaluation de l'espace de conception **39**
4.4 Conclusion **45**

4.1 Introduction

Nous avons présenté dans les chapitres précédents les intérêts d'un geste *déictique* (paragraphe 2.2) afin de sélectionner un objet physique augmenté via un geste de pointage. La sélection par pointage pouvant rencontrer des difficultés dues à des limitations du monde physique (paragraphe 3.2), nous nous intéressons au pointage volumique, c'est-à-dire utilisant un volume de sélection. Ce volume pouvant capturer plusieurs objets physiques, une étape de désambigüisation est potentiellement nécessaire avant de confirmer la sélection. Plusieurs approches existent afin d'effectuer cette désambigüisation (paragraphe 3.3). Nous focalisons sur l'approche *interactive*, c'est-à-dire exécutée par l'utilisateur.

L'objectif de ce chapitre est d'organiser les techniques de désambigüisation interactives existantes selon des options de conception permettant ainsi de les décrire, de les comparer et de créer de nouvelles techniques.

Nous présentons un espace de conception pour capturer les spécificités de l'étape de désambigüisation interactive qui suit un geste de pointage dans le monde physique. Cet espace de conception, composé de 11 axes, adopte deux points de vue : l'un pour caractériser l'aspect lié à l'interaction (Groupe *Interaction*), l'autre pour caractériser les prérequis du système de désambigüisation à développer (Groupe *Système de Désambigüisation*). Ce dernier est lui-même composé de deux sous-groupes suivant deux questions : *Quoi* et *Où* (Figure 4.1). Ces deux sous-groupes sont spécifiques au contexte de l'interaction dans le monde physique. En effet, ces axes caractérisent les connaissances que le système doit posséder et maintenir à propos du monde physique. Le groupe *Système de Désambigüisation* adopte donc un point de vue technique en identifiant les implications du groupe *Interaction* sur le système global.

Il est à noter que nous ne cherchons pas à décrire en détail la modalité d'entrée utilisée pour l'étape de désambigüisation. En effet, en fonction de la modalité d'entrée choisie (une entrée tangible, gestuelle, via le regard ou une commande vocale par exemple), des espaces de conception existants peuvent approfondir sa description en prenant en compte des caractéristiques non spécifiques à la désambigüisation (par exemple pour l'interaction tangible [Fishkin, 2004] ou bien l'interaction via des commandes vocales [Bernsen *et al.*, 1998]). Le fait de tirer profit d'espaces de conception existants et de faciliter les interconnexions entre les différentes approches est un attribut important de notre espace de conception.

Publication liée à ce chapitre

William Delamare, Céline Coutrix et Laurence Nigay. "Designing Disambiguation Techniques for Pointing in the Physical World", in proceedings of the 5th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2013). London, UK, 2013, p. 197–206.

Dans la suite de ce chapitre, nous présentons tout d'abord les axes de notre espace de conception répartis en deux groupes (*Interaction* dans le paragraphe 4.2.1 et *Système de Désambigüisation* dans le paragraphe 4.2.2) et les liens existants entre ces deux groupes dans le paragraphe 4.2.3). Nous positionnons ensuite cet espace de conception par rapport à l'existant en mettant en valeur comment nos travaux intègrent les études précédentes, ou leur sont complémentaires (paragraphe 4.2.4).
Nous terminons l'exposé de notre espace de conception par l'étude de son pouvoir *descriptif* (capacité à décrire les techniques de désambigüisation existantes). De plus dans le chapitre suivant (chapitre 5)), nous étudions son pouvoir *génératif* (capacité à faciliter la création

de nouvelles techniques de désambigüisation).

4.2 Espace de Conception

Nous présentons les axes de notre espace de conception au regard de nos deux groupes : *Interaction* et *Système de désambigüisation*.

4.2.1 Groupe *Interaction*

Du point de vue de l'interaction, le concepteur peut exploiter deux caractéristiques : l'affichage des cibles (c'est-à-dire l'espace de représentation) et le contrôle par l'action de l'utilisateur (c'est-à-dire l'espace de contrôle)

4.2.1.1 Espace de représentation

Durant l'étape de pointage, l'utilisateur porte son attention sur les objets physiques. Durant l'étape de désambigüisation, l'utilisateur peut se concentrer sur les objets physiques ou bien porter son attention sur une représentation numérique de ces objets (les deux extrémités de l'axe 1, Figure 4.1). Entre ces deux extrémités, l'affichage peut être mixte, augmentant l'objet physique avec des informations numériques supplémentaires.

Augmenter les objets physiques ou bien utiliser une représentation numérique peut faciliter la sélection si les objets physiques sont trop petits, éloignés de l'utilisateur ou bien trop proches les uns des autres. Cette augmentation ou cette représentation numérique peut être spatialement découplée des objets physiques pointés, comme lors de l'utilisation d'une liste affichant le nom des objets pointés sur un dispositif mobile par exemple [Välkkynen *et al.*, 2006]. Toutefois, ce type de solution divise l'attention de l'utilisateur entre l'environnement physique et la représentation numérique, créant ainsi un coût supplémentaire lors du changement d'attention de l'utilisateur (discontinuité perceptive définie par [Dubois *et al.*, 2002]).

Pour surmonter cette limitation, l'information numérique peut maintenir une relation spatiale avec les objets physiques. RFIG Lamp [Raskar *et al.*, 2004] par exemple projette des disques sur les cibles physiques pointées pendant l'étape de désambigüisation. Ces disques étant plus grands et plus visibles que les capteurs sur les objets, l'utilisateur peut ensuite sélectionner l'objet désiré en visant le disque correspondant avec un laser.

Une autre façon d'éviter ce changement d'attention est de proposer une solution permettant à l'utilisateur d'effectuer la désambigüisation en se concentrant sur les objets physiques eux-mêmes par exemple. Il est envisageable qu'à l'avenir, des systèmes permettront de modifier l'agencement de l'espace physique. Par exemple, ZeroN permet de bouger des

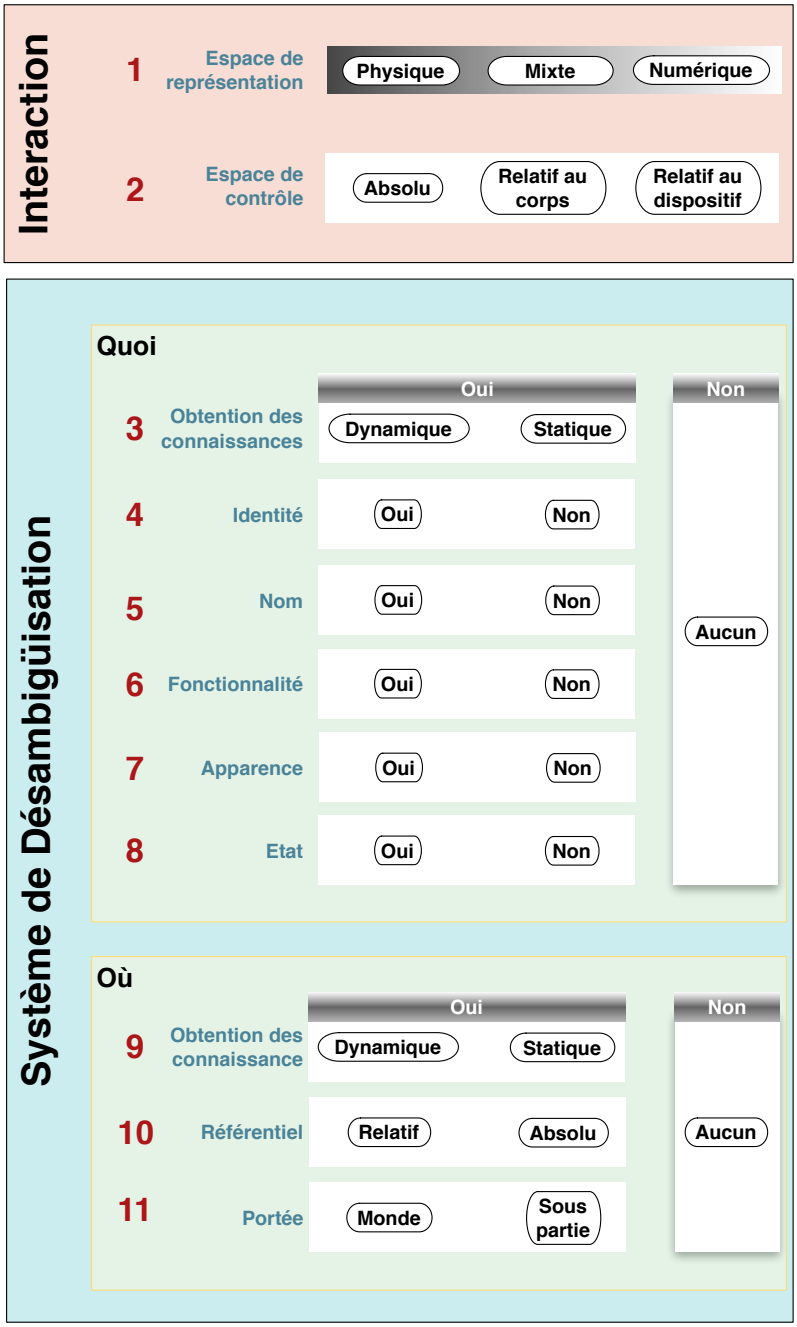


FIGURE 4.1 – Espace de conception avec 11 axes organisés selon deux groupes : *Interaction* et *Système de Désambigüisation*. Ce dernier est lui-même composé de deux sous-groupes reliés à deux questions : *Quoi* et *Où*.

objets physiques [Lee *et al.*, 2011]. Une technique de désambigüisation pourrait utiliser ce principe pour ré-agencer les cibles de façon à faciliter la sélection de l'objet désiré comme les techniques en environnements numériques 3D telles que Flower Ray [Grossman

et Balakrishnan, 2006] ou SQUAD [Kopper *et al.*, 2011]. On peut aussi envisager de nouveaux matériaux permettant d'augmenter la taille des objets physiques en vue de faciliter leur sélection [Yao *et al.*, 2013].

4.2.1.2 Espace de contrôle

La désambigüisation des objets pré-sélectionnés peut être facilitée par une modification du contrôle, c'est-à-dire des actions de l'utilisateur et de leur interprétation par le système. Cette approche peut être particulièrement utile si on suppose que l'espace d'affichage (à savoir le monde physique) ne peut être modifié.

Pour cela, le concepteur peut choisir de dissocier le contrôle des objets physiques en modifiant le référentiel dans lequel s'inscrit l'entrée de la désambigüisation (axe 2 de la Figure 4.1). Cet axe a été introduit pour caractériser un geste de pointage 3D [Cockburn *et al.*, 2011], et est aussi applicable au cas de techniques de désambigüisation. En effet, l'axe *Espace de contrôle* caractérise la manière dont le lien absolu entre le geste de pointage et les objets physiques est transformé pour l'étape de désambigüisation.

Premièrement, si l'action de l'utilisateur est dirigée vers les objets physiques, le référentiel est dit *absolu*. Dans ce cas, la technique de désambigüisation devrait se baser sur une modification de l'affichage (axe 1) afin de faciliter la sélection d'un objet.

Deuxièmement, si l'action de l'utilisateur est effectuée autour d'une partie de leur corps, le référentiel est dit *relatif au corps* de l'utilisateur. Les cibles sont dissociées des objets physiques et ré-organisées de manière abstraite selon une origine définie par une partie du corps. Par exemple, les objets physiques pointés peuvent être représentés par des cibles abstraites placées verticalement des hanches à la tête. Dans cet exemple, l'utilisateur désambigüise la sélection en bougeant son bras de bas en haut le long de son corps. L'espace de commande définit donc un espace d'entrée que l'utilisateur doit mentalement relier au monde physique.

Troisièmement, si l'action de l'utilisateur est effectuée avec un dispositif comme point de référence, le référentiel est dit *relatif à un dispositif*. Les cibles sont dissociées des objets physiques et ré-organisées de manière abstraite selon une position neutre du dispositif comme origine. Par exemple, les objets physiques peuvent être abstraits puis ré-organisés à la manière du cadrant d'une horloge. Dans cet exemple, l'utilisateur doit alors tourner le dispositif afin de sélectionner l'objet désiré. Comme dans le cas d'un référentiel *relatif au corps*, un référentiel *relatif à un dispositif* implique un espace d'entrée que l'utilisateur doit mentalement relier au monde physique.

Au delà des aspects interactionnels, notre espace de conception considère aussi les aspects techniques en incluant un ensemble d'axes noté *Système de Désambigüisation* afin de spécifier les connaissances que le système doit obtenir.

4.2.2 Groupe *Système de Désambigüisation*

Nous considérons deux catégories de connaissances que le système de désambigüisation pourrait obtenir, en vue de permettre ou de faciliter l'étape de désambigüisation. Ces connaissances concernent les objets eux-mêmes (sous-groupe *Quoi*) ainsi que leur localisation spatiale (sous-groupe *Où*).
Nous définissons tout d'abord comment le système obtient ces connaissances (de manière *dynamique* ou *statique*) puis détaillons les deux sous-groupes *Quoi* et *Où* dans le cas d'un système de désambigüisation.

4.2.2.1 *Obtention des connaissances*

Comme représentés par les axes 3 et 9 de la Figure 4.1, nous distinguons deux manières d'obtenir des informations sur le monde physique : celle obtenue lors d'une phase d'installation et celle obtenue pendant la phase d'interaction. Nous distinguons la manière d'obtenir les informations d'identité (axe 3) de celles concernant les informations spatiales (axe 9).
Les systèmes utilisant des marqueurs 2D attachés aux objets tels que Freeze-Set-Go [Lee *et al.*, 2009] ont besoin de déclarer et d'enregistrer ces marqueurs afin de pouvoir les identifier ensuite par vision par ordinateur.
Au contraire, des systèmes tels que RFIG Lamps [Raskar *et al.*, 2004] obtiennent des informations pendant que l'utilisateur est en train d'interagir. En effet, après l'étape de pointage, le système utilise des capteurs de lumière sur les objets et une projection lumineuse encodée afin de cartographier l'espace physique et ainsi avoir la position de objets pointés.

La connaissance obtenue de manière statique oblige l'utilisateur à configurer ainsi qu'à maintenir le système. Par exemple, l'utilisateur peut être amené à nommer les objets physiques interactifs, mais aussi à déterminer leur emplacement physique. Il devra aussi mettre à jour ces informations en cas de déplacement, d'ajout ou de suppression d'objets interactifs.
Même si une connaissance obtenue de manière dynamique est plus souple, elle peut cependant être moins robuste. Par exemple, RFIG Lamps permet de déterminer l'emplacement des objets pointés grâce à une projection lumineuse encodée et des capteurs de lumières attachés aux objets [Raskar *et al.*, 2004]. Ces capteurs peuvent alors être masqués du point de vue du projecteur tenu dans la main, mais visibles du point de vue de l'utilisateur.

Les informations sur l'identité des objets (sous-groupe *Quoi*) et les informations spatiales (sous-groupe *Où*) peuvent être obtenues par le système de manière différente (axes 3 et 8 Figure 4.1). Il en est de même pour chaque type d'information sur l'identité (les axes du sous-groupe *Quoi*) et chaque type d'information spatiale (les axes du sous-groupe *Où*). Cependant, une option de conception prometteuse est l'obtention de connaissance

dynamique, évitant ainsi une phase d'installation ainsi que de potentielles mises-à-jour de la part de l'utilisateur. Ainsi, nous qualifions globalement l'obtention des connaissances sur l'identité des objets (*Quoi*) et les connaissances spatiales (*Où*) de statique si l'une de leurs informations est obtenue de manière statique. Par conséquent, nous la qualifions de dynamique si toutes leurs informations sont obtenues de manière dynamique.

4.2.2.2 Sous-groupe *Quoi*

Si le concepteur choisit de fournir une représentation numérique comme espace d'affichage (axe 1 Figure 4.1), des informations à propos des objets physiques peuvent améliorer la continuité cognitive entre l'objet et sa représentation [Dubois *et al.*, 2002]. Par exemple, représenter la lampe de chevet par sa photographie plutôt que par "lampe Num. 1" améliore cette continuité cognitive. La continuité cognitive fait référence à l'interprétation que fait l'utilisateur de la représentation numérique. Les informations du sous-groupe *Quoi* ont une influence sur la représentation numérique des objets sélectionnés.
Le système peut aussi n'avoir aucune information sur les objets physiques. Nous incluons donc l'option *aucun* dans notre espace de conception. Dans le cas où le système de désambigüisation aurait certaines de ces informations, nous distinguons :

1. L'*identité* des objets (axe 4 Figure 4.1). Cela permettrait au système de savoir quels sont les objets pré-sélectionnés et ainsi de signaler à l'utilisateur que plusieurs objets sont dans le volume de sélection par exemple.
   Par exemple, en utilisant PICOntrol [Schmidt *et al.*, 2012], c'est à l'utilisateur de décider s'il doit désambigüiser ou non, puisque le système ne connait pas l'identité des objets pointés. Au contraire, avec HOBS [Zhang *et al.*, 2014], le système connait l'identité des objets et peut donc proposer l'étape de désambigüisation si plusieurs objets sont détectés dans le volume de sélection.
2. Le *nom* des objets (axe 5 Figure 4.1). Cela permettrait au système de désambigüisation d'afficher le nom des objets de manière explicite : par exemple "lampe de chevet" au lieu de "uid31286".
3. Les *fonctionnalités* des objets (axe 6 Figure 4.1). Cela permettrait au système de désambigüisation d'afficher des informations supplémentaires à l'utilisateur, par exemple si celui-ci vise deux sources de lumières, mais aimerait contrôler celle qui a une intensité modifiable. Faire savoir à l'utilisateur les fonctionnalités des objets physiques peut alors faciliter la désambigüisation en aidant à l'interprétation des représentations numériques.
4. L'*apparence* des objets (axe 7 Figure 4.1). Ce type d'information comprend la forme, la couleur, la texture, etc. Cela permettrait au système de désambigüisation de diminuer la discontinuité cognitive en réduisant l'écart entre l'objet et sa représentation graphique par exemple.
5. L'*état* des objets (axe 8 Figure 4.1). Ce type d'information comprend la consommation électrique courante, son état allumé ou éteint, etc. Par exemple, si l'utilisateur

est en train de viser deux lampes, une allumée et l'autre éteinte, avec l'intention de contrôler celle allumée, représenter l'état courant des objets visés peut alors aider à la désambigüisation. Représenter cet état peut aussi faciliter la compréhension de la représentation graphique et ainsi aider encore à la désambigüisation.

Chacun de ces axes (*identité, nom, fonctionnalités, apparence, état*) peuvent avoir plusieurs niveaux de connaissance. Pour l'*apparence* par exemple, le système de désambigüisation pourrait avoir des informations sur la forme de l'objet, mais aussi sur sa couleur (3D), ou seulement un niveau de clarté (1D), etc. Afin d'éviter un niveau de description trop fin qui rendrait difficile la description des techniques de désambigüisation ainsi que la comparaison de ces techniques entre elles, nous considérons seulement deux valeurs : *oui* et *non*. Cette approche permet une classification simple des connaissances à propos des objets, et suffit pour comparer les systèmes existants.

4.2.2.3 Sous-groupe *Où*

Comme dans l'axe *Etendue des connaissances sur le monde* initialement défini par [Milgram et Kishino, 1994], le système peut n'avoir aucune information de localisation des objets physiques (valeur *aucun* dans le sous-groupe *Où* de la Figure 4.1). Si le système maintient des informations spatiales, statiques ou dynamiques, alors nous étendons cet axe avec la définition de deux nouveaux axes :

1. *Référentiel : relatif vs absolu* (axe 10 Figure 4.1) : Dans un référentiel relatif, le système sait comment les objets sont placés les uns par rapport aux autres. Par exemple, un système pourrait utiliser une carte simplifiée du monde physique en ne tenant compte que des relations gauche/droite entre les objets. Un référentiel relatif est bien adapté à une désambigüisation utilisant une stratégie de *parcours* (paragraphe 3.3) de l'ensemble des objets pré-sélectionnés.
   Dans un référentiel absolu, le système possède des coordonnées absolues dans un référentiel commun. RFIG Lamps par exemple détermine les coordonnées (x, y) afin de projeter des informations numériques sur les objets physiques [Raskar *et al.*, 2004]. Un référentiel absolu est bien adapté à une désambigüisation utilisant une stratégie de *désignation* (paragraphe 3.3), c'est-à-dire lorsque l'utilisateur doit directement désigner l'objet voulu dans l'ensemble d'objets pré-sélectionnés.
2. *Portée : monde vs sous-partie* (axe 11 Figure 4.1) : Le système peut avoir des informations sur tous les objets du *monde* physique. Avec un système tel que l'OptiTrack[1] par exemple, composé de caméras infrarouges pour détecter des objets physiques grâce à des marqueurs réflectifs, le système peut connaître l'emplacement et l'orientation de tous les objets.
   Le système peut aussi n'avoir des informations que sur une *sous-partie* des objets du monde physique. Par exemple, HOBS [Zhang *et al.*, 2014] détermine dynamiquement des informations spatiales relatives des objets pointés seulement en

1. http://www.naturalpoint.com/optitrack/ (lien valide en septembre 2015)

se basant sur l'intensité du signal infrarouge reçu par les objets. Cette intensité permet au système de définir une relation d'ordre entre les objets en fonction de la distance les séparant du centre du rayon infrarouge.

4.2.3 Liens entre les groupes *Interaction* et *Système de Désambigüisation*

Tout en décrivant les 11 axes de notre espace de conception, nous avons mis en évidence certains liens existants entre les groupes *Interaction* et *Système de Désambigüisation*. En effet, faire des choix quant à l'interaction pour désambigüiser le pointage a des conséquences sur le système à développer, en particulier les connaissances sur le monde physique dont le système aura besoin :

— Axe 1 du groupe *Interaction* : *Espace de représentation*. En choisissant de fournir une représentation numérique des objets, le concepteur fait le choix de devoir obtenir des informations sur le monde physique afin d'améliorer cette représentation numérique et ainsi faciliter l'étape de désambigüisation. Par exemple, des informations sur l'identité des objets (axes du sous-groupe *Quoi*) peuvent réduire la discontinuité cognitive entre les objets physiques et leur représentation. Il en va de même pour les informations spatiales (axes du sous-groupe *Où*) afin d'établir une cohérence entre la représentation graphique et les emplacements physiques des objets.

— Axe 2 du groupe *Interaction* : *Espace de contrôle*. Si le concepteur choisi un référentiel autre que *absolu*, c'est-à-dire *relatif au corps* ou *à un dispositif*, des informations quant aux positions physiques des objets (axes du sous-groupe *Où*) peuvent être utiles pour coupler l'entrée de la désambigüisation aux positions des objets et ainsi minimiser les transformations spatiales que doit mentalement effectuer l'utilisateur.

Notre espace de conception sépare ces deux aspects en deux groupes d'axes correspondant à deux points de vue qui sont néanmoins liés : un ensemble d'axes pour caractériser l'interaction après l'étape de pointage, un autre pour les conditions nécessaires du système de désambigüisation correspondant. De plus, un niveau non présenté dans l'espace concerne l'impact qu'auront ces choix (interaction et système) sur les technologies à adopter pour déployer le système final.

Avec ces trois liens (interaction, système, technologie), il convient de noter que la connaissance du monde physique par le système de désambigüisation peut être une contrainte pour l'utilisateur. Nommer les objets ou bien mettre à jour des coordonnées spatiales peut être une tâche fastidieuse qu'il vaudrait mieux éviter. Toutefois, si le concepteur propose un système autonome (c'est-à-dire une valeur *dynamique* pour les axes 3 et 9), cela pourrait impliquer des infrastructures importantes et/ou trop onéreuses ou être tout simplement infaisable pour un déploiement en situation réelle.

4.2.4 Positionnement : espaces de conception existants

Dans notre espace de conception, nous avons utilisé ou adapté des axes d'espaces de conception existants issus de trois domaines :

— désambigüisation dans un environnement numérique 3D,
— interaction dans le monde physique, et
— relation entre les mondes physique et numérique.

Nous décrivons par la suite les éléments de ces espaces existants que nous mettons explicitement en relation avec les axes de notre espace de conception (paragraphe 4.2.1 et paragraphe 4.2.2).

4.2.4.1 Désambigüisation en environnements numériques

Des techniques de désambigüisation ont été étudiées avec des affichages volumiques [Grossman et Balakrishnan, 2006]. Dans ce contexte, les techniques servaient à désambigüiser la sélection faite avec un rayon pouvant intersecter plusieurs cibles numériques. Afin d'explorer plusieurs méthodes de désambigüisation interactives (chemin $(t1^*.t2.t3^*.t4)$ de l'automate de la Figure 3.2, paragraphe 3.3), les auteurs considèrent la différence entre des mécanismes *simultanés* et *séquentiels*.
Les mécanismes simultanés impliquent une phase de pointage et une phase de désambigüisation effectuées en parallèle, c'est-à-dire que l'action de $t3$ inclue du pointage et de la désambigüisation. Par exemple, le Depth Ray [Grossman et Balakrishnan, 2006] permet à l'utilisateur de contrôler l'origine et la direction du rayon (pointage) tout en manipulant un marqueur de profondeur attaché au rayon (désambigüisation). Ce marqueur est relié à la distance séparant la main de l'utilisateur de la surface de l'affichage volumique. Lors de la validation, la cible intersectée par le rayon et la plus proche de ce marqueur est sélectionnée.
Les mécanismes séquentiels impliquent une phase de pointage temporellement distincte de la phase de désambigüisation, c'est-à-dire que l'action de $t3$ n'inclue pas de pointage et que $t2$ marque la séparation entre le pointage $t1$ et la désambigüisation $t3$. Par exemple, le Lock Ray [Grossman et Balakrishnan, 2006] fige les mouvements du rayon (pointage) avant de permettre de bouger le marqueur de profondeur (désambigüisation).
Nous n'avons pas intégré cette dimension dans notre espace de conception puisque (1) elle est abstraite dans notre définition de désambigüisation interactive de notre automate et (2), la caractérisation *simultané/séquentiel* est une conséquence de choix de conception de notre espace. Par exemple, choisir une *représentation numérique* implique d'avoir effectué le pointage en amont de l'affichage. Cette dimension est donc inclue comme conséquence de nos axes.

Une autre étude définit un axe se rapprochant de la distinction *simultané/séquentiel* dans le cas de techniques de désambigüisation en environnement numérique [Bacim *et al.*, 2013]. Cet axe caractérise le *type* de la technique : *continu* ou *discret*. L'option 'continu'

peut être directement reliée à l'option *simultané* précédemment décrite. En revanche, l'option 'discret', bien que reliée à l'option 'séquentiel', permet en plus d'inclure la notion de répétition dans le mécanisme de désambigüisation. On retrouve cette notion avec l'étoile du chemin définit par notre machine à états finis Figure 3.2 pour les techniques de désambigüisation interactives : $(t1^*.t2.t3^*.t4)$.

L'espace de conception de [Bacim *et al.*, 2013] contient trois axes supplémentaires :

— L'*affichage* des objets, qui peut être fait en contexte ou hors-contexte.
Cette dimension est inclue dans notre espace de conception avec l'axe *Espace de représentation* du groupe *Interaction* (paragraphe 4.2.1.1). Plutôt que de seulement considérer les options en-contexte ou hors-contexte, notre axe explicite différentes options (physique, mixte, numérique) directement reliées à notre problématique, à savoir la désambigüisation dans *le monde physique*.

— Le *critère de désambigüisation*, caractérisant si la désambigüisation est effectuée dans le *contexte spatial d'origine*, c'est-à-dire choisir une cible dans son contexte d'origine, ou bien *hors du contexte spatiale d'origine*, comme la sélection à partir d'un menu. La désambigüisation peut aussi se baser sur des *attributs des objets*, tels que la couleur ou bien la forme.
Cet axe n'est pas orthogonal au précédent, puisque un critère de désambigüisation en contexte nécessitera à priori des objets aussi affichés en contexte par exemple. Cette dimension est néanmoins comprise dans notre axe *Espace de Contrôle* du groupe *Interaction* (paragraphe 4.2.1.2). Notre axe présente l'avantage de s'intéresser au référentiel de l'entrée utilisée pour la désambigüisation afin d'être orthogonal à l'axe *Espace de représentation*.

— La *stratégie*, qui peut être *imposée* par la technique, ou bien *choisie* par l'utilisateur.
Afficher le nom des objets à désambigüiser dans une liste par exemple impose l'étape de désambigüisation à l'utilisateur [Välkkynen *et al.*, 2006]. Au contraire, une technique permettant de contrôler l'angle d'ouverture du cône de sélection laisse le choix à l'utilisateur d'utiliser ou non cette action. C'est le cas de l'extension envisagée pour GesturePen [Swindells *et al.*, 2002] permettant de pointer des objets grâce à un rayon infrarouge.
Comme pour la dimension *simultané/séquentiel*, cette dimension résulte des choix de conception inclus dans notre espace. Par exemple, en choisissant de changer l'espace de représentation (comme avec une liste) ou bien l'espace de contrôle (comme avec un dispositif dédié), le concepteur n'a plus le choix quant à la *stratégie*.

Ces espaces de conception pour les techniques de désambigüisation en environnements numériques définissent des dimensions pertinentes qui contribuent parallèlement à notre étude, à savoir les techniques de désambigüisation dans le monde physique. Ces travaux constituent une base pour caractériser la conception et l'ingénierie de l'étape de désambigüisation interactive. Notre espace de conception inclut les dimensions pertinentes à notre problématique en adaptant les options de conception existantes afin de prendre en compte les spécificités du monde physique.

Nous nous sommes inspirés d'études portant sur un domaine parallèle au notre (interaction dans le monde physique [Cockburn *et al.*, 2011] et intégration d'affichages mixtes [Milgram et Kishino, 1994]) afin de compléter notre espace de conception en assemblant les dimensions pertinentes à l'étape de désambigüisation dans le monde physique. Ces études concernent le groupe *Interaction* [Cockburn *et al.*, 2011] ainsi que le groupe *Système de Désambigüisation* [Milgram et Kishino, 1994].

4.2.4.2 Interaction dans le monde physique

Dans leur étude, [Cockburn *et al.*, 2011] définissent un espace de conception dédié au geste de pointage effectué dans le monde physique. Parmi les cinq dimensions proposées, une d'entre elles caractérise le référentiel dans lequel est effectué le pointage vers l'objet physique. Ce référentiel peut être :

- *Absolu*, c'est-à-dire que l'utilisateur pointe vers l'emplacement physique des objets,
- *Relatif à un objet externe*, c'est-à-dire que l'emplacement pointé dépendra de l'emplacement d'un objet (comme lorsque l'on souhaite pointer un objet dans une voiture en mouvement),
- *Relatif au corps* de l'utilisateur, c'est-à-dire que le corps de l'utilisateur est considéré comme étant l'origine du référentiel dans lequel sont exécutés les mouvements (comme lorsque l'on souhaite viser à notre gauche par exemple, peu importe l'emplacement et l'orientation du corps), et
- *Relatif à un dispositif*, c'est-à-dire que le dispositif de pointage est lui-même l'origine du référentiel (comme lorsque l'on souhaite viser vers sa gauche, peu importe son emplacement).

Notre axe *Espace de contrôle* du groupe *Interaction* (paragraphe 4.2.1.2) est directement inspiré de cet axe. Cet axe décrit l'étape de désambigüisation qui suit un pointage *absolu* décrit dans [Cockburn *et al.*, 2011]. Cela permet en effet d'identifier la différence de référentiel entre les deux étapes. Nous avons cependant omis l'option *relatif à un objet externe* qui n'aurait pas eu de sens lorsque l'on souhaite désambigüiser un pointage *absolu* : si l'on reprend l'exemple du pointage dans une voiture en mouvement, la désambigüisation se fera elle aussi dans la voiture en mouvement.

4.2.4.3 Relation entre les mondes physique et numérique

L'axe *Etendue des Connaissances sur le Monde* [Milgram et Kishino, 1994] a été introduit pour caractériser les connaissances qu'avait le système du monde physique lorsque l'on cherche à interagir avec un système mixte. Plus particulièrement, l'axe caractérise l'intégration de représentations physiques et numériques, comme dans le cas de la Réalité Augmentée. Cette dimension de conception définit un continuum allant d'un monde complètement modélisé à un monde nullement modélisé (Figure 4.2). Entre les deux extrêmes, le monde peut être partiellement modélisé grâce à des informations concernant

la localisation et/ou l'identité des données du monde physique, c'est-à-dire *où* et *quoi*. Un système n'ayant que des informations de localisation pourrait précisément afficher des données numériques sur les objets physiques. Un système n'ayant que des informations sur l'identité des objets pourrait précisément définir l'augmentation de ces objets, mais sans pouvoir définir un emplacement, une orientation ou bien une échelle correctes.
Notre espace de conception intègre cet axe en l'affinant de deux manières : (1) définition de l'obtention de ces connaissances (*dynamique* et *statique*) et (2), séparation du *quoi* et du *où* afin d'obtenir une granularité plus fine sur le niveau de ces connaissances.

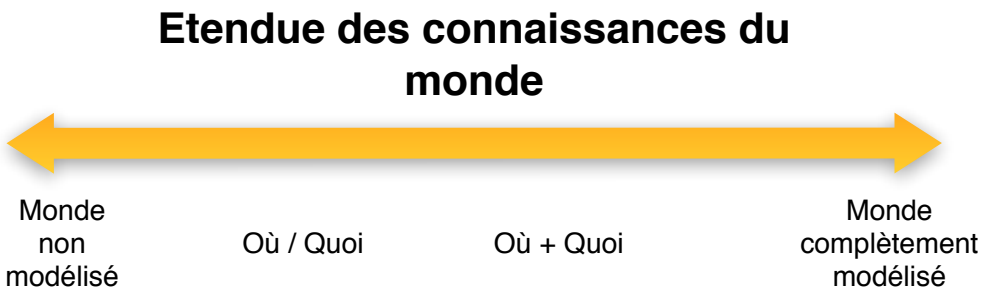


FIGURE 4.2 – Axe *Etendue des Connaissances sur le Monde* de [Milgram et Kishino, 1994].

Résumé

Nous avons présenté un espace conception composé de 11 axes organisés selon deux groupes : *Interaction* et *Système de Désambigüisation*. Le premier caractérise les changements de représentation et de contrôle pouvant s'opérer entre l'étape de pointage dans le monde physique et l'étape de désambigüisation. Le second caractérise les connaissances dont a besoin le système de désambigüisation : celles à propos des objets physiques et celles à propos de leur localisation.

Notre espace de conception adapte et étend des espaces existants pour considérer les spécificités de la désambigüisation dans le monde physique. Notre espace s'inspire de trois domaines différents : la désambigüisation en environnements numériques 3D, l'interaction dans le monde physique et la relation entre les mondes physique et numérique.

4.3 Evaluation de l'espace de conception

Afin d'évaluer notre espace de conception, nous considérons trois dimensions [Beaudouin-Lafon, 2004] :

— *Pouvoir descriptif* : capacité de l'espace de conception à pouvoir décrire toutes les techniques existantes.

— *Pouvoir évaluatif* : capacité de l'espace de conception à pouvoir aider le concepteur à choisir parmi un ensemble d'options.
— *Pouvoir génératif* : capacité de l'espace de conception à faciliter la conception de nouvelles techniques.

La complétude d'un espace de conception ne peut être définitivement prouvée. Toutefois, le pouvoir descriptif garantit que toutes les techniques de désambigüisation existantes sont décrites et distinguées au sein de notre espace. Nous détaillons ce pouvoir descriptif dans la suite de cette section.
Le pouvoir évaluatif de notre espace de conception est un travail à effectuer une fois que l'espace de conception est créé. Il est difficile de fournir des mesures qualitatives et/ou quantitatives permettant d'évaluer les options de conception tout de suite après leur définition. En effet, nous ne pouvons quantifier les options de conception puisque évaluer leurs impacts s'effectue après avoir défini les axes. Il s'agit donc d'un travail qui doit être effectué *a posteriori* en comparant différentes options de conception lors d'expériences contrôlées.
Enfin, le pouvoir génératif sera discuté en détail dans le chapitre 5, avec la conception et l'évaluation expérimentale de deux techniques de désambigüisation.

Une technique de désambigüisation pour la sélection d'un objet physique définit un chemin dans cet espace multidimensionnel. Les 11 axes composant notre espace de conception permettent alors de précisément spécifier une technique donnée en représentant son chemin correspondant ou bien de concevoir une nouvelle technique de désambigüisation en étudiant des options inexplorées. Avec notre espace de conception, les techniques de désambigüisation sont dites similaires si elles ont les mêmes caractéristiques (c'est-à-dire le même chemin) et différentes si elles présentent au moins une différence sur l'un des 10 axes. Afin de décrire et comparer des techniques de désambigüisation en utilisant notre espace de conception, nous supposons qu'une description correcte et complète est disponible pour chacune d'entre elles.

Nous décrivons ici les techniques de désambigüisation existantes au regard de nos 11 axes. Ces techniques sont chacune décrite par un chemin à la Figure 4.3.
Nous justifions par la même occasion notre intérêt pour les techniques de désambigüisation laissant l'attention de l'utilisateur dans le monde physique. Ces techniques seront discutées en détail dans le chapitre suivant (chapitre 5). En effet, nous montrons que :

1. les techniques de désambigüisation laissant l'attention de l'utilisateur dans le monde physique proposées dans la littérature n'ont été ni implémentées, ni évaluées, et
2. la technique de désambigüisation proposée le plus souvent dans la littérature est une liste affichant le nom des objets pré-sélectionnés, obligeant les utilisateurs à se concentrer sur une représentation numérique des objets.

GesturePen propose une solution de pointage basée sur un rayon infrarouge [Swindells *et al.*, 2002]. Bien que le système ne propose pas d'étape de désambigüisation dans le cas où plusieurs capteurs infrarouges répondraient à l'émission infrarouge du stylet tenu par

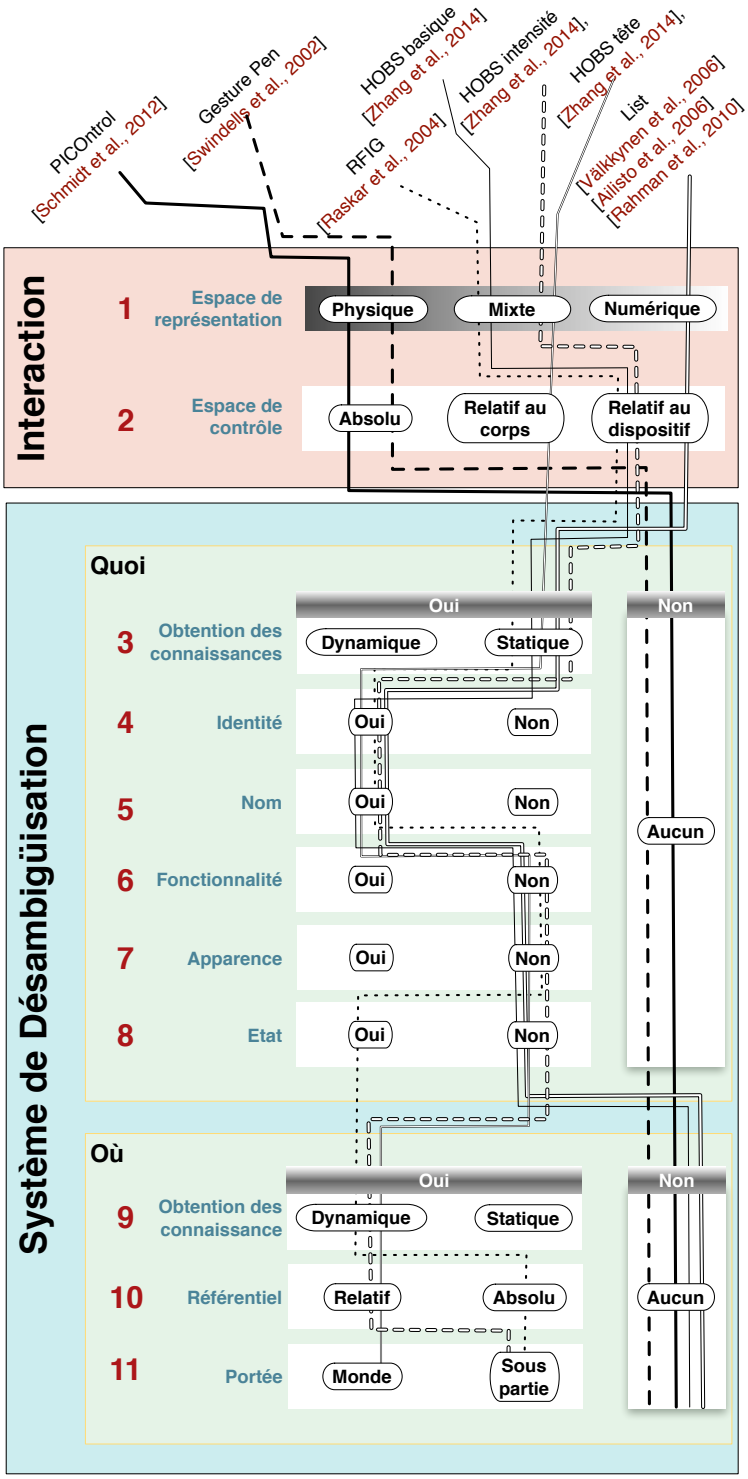


FIGURE 4.3 – Pouvoir descriptif de notre espace de conception. Les techniques de désambigüisation existantes sont représentées par des chemins.

l'utilisateur, les auteurs rapportent l'idée de l'un des participants. Ce dernier propose de rajouter un bouton afin de contrôler explicitement la longueur et/ou la largeur du rayon infrarouge émis par GesturePen. Cette solution propose donc de laisser l'attention de l'utilisateur dans le monde physique avec un référentiel absolu quant à l'espace de contrôle (l'interaction est toujours dirigée vers les objets). La technique de désambigüisation proposée pour GesturePen n'a besoin d'aucune information sur les objets physiques ou bien leur emplacement dans l'environnement.
Le fait de proposer un référentiel *absolu* pour la désambigüisation signifie que l'utilisateur peut retrouver les difficultés d'un pointage basé sur rayon s'il réduit trop l'ouverture du volume de pointage. D'autant plus qu'un faisceau infrarouge non visible à l'œil nu peut être difficile à modifier sans retour d'information visuel.
Les auteurs ont comparé expérimentalement GesturePen (pointage dans le monde physique) à la sélection d'objets physiques via une liste affichée sur un écran, forçant ainsi l'utilisateur à se concentrer sur une représentation numérique des objets. Néanmoins, (1) les participants devaient se rapprocher suffisamment des objets afin de lire leurs noms, et (2) seule l'étape de pointage a été évaluée, non l'étape de désambigüisation.

PICOntrol [Schmidt *et al.*, 2012] propose un mécanisme de désambigüisation similaire à GesturePen. PICOntrol propose de pointer les objets physiques grâce à un projecteur et des capteurs attachés aux objets. L'image projetée étant encodée, l'objet attaché à un capteur recevant l'information peut directement être contrôlé en fonction de la partie graphique positionnée sur ce capteur. Chaque élément de la projection graphique, tels que les boutons par exemple, sont encodés afin que les capteurs puissent identifier l'action à effectuer. Les auteurs proposent de réduire de manière mécanique ou optique la taille de la projection si jamais plusieurs capteurs intersectent l'interface graphique projetée.
Ce système propose donc une désambigüisation dans le monde physique avec un référentiel absolu et n'ayant besoin d'aucune information sur le monde physique. Cette solution de désambigüisation, comme celle proposée pour GesturePen, n'a cependant pas été implémentée, ni évaluée.
De plus, comme avec GesturePen, cette solution implique un référentiel *absolu* pour l'espace de contrôle pouvant amener à des problèmes de précision si la projection est trop réduite.

RFIG Lamps propose une désambigüisation fournissant de l'information numérique co-localisée avec les objets physiques [Raskar *et al.*, 2004]. RFIG Lamps utilise un pointage volumique grâce à un lecteur RFID intégré dans le dispositif de pointage. La communication établie permet alors au système de connaître l'identité des objets ainsi que des informations quant à leur état enregistrées de manière statique. Le dispositif projette ensuite deux quadrillages encodés sur les objets (sous-partie du monde) afin de dynamiquement déterminer leur position spatiale. Une fois les positions connues, le dispositif projette des cercles lumineux sur les objets physiques (espace de représentation mixte). L'utilisateur peut ensuite sélectionner le cercle voulu grâce à un pointeur laser intégré au dispositif et une stabilisation de la projection permettant ce pointage sans

modifier la position des cercles dans le monde physique.
Bien que ce système de désambigüisation permette à l'utilisateur de garder son attention sur les objets physiques, il a été conçu pour un contexte particulier, ici un entrepôt. En effet, les objets physiques concernés sont donc des boîtes en carton propices à la projection numérique. Nous ne pouvons cependant pas supposer que tous les objets physiques puissent disposer d'une surface plane adéquate. De plus, l'étape de désambigüisation faite avec le laser utilise un référentiel *absolu* - l'interaction est toujours dirigée vers les objets physiques. Or, les auteurs rapportent que la projection ainsi que la visée avec le laser souffrent de vibrations et tremblements.

HOBS propose aussi un espace de représentation mixte en affichant le nom des objets visés grâce à des lunettes de Réalité Augmentée [Zhang *et al.*, 2014]. Les auteurs proposent trois versions de leur technique de désambigüisation.
La première version (*HOBS basique*) utilise un mécanisme de désambigüisation connaissant l'identité et le nom des objets visés. Cette version basique affiche le nom de tous les objets ayant reçu le signal infrarouge (donc avec une étape de configuration) par ordre alphabétique. L'utilisateur peut ensuite naviguer dans cette liste en passant son doigt sur la partie interactive des lunettes - donc dans un référentiel relatif au dispositif.
La seconde version (*HOBS intensité*) est similaire à la version basique mais récupère dynamiquement des informations spatiales relatives grâce à l'intensité du signal reçu par les cibles - donc seulement une sous-partie du monde physique. La liste affichée peut donc réordonner les objets selon s'ils sont plus ou moins proches de la direction visée.
La troisième version (*HOBS tête*) utilise quant à elle des informations spatiales relatives du monde physique dans son entier. Ces informations sont obtenues dynamiquement sans calibration explicite de la part de l'utilisateur et sont utilisées pour créer une carte d'adjacence des objets. En effet, le système utilise l'intensité de signaux infrarouges ainsi que les capteurs de mouvements intégrés aux lunettes afin de balayer l'environnement physique et en déduire la position des objets en arrière-plan de l'activité principale de l'utilisateur. La carte d'adjacence des objets est ensuite affichée afin que l'utilisateur puisse naviguer d'un objet à l'autre grâce à des petits mouvements de la tête dans la direction voulue, donc avec un référentiel relatif au corps de l'utilisateur.
Les auteurs montrent ainsi trois résultats principaux. Premièrement, le pointage proposé grâce aux lunettes est préférable à la sélection dans une liste. Deuxièmement, un ordonnancement en 1D de l'affichage selon la distance des objets à la direction pointée est plus efficace qu'un ordonnancement alphabétique lors d'une navigation effectuée avec le doigt. Finalement, les utilisateurs préfèrent un ordonnancement 2D et une navigation effectuée avec la tête. Malheureusement, ce dernier résultat n'est validé que par une étude préliminaire et mérite d'être approfondi. De plus, ce troisième prototype (version *tête*) diffère selon deux aspects des précédents (versions *basique* et *intensité*) : l'affichage en lui-même (1D ou 2D) et le référentiel d'entrée utilisé pour la désambigüisation : *relatif au dispositif* ou bien *relatif au corps* de l'utilisateur. Il est donc difficile de juger si un référentiel est plus bénéfique que l'autre au regard des préférences des utilisateurs.

Certains systèmes proposent de présenter le nom des objets sur une liste affichée sur l'écran d'un téléphone mobile. C'est le cas du système proposé par [Välkkynen *et al.*, 2006] ou [Ailisto *et al.*, 2006] ou bien suggéré par [Rahman *et al.*, 2010]. Seulement l'une de ces techniques de désambigüisation a été implémentée [Välkkynen *et al.*, 2006]. Le mécanisme de désambigüisation utilise donc un espace de représentation numérique avec un référentiel relatif au dispositif puisque l'utilisateur doit interagir sur l'écran pour effectuer la désambigüisation. Bien que non mentionné dans leur étude, nous supposons que l'identité des objets ainsi que les noms affichés sont obtenus de manière statique par le système de désambigüisation. En effet, le système fonctionne grâce à une communication infrarouge, nécessitant donc une installation et une configuration au préalable. Nous notons cette technique "List" sur la Figure 4.3.
Malheureusement, les auteurs ont seulement rapporté la satisfaction des participants lorsqu'ils devaient désambigüiser dans la liste durant l'évaluation expérimentale.

La Figure 4.3 synthétise notre analyse des techniques de désambigüisation existantes en les représentant par des chemins au sein des options de conception de notre espace. Nous notons que (1) toutes les techniques sont décrites par notre espace de conception et (2), des techniques conceptuellement proches telles les versions d'HOBS [Zhang *et al.*, 2014] sont différentiables grâce à nos axes.
Cette représentation visuelle synthétique permet de tirer des conclusions :

- Seuls PICOntrol et GesturePen n'utilisent pas de connaissance à propos des objets physiques.
- Tous les systèmes utilisant des connaissances du groupe *Quoi* obtiennent ces connaissances de manière statique.
- Aucun système n'exploite des informations concernant les fonctionnalités ou l'apparence des objets physiques.
- Seulement deux systèmes (RFIG Lamps et *HOBS tête*) utilisent des connaissances spatiales du monde physique qui sont de plus obtenues de manière dynamique.

Cette vue synthétique de l'existant permet d'identifier de nouveaux chemins et donc de créer de nouvelles techniques de désambigüisation. L'aspect génératif de notre espace de conception fait l'objet du chapitre suivant.

Résumé

Nous évaluons notre espace de conception en nous concentrant sur deux dimensions :

- Nous illustrons le pouvoir descriptif en classant toutes les techniques de désambigüisation existantes dans notre espace de conception en les représentant grâce à des chemins. Cette représentation permet d'identifier les différences et similarités entre les techniques. Elle permet aussi d'identifier des options inexplorées jusqu'ici.
- Nous illustrons brièvement le pouvoir génératif en tirant rapidement des conclusions de la représentation visuelle synthétique fournie par notre espace

de conception. Le chapitre suivant traite du pouvoir génératif de notre espace avec la conception et l'évaluation de deux nouvelles techniques de désambigüisation.
— Le pouvoir évaluatif est une perspective pour de futurs travaux.

Nous avons aussi montré que :
— les techniques de désambigüisation existantes laissant l'attention de l'utilisateur dans le monde physique n'ont été ni implémentées, ni évaluées, et peuvent théoriquement souffrir des même limitations qu'un pointage absolu, et
— la technique de désambigüisation récurrente est une liste servant à afficher le nom des objets, obligeant ainsi les utilisateurs à se concentrer sur une représentation numérique des objets physiques.

4.4 Conclusion

Nous avons présenté et décrit un espace de conception dédié aux techniques de désambigüisation suivant un geste de pointage dans le monde physique. Cet espace de conception se veut être un support pour la conception et la recherche liées aux interactions avec et dans le monde physique. Contrairement à une approche basée sur la technologie, notre espace de conception caractérise la sélection d'objets physiques en considérant deux points de vue complémentaires : l'interaction et les connaissances du système de désambigüisation à propos des objets physiques augmentés (informations sur l'identité ou l'emplacement des objets).

Notre espace de conception capitalise les travaux effectués dans différents domaines. En particulier, pour l'aspect lié à l'interaction, notre espace de conception affine les deux principales approches pour faciliter le pointage dans une interface graphique : l'espace de représentation et l'espace de contrôle [Balakrishnan, 2004]. Pour l'aspect lié au système, notre espace de conception étend des taxonomies existantes dédiées aux systèmes de réalités mixte [Milgram et Kishino, 1994]. Cela contribue à renforcer la pertinence des concepts intégrés à notre espace de conception.

Nous avons souligné le pouvoir descriptif de notre espace de conception en classant précisément les techniques existantes. Ce travail doit être mis à jour puisque de nouvelles techniques sont susceptibles d'apparaitre à l'avenir. Pour le moment, la description des techniques existantes est réussie puisque toutes sont décrites par un chemin et toutes sont décrites par un chemin différent parmi nos options de conception.

Le chapitre suivant traite du pouvoir génératif de notre espace de conception. En effet notre espace permet de créer de nouveaux chemins, c'est-à-dire de nouvelles techniques, en

créant des combinaisons d'options encore non explorées. Pour cela, nous nous concentrons sur le groupe *Interaction* et ses deux axes : *espace de représentation* et *espace de contrôle*. Nous voulons étudier l'impact de l'attention de l'utilisateur (c'est-à-dire sur les objets physiques ou bien une représentation graphique) sur les performances obtenues avec différentes techniques de désambigüisation. Pour étudier cet impact, nous présentons deux nouvelles techniques laissant l'attention de l'utilisateur dans le monde physique mais avec des référentiels différents qui soient autre que le référentiel absolu : *relatif au corps* et *relatif au dispositif*.

5 Techniques de désambigüisation proposées : P2Roll et P2Slide

Sommaire

5.1 Introduction **48**
5.1.1 Rechercher des chemins inexplorés dans notre espace de conception 48
5.1.2 Propriétés des actions 48
5.1.3 Désambigüisation et loi de Fitts 50
5.2 Conception de P2Roll et P2Slide **51**
5.2.1 Quasi-mode 51
5.2.2 Physical Pointing Roll : P2Roll 52
5.2.3 Physical Pointing Slide : P2Slide 54
5.3 Evaluation expérimentale **56**
5.3.1 Technique de référence : liste 56
5.3.2 Hypothèses 57
5.3.3 Equipement 58
5.3.4 Participants 59
5.3.5 Procédure 59
5.3.6 Protocole expérimental 60
5.3.7 Résultats 61
5.3.7.1 H1a : Garder l'attention de l'utilisateur dans le monde physique est plus efficace que de forcer l'utilisateur à porter son attention sur une représentation numérique 62
5.3.7.2 H1b : Les utilisateurs préfèrent garder leur attention dans le monde physique plutôt que de porter leur attention sur une représentation numérique 66
5.3.7.3 H2 : Les petits mouvements ne sont pas forcément plus efficaces 66
5.4 Conclusion sur l'évaluation expérimentale **67**
5.5 Travaux futurs à partir de l'évaluation expérimentale **68**

5.1 Introduction

Afin d'illustrer le pouvoir génératif de notre espace de conception (chapitre 4), nous l'utilisons pour créer de nouvelles techniques de désambigüisation. Pour cela, nous nous concentrons sur l'exploration du groupe *Interaction* (paragraphe 4.2.1) et de ses deux axes : *Espace de représentation* et *Espace de contrôle*.

Nous présentons tout d'abord les concepts importants considérés pour la conception et l'évaluation de nos deux nouvelles techniques de désambigüisation. Par la suite, nous présentons nos deux techniques (paragraphe 5.2) ainsi que leur évaluation expérimentale (paragraphe 5.3). Nous terminons en concluant sur nos travaux (paragraphe 5.4) et en présentant des pistes de travaux futurs (paragraphe 5.5).

5.1.1 Rechercher des chemins inexplorés dans notre espace de conception

Nous fixons les axes du groupe *Système de Désambigüisation* afin de ne faire varier que les axes du groupe *Interaction*. Pour cela, nous considérons que notre système peut avoir accès à toutes les informations sur les objets et leurs positions (paragraphe 4.2.2). Nous pouvons ainsi évaluer expérimentalement les options de conception du groupe *Interaction* sans considérer les conséquences sur le groupe *Système de Désambigüisation* décrites dans le paragraphe 4.2.3.

Nous voulons étudier l'effet qu'a l'attention de l'utilisateur sur les performances de la désambigüisation (axe 1, espace de représentation, paragraphe 4.2.1.1). En effet, l'étape de désambigüisation peut être effectuée en laissant l'attention de l'utilisateur sur les objets physiques pointés, ou bien sur une représentation numérique de ces objets, comme une liste affichée sur un écran. Nous cherchons à connaître les avantages et inconvénients de ces deux options (les extrémités de l'axe 1).

Le second aspect lié à l'interaction mis en évidence par notre espace de conception est l'espace de contrôle (axe 2, paragraphe 4.2.1.2). Nous voulons explorer les deux options permettant de ne plus se placer dans un référentiel *absolu* (c'est-à-dire dirigé vers les objets physiques), à savoir un référentiel *relatif au corps* ou bien *relatif à un dispositif*. Afin de choisir des modalités d'interaction en entrée correspondant à ces deux options, il nous faut examiner les propriétés physiques des membres du corps pour décider des actions de l'utilisateur pour la désambigüisation.

5.1.2 Propriétés des actions

Les performances d'une action donnée dépendent du membre avec lequel elle est exécutée, mais aussi des articulations impliquées ou les muscles utilisés. Cependant, en nous

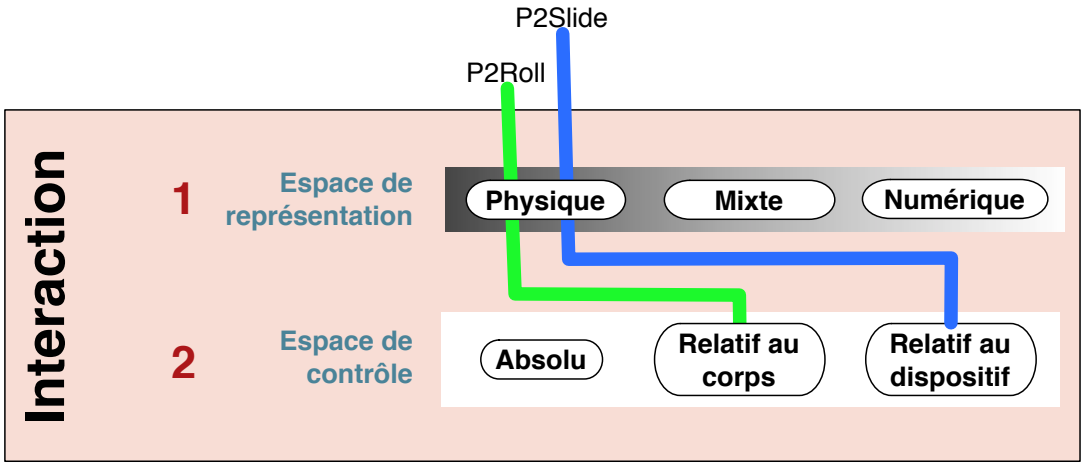


FIGURE 5.1 – Exploration de notre espace de conception. Axe 1 : création de technique de désambigüisation laissant l'attention de l'utilisateur dans le monde physique. Axe 2 : exploration de deux options non évaluées.

basant sur la morphologie du membre supérieur, puisque l'étape de pointage implique les articulations de l'épaule et du coude, nous nous intéressons aux articulations non-utilisées situées au niveau du poignet et des doigts.

Il existe un compromis entre la vitesse et la précision des actions faisant intervenir un membre donné. Cela dépend de :

— la bande passante du membre [Card *et al.*, 1991], c'est-à-dire la quantité d'information que les actions exécutées avec ce membre peuvent transmettre, et
— les tremblements [Davis et O'Connell, 2000] que les utilisateurs peuvent avoir en utilisant ce membre.

Il est difficile de prédire la bande passante et les tremblements de chaque membre [Balakrishnan et MacKenzie, 1997]. En effet, les résultats existants souffrent d'un grand nombre de différences entre les tâches, les protocoles d'expérimentation, les effets de bord tels que l'apprentissage entre les conditions d'expérimentation ainsi que les muscles et les articulations concernés par les études. Il est donc difficile d'être certain quant au membre à choisir pour une tâche donnée.

Il apparait que même avec des résultats non comparables entre eux, le poignet [Balakrishnan et MacKenzie, 1997] et les doigts [Zhai *et al.*, 1996] ont été identifiés comme de bons candidats en entrée d'interaction. Les deux techniques que nous avons conçues exploitent l'interaction avec le poignet et le doigt : P2Roll (Physical Pointing Roll) utilise un roulis du poignet pour naviguer parmi les cibles pré-sélectionnées, et P2Slide (Physical Pointing Slide) utilise un mouvement du doigt pour naviguer parmi les cibles pré-sélectionnées (Figure 5.1).

5.1.3 Désambigüisation et loi de Fitts

La loi de Fitts permet la modélisation d'une tâche de pointage [Fitts, 1954]. Bien que son lien avec la théorie de l'information puisse être critiquée [Drury *et al.*, 1987], la loi de Fitts a conduit à de nombreuses études dans le domaine de l'IHM, que ce soit pour valider le modèle mathématique dans des conditions particulières (comme une entrée via un geste de roulis du poignet [Crossan *et al.*, 2008]), pour l'étendre à de nouveaux contextes (comme le pointage à distance [Kopper *et al.*, 2010] ou bien le pointage via un écran de téléphone [Rohs et Oulasvirta, 2008]) ou encore pour améliorer des techniques de pointage [Balakrishnan, 2004]. C'est aussi grâce à cette loi que la bande passante d'un membre peut être étudiée.

Nous cherchons avant tout à savoir si les techniques laissant l'attention de l'utilisateur dans le monde physique méritent d'être étudiées plus en détail. Pour cela, notre évaluation expérimentale n'impose aucune contrainte quant à la position de départ du pouce ou bien la rotation de départ du poignet. Or, ces contraintes sont nécessaires lorsque l'on souhaite modéliser une technique de pointage grâce à la loi de Fitts. Cette modélisation est une perspective de recherche intéressante dans le cas où nos techniques se révèlent appréciées des utilisateurs et montrent de meilleures performances que les techniques déjà existantes, c'est-à-dire celles qui imposent un changement d'attention de la part de l'utilisateur, des objets physiques vers une représentation graphique.

Résumé

Notre objectif au regard de notre espace de conception est de concevoir des techniques de désambigüisation laissant l'attention de l'utilisateur dans le monde physique. Les deux techniques conçues - P2Roll et P2Slide - utilisent deux modalités d'interaction en entrée différentes pour l'étape de désambigüisation suivant un pointage avec le bras : respectivement mouvements du poignet et mouvements du pouce.
Cela se traduit par une méthode de conception systématique dans notre espace de conception via l'exploration des options des axes du groupe *Interaction*, à savoir l'*Espace de représentation* (physique *vs* numérique) et l'*Espace de contrôle* (relatif au corps vs relatif à un dispositif).

A notre connaissance, il n'existe pas d'étude explorant les avantages et inconvénients d'une attention visuelle maintenue sur les objets physiques pendant l'étape de désambigüisation. De plus, comme décrit précédemment dans le paragraphe 4.3, les techniques de désambigüisation laissant l'attention sur les objets physiques (GesturePen [Swindells *et al.*, 2002] et PICONtrol [Schmidt *et al.*, 2012]) (1) ont choisi une technique de désambigüisation perdant les avantages de la sélection volumique en retournant vers un pointage basé sur un rayon et ses potentiels inconvénients et (2) n'ont été ni implémentées ni évaluées. C'est le défi que nous désirons relever : garder les avantages d'une sélection volumique en

explorant différentes options pour l'espace de contrôle (référentiel *relatif au corps* ou bien *relatif à un dispositif*) pour la conception de nos techniques P2Roll et P2Slide.

Dans la section suivante, nous décrivons nos deux techniques visant à maintenir l'attention de l'utilisateur dans le monde physique.

Publication liée à ce chapitre

William Delamare, Céline Coutrix et Laurence Nigay. "Mobile Pointing Task in the Physical World : Balancing Focus and Performance while Disambiguating", in proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2013). Munich, Germany, 2013, p. 89–98.

5.2 Conception de P2Roll et P2Slide

Nous présentons les deux techniques de désambigüisation conçues pour la sélection d'objets physiques augmentés. Nous supposons que ces objets permettent de fournir un retour d'information visuel avec une ou des LEDs. Cette augmentation est communément adoptée dans de nombreux travaux [Swindells *et al.*, 2002] [Zhang *et al.*, 2014].

5.2.1 Quasi-mode

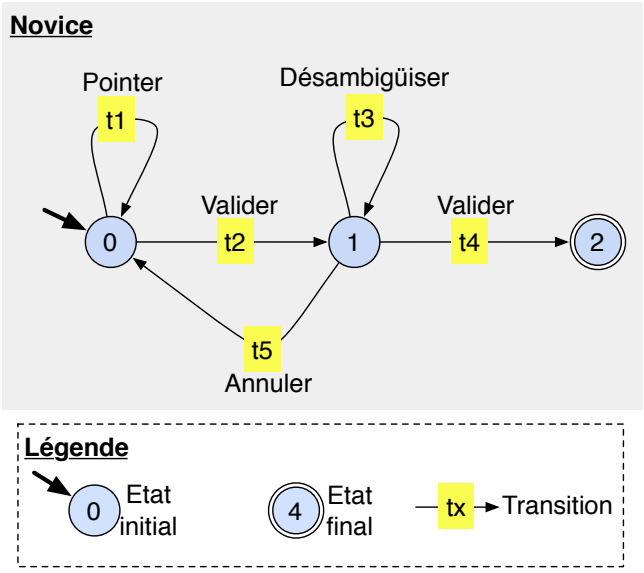


FIGURE 5.2 – Automates de nos techniques illustrant le comportement attendu d'utilisateurs novices.

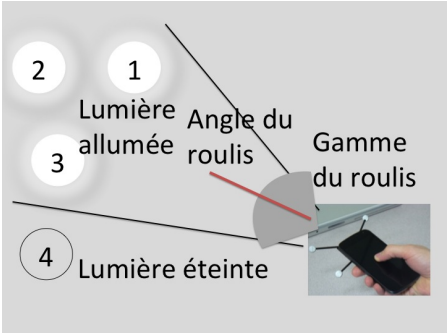
Les deux techniques utilisent des actions séquentielles pour le pointage et la désambigüisation grâce à un smartphone : $t1$ et $t3$ séparés par $t2$ sur la Figure 5.2. Le pointage volumique se fait en pointant dans la direction de l'objet physique à sélectionner avec un smartphone. L'étape suivante de désambigüisation peut être considérée comme un quasi-mode : l'utilisateur entre dans ce mode en touchant et en maintenant son doigt sur l'écran tactile du téléphone. Dans ce mode, la direction du pointage est figée, c'est-à-dire que le cône de sélection ne bouge plus en fonction de la direction visée par le téléphone. L'utilisateur quitte le mode (c'est-à-dire valide sa sélection) en retirant son doigt de l'écran ($t4$), déverrouillant ainsi la direction du pointage.

Si jamais la cible voulue ne se trouve pas dans le volume de sélection lorsque l'utilisateur entre dans l'étape de désambigüisation, il peut alors quitter le mode en glissant son doigt vers le bas de l'écran ($t5$). Ainsi, lorsque le doigt ne touchera plus l'écran, la sélection ne sera pas validée et la direction du pointage sera déverrouillée.

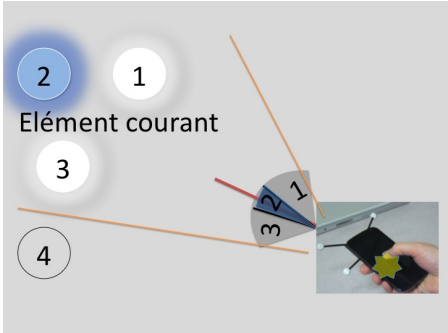
5.2.2 Physical Pointing Roll : P2Roll

P2Roll utilise un roulis du poignet afin de naviguer parmi les éléments pré-sélectionnés durant l'étape de pointage (Figure 5.3). Après le geste de pointage (Figure 5.3a), l'utilisateur peut verrouiller la direction pointée en touchant l'écran du téléphone (Figure 5.3b). A ce stade, les objets dans le volume de sélection sont associés à des plages angulaires de sorte que lorsque l'angle du poignet est dans une zone donnée, l'objet associé fournit un retour d'information visuel à l'utilisateur en s'éclairant d'une couleur différente (bleue) de celle utilisée pour signaler que les objets sont dans le volume de sélection (blanche). Rien n'est affiché sur l'écran, l'utilisateur se concentre uniquement sur les objets physiques et leur retour d'information. Une fois que l'objet voulu est l'objet courant, c'est-à-dire que l'angle du poignet est dans la zone angulaire correspondant à cet objet, l'utilisateur peut valider sa sélection en retirant le pouce de l'écran.

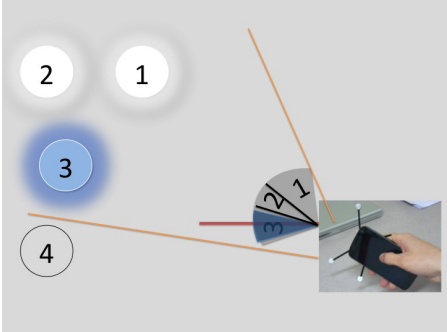
Des travaux précédents ont montré que les utilisateurs pouvaient confortablement utiliser le roulis du poignet sur 130 ° en tenant un laser, à savoir entre -70 ° et 50 ° [Qin *et al.*, 2010]. Toutefois, il y a une différence entre tenir un laser et tenir un téléphone en main. Nous avons donc conduit une étude pilote (7 participants) montrant que les sujets droitiers utilisaient confortablement un roulis entre -80 ° et 50 ° . Nous avons par la suite utilisé des valeurs symétriques pour les gauchers.
Nous avons discrétisé cette gamme angulaire de façon quadratique (Figure 5.3b, Figure 5.3c, Figure 5.3d). Il a en effet été montré qu'une discrétisation quadratique permettait une meilleur précision qu'une discrétisation linéaire pour le roulis du poignet [Rahman *et al.*, 2009]. Il en résulte qu'au lieu d'avoir des zones angulaires de même taille, les zones sont de plus en plus larges à mesure qu'elles s'écartent du centre. Nous avons empiriquement optimisé cette discrétisation puis l'avons validée via des expériences pilotes. Enfin, chaque zone discrétisée est ensuite reliée à un objet physique et est ordonnée selon la position



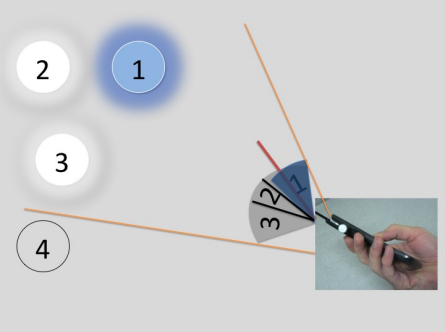
(a) Pointer : volume de sélection



(b) Toucher : volume de sélection figé



(c) Roulis vers la gauche : l'élément sur la gauche devient l'élément courant



(d) Roulis sur la droite : l'élément sur la droite devient l'élément courant

FIGURE 5.3 – Illustration de P2Roll. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. L'élément courant est l'objet 2 puisque l'angle du roulis est dans sa zone. (c) L'utilisateur effectue un roulis vers la gauche. L'objet 3 devient alors l'élément courant. (d) L'utilisateur effectue un roulis vers l'extrémité droite de sorte que l'élément courant passe de 3 à 1 en passant par 2.

des objets sur l'axe horizontal perpendiculaire à la direction du pointage.

Nous faisons l'hypothèse que les utilisateurs expérimentés (c'est-à-dire habitués à la technique et à leur environnement physique) pourront anticiper l'angle du roulis du poignet de façon à ce que lorsqu'ils terminent l'étape de pointage, cet angle soit directement dans la zone angulaire correspondant à l'objet voulu (Figure 5.4). Ce serait un gain de temps puisqu'une simple action de *toucher/relâcher* $t3$ serait nécessaire pour la sélection, les deux actions (pointage et roulis) étant effectuées en parallèle $(t1^*.t2^*)^*$. Dans le monde réel, ce comportement expert peut être attendu dans des situations récurrentes, telles que la sélection de la télévision depuis le canapé.
Il est à noter que seules les *actions physiques*, sont effectuées en parallèle, c'est-à-dire le geste de pointage et le geste de roulis, et non les étapes de pointage et de désambigüisation elles-mêmes, telles que décrit lors de la distinction *simultanée/séquentielle* du paragraphe 4.2.4.1. Bien que nous cherchions à exploiter la proprioception de l'utilisateur afin que celui-ci ait l'impression d'effectuer les deux étapes (pointage et désambigüisation)

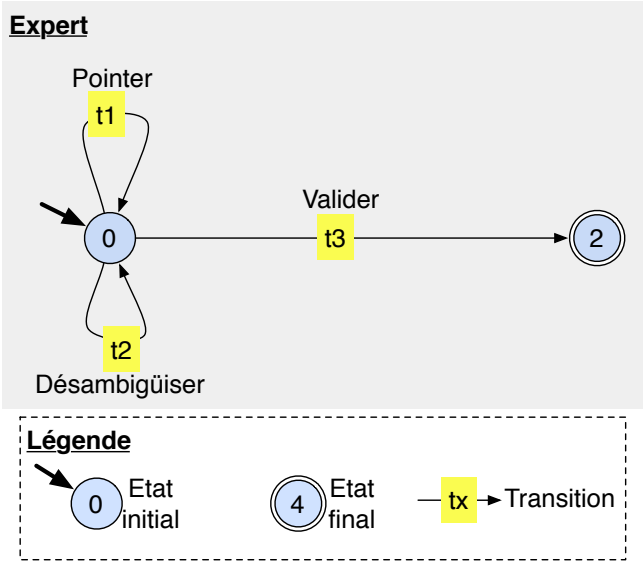


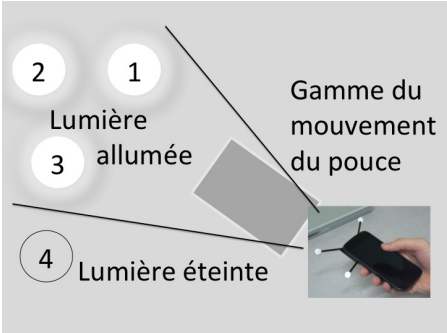
FIGURE 5.4 – Automates de nos techniques illustrant le comportement attendu d'utilisateurs experts.

en même temps, ces deux étapes restent en essence séparées. En effet, le roulis du poignet effectué pendant l'étape de pointage ne permet pas de faire une désambigüisation puisque (1) il n'existe aucune relation entre le roulis et les objets physiques tant que le pointage n'est pas validé et (2) aucun retour d'information ne peut être fourni à l'utilisateur.

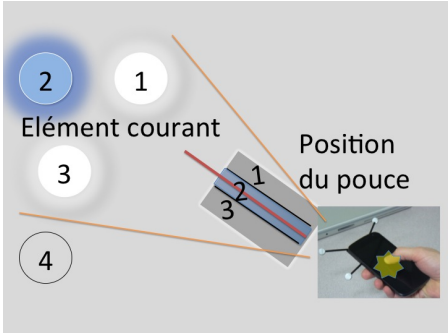
5.2.3 Physical Pointing Slide : P2Slide

P2Slide utilise un glissement horizontal du pouce sur l'écran du téléphone afin de naviguer parmi les éléments pré-sélectionnés (Figure 5.5). Nous avons choisi un mouvement horizontal plutôt que vertical puisqu'il a été montré que le pouce avait une bande passante équivalente ou meilleure avec des mouvements horizontaux [Wobbrock *et al.*, 2008]. Comme P2Roll, rien n'est affiché sur l'écran du téléphone : la navigation s'effectue avec une entrée sur le dispositif et une sortie sur les objets physiques augmentés.
Après le geste de pointage (Figure 5.5a) allumant les objets visés en blanc, l'utilisateur peut verrouiller le volume de sélection en touchant l'écran tactile (Figure 5.5b). A ce stade, les objets physiques sont associés à une zone verticale de l'écran. Ces zones verticales sont horizontalement réparties de manière égale sur toute la largeur de l'écran. Chaque zone est ensuite reliée à un objet physique et ordonnée selon l'axe horizontal perpendiculaire à la direction du pointage. Lorsque le doigt est dans une zone donnée, l'objet associé fournit un retour d'information visuel en s'allumant avec une couleur bleue (Figure 5.5b, Figure 5.5c, Figure 5.5d).

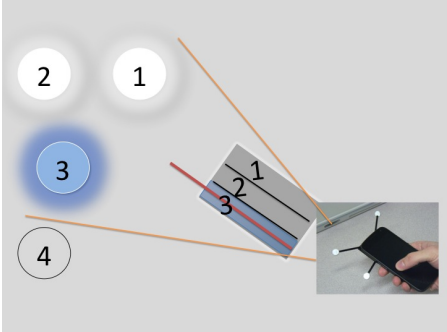
Comme pour P2Roll, nous faisons l'hypothèse qu'un utilisateur expert pourra anticiper



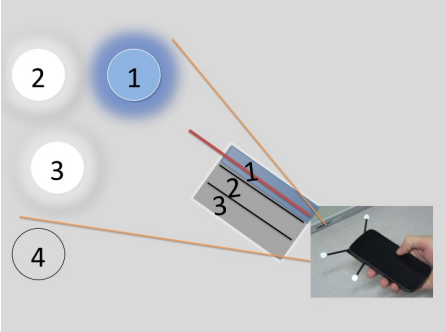
(a) Pointer : volume de sélection



(b) Toucher : volume de sélection figé



(c) Glissement du pouce vers la gauche : l'élément sur la gauche devient l'élément courant



(d) Glissement du pouce vers la droite : l'élément sur la droite devient l'élément courant

FIGURE 5.5 – Illustration de P2Slide. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. L'élément courant est l'objet 2 puisque le pouce est dans sa zone. (c) L'utilisateur effectue un glissement du pouce vers la gauche. L'objet 3 devient alors l'élément courant. (d) L'utilisateur effectue un glissement du pouce vers l'extrémité droite de sorte que l'élément courant passe de 3 à 1 en passant par 2.

la position du doigt sur l'écran pour sélectionner une zone avant même de finir l'étape de pointage (Figure 5.4). Ainsi, une simple action *toucher/relâcher* dans la bonne zone sera requise pour sélectionner un objet.

Résumé

Nous avons présenté la conception de deux techniques de désambigüisation laissant l'attention de l'utilisateur dans le monde physique :

- P2Roll (Physical Pointing Roll) utilise un roulis du poignet pour naviguer parmi les éléments pré-sélectionnés durant l'étape de pointage, et
- P2Slide (Physical Pointing Slide) utilise un mouvement horizontal du doigt sur l'écran tactile d'un téléphone pour naviguer parmi les éléments pré-sélectionnés durant l'étape de pointage.

Les deux techniques utilisent un quasi-mode permettant de figer le volume de

sélection pendant l'étape de désambigüisation.

5.3 Evaluation expérimentale

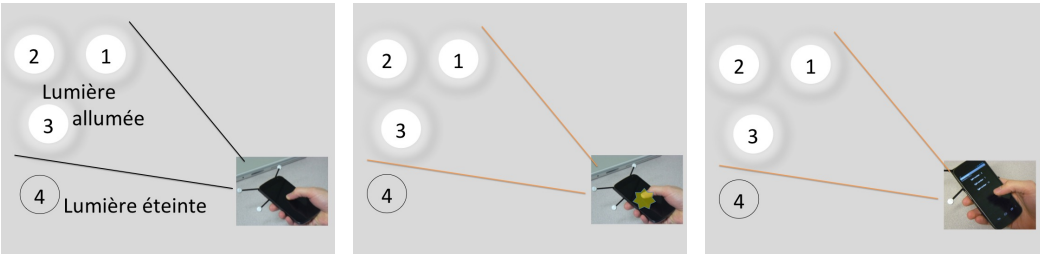
Le but de notre expérimentation est d'évaluer la relation entre l'attention de l'utilisateur et les performances durant la phase de désambigüisation de pointage. Nous étudions donc les différences de performance de deux classes de techniques :

— Les techniques qui impliquent un changement d'attention de l'utilisateur vers une représentation numérique des objets physiques. Une technique communément utilisée, à savoir une liste de noms affichée sur l'écran d'un dispositif mobile, est utilisée comme point de comparaison.

— Les techniques qui maintiennent l'attention de l'utilisateur sur les objets physiques. Dans notre expérimentation, nous avons évalué P2Roll et P2Slide. Pour cette classe de techniques, nous considérons aussi le second axe du groupe *Interaction* de notre espace de conception : l'espace de contrôle :
  - P2Roll utilise un référentiel *relatif au corps*, à savoir le poignet,
  - P2Slide utilise un référentiel *relatif au dispositif*, à savoir l'écran tactile du téléphone.

Les performances des techniques de désambigüisation sont évaluées à travers le temps de sélection et le taux d'erreur. Nous considérons aussi plusieurs niveaux de difficulté : la densité d'objets physiques dans l'environnement.

5.3.1 Technique de référence : liste



(a) (b) (c)

FIGURE 5.6 – Illustration de la technique de référence Liste. (a) L'étape de pointage. (b) Le début de l'étape de désambigüisation lorsque le doigt touche l'écran. A ce stade, le pointage est figé. (c) L'apparition de la liste sur l'écran du téléphone.

L'état de l'art fait ressortir une technique de désambigüisation couramment proposée, à savoir l'affichage du noms des objets pré-sélectionnés dans une liste sur dispositif mobile (paragraphe 4.3). Nous avons donc choisi cette technique comme point de comparaison afin d'évaluer nos techniques de désambigüisation.

Tout d'abord, l'utilisateur pointe vers une cible (Figure 5.6a). Une fois que le volume de sélection est figé par un contact sur l'écran (Figure 5.6b), l'utilisateur peut regarder ce qui est affiché dans une position confortable sans avoir à maintenir son geste de pointage (Figure 5.6c). L'utilisateur peut ensuite appuyer sur le nom voulu ou bien faire défiler la liste jusqu'à ce que le nom apparaisse à l'écran pour être sélectionné. Si le pointage doit être annulé (cible non présente dans le volume sélectionné), l'utilisateur peut appuyer sur le bouton physique "retour" du téléphone et recommencer le pointage (Figure 5.7).
Nous avons choisi une représentation des objets via des noms textuels puisque (1) des images auraient pu être mal interprétées, d'autant plus si les objets se ressemblent et (2), des noms textuels peuvent facilement être classés alphabétiquement afin de faciliter la recherche dans la liste. Nous avons choisi l'ordre alphabétique qui permet d'avoir un affichage cohérent entre chaque sélection. Ce choix est donc favorable à cette technique. Pour le développement de la liste en elle-même, nous avons utilisé le widget standard Android *liste verticale* composée d'éléments de taille standard sur Android. Ainsi, l'utilisateur n'a besoin de faire défiler la liste que lorsqu'elle contient plus de huit éléments.

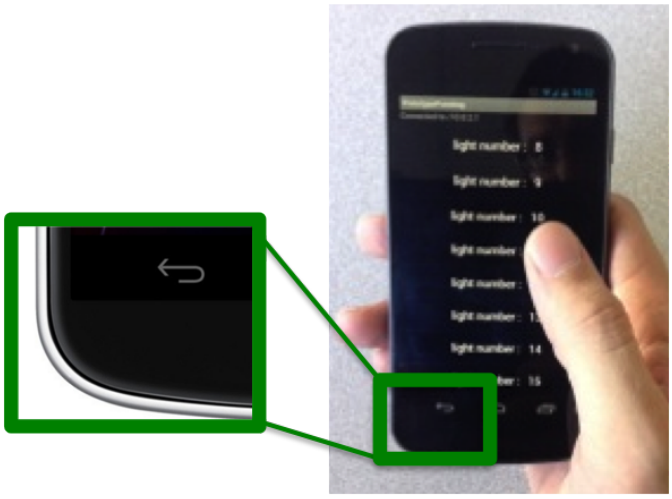


FIGURE 5.7 – Bouton physique *retour* par défaut du téléphone pour annuler l'étape de pointage avec la liste.

5.3.2 Hypothèses

Nous formulons les trois hypothèses suivantes pour notre expérience. Deux hypothèses concernent la comparaison entre nos classes de techniques (selon l'attention de l'utilisateur, axe 1 de notre espace de conception) et une hypothèse concerne la comparaison entre nos deux techniques (selon le référentiel utilisé en entrée, axe 2 de notre espace de conception).

— *H1a* :Les techniques laissant l'attention de l'utilisateur sur les objets physiques ont de meilleures performances que la technique forçant l'utilisateur à se concentrer sur une représentation numérique des objets.

— *H1b* : Les utilisateurs préfèrent garder leur attention sur les objets physiques pendant l'étape de désambigüisation.
— *H2* : La technique utilisant des mouvements de petite amplitude, à savoir un glissement du pouce (P2Slide) a de meilleures performances que la technique utilisant une plus grande amplitude dans le mouvement, à savoir un roulis du poignet (P2Roll).

5.3.3 Equipement

Les cibles sont des balles de ping-pong intégrant des LEDs RVB contrôlées par des micro-contrôleurs Arduino [1]. Les balles de ping-pong ont deux rôles : diffuser la lumière des LEDs et permettre l'inscription de nombres servant de noms des objets affichés par la liste (Figure 5.8). Nous avons choisi des numéros de façon à ce qu'ils puissent être lus de plus loin que du texte (assez de place sur les balles de ping-pong) et être facilement ordonnés dans la liste. Ces numéros sont placés au hasard, sans aucune signification quant à leur disposition.

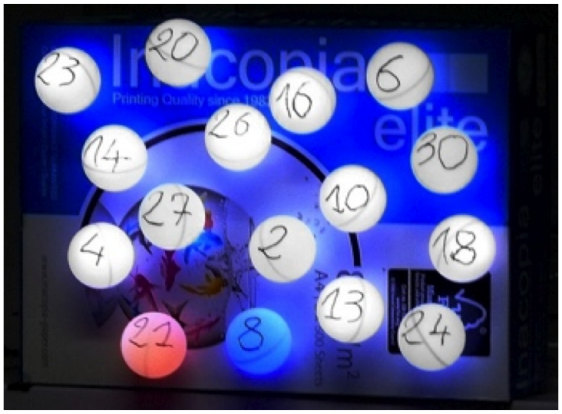


FIGURE 5.8 – Plateau avec 16 objets physiques (balles de ping-pong) intégrant des LEDs. La numéro 21 (rouge) représente la cible de l'essai en cours, la numéro 8 (bleue) représente l'élément sélectionné courant.

La désambigüisation est utile lorsque plusieurs objets sont pré-sélectionnés par le pointage. Nous voulons évaluer l'impact du nombre d'objets présents dans le volume de sélection. Afin de fournir des densités de cibles différentes, des ensembles de 2, 4, 8 et 16 balles sont fixés sur des panneaux de 31 cm x 22 cm. Les mêmes densités ont été utilisées pour évaluer des techniques de pointage en Réalité Augmentée via des lentilles magiques (Magic Lens) [Rohs *et al.*, 2009]. Nous avons choisi 16 objets pour la densité maximale puisque des travaux existants ont montré que c'était la limite maximale pour un roulis du poignet [Rahman *et al.*, 2009]. Cette limite convient aussi à P2Slide : la précision que l'on peut atteindre avec l'index sur une surface est de 150 dpi, soit des cibles de 17

1. http://www.arduino.cc/ (lien valide en septembre 2015)

mm [Berard et Rochet-capellan, 2012]. Malheureusement, ce résultat concerne l'index et non le pouce. Néanmoins, la limite de 16 cibles implique une discrétisation de l'écran en zones de 3,5 mm, soit 20 fois plus grandes que la limite pour l'index. Nous faisons l'hypothèse que c'est suffisant pour le pouce. Des études précédentes ont considéré des densités de 4 [Rukzio *et al.*, 2007] [Välkkynen *et al.*, 2006] ou 5 objets [Swindells *et al.*, 2002]. Une enquête auprès de 13 personnes nous indique que 6 objets sont susceptibles d'être désambigüisés chez eux en moyenne (écart type = 2). 16 objets dans le volume de sélection représentent donc un maximum réaliste dans un habitat d'aujourd'hui.

Afin de capturer la position et l'orientation du téléphone, nous utilisons le système OptiTrack [2] avec 5 caméras V100 :R2 et le logiciel Arena [3]. Le système fournit des informations sur des marqueurs attachés au téléphone à 100Hz et fonctionne sur un poste de travail Core 2 duo 3,06 Ghz. Ces marqueurs permettent d'obtenir la position et l'orientation du téléphone dans l'espace.

Afin d'éviter que la forme du téléphone soit une variable parasite, nos trois techniques utilisent le même téléphone : un Galaxy Nexus I9250, utilisé pour capturer les contacts des doigts ainsi que pour afficher la liste. Les trois techniques sont implémentées en python et fonctionnent sur un ordinateur Intel Core i7 2,2 Ghz.

5.3.4 Participants

Six hommes et six femmes, âgés de 22 à 33 ans (moyenne = 26.75, écart type = 2.70) ont bénévolement participé à l'expérience. Un des participants était gaucher. Huit participants étaient habitués à un téléphone avec un écran tactile. Tous les participants ont manipulé le téléphone de leur main dominante.

5.3.5 Procédure

Durant l'expérience, les participants doivent sélectionner un objet physique, à savoir une balle de ping-pong intégrant une LED. Ils sont assis à 2.1 m en face des quatre panneaux (correspondant aux 4 densités différentes), exactement au milieu de la rangée de panneaux (Figure 5.9). Bien que les trois techniques de désambigüisation soient mobiles, les participants sont assis pour minimiser la fatigue durant l'expérience relativement longue (environ 90 minutes). Un panneau spécial n'ayant qu'une seule cible non numérotée sert à marquer la *cible de départ* à partir de laquelle les participants commencent un nouvel essai. Ce panneau spécial est situé à droite de la rangée de panneaux comme le montre la Figure 5.9. Ceux-ci sont espacés de 35 cm les uns des autres.

Pour commencer un essai, les participants doivent pointer vers la *cible de départ*, sans

2. http://www.naturalpoint.com/optitrack/ (lien valide en septembre 2015)
3. http://www.naturalpoint.com/optitrack/products/arena/ (lien valide en septembre 2015)

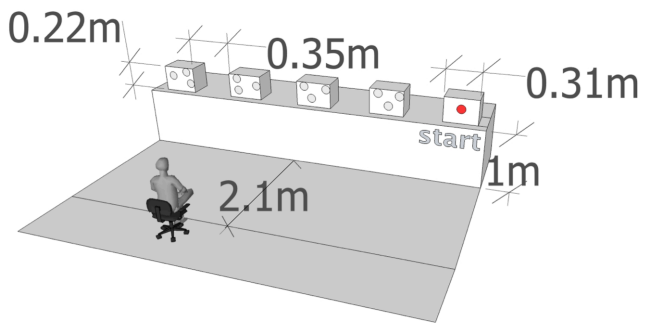


FIGURE 5.9 – Configuration de l'expérience.

contrainte quant à l'angle du poignet ou bien la position du doigt sur l'écran. Une fois pointée, la cible de départ devient verte pour indiquer que la sélection peut être validée. Après cette sélection, une autre cible choisie au hasard parmi les 30 balles de ping-pong devient rouge. Lorsqu'un panneau est visé, toute ses cibles s'allument (blanc) pour indiquer qu'elles se trouvent dans le volume de sélection. Une étude préliminaire nous a montré que lorsque la cible était sur le bord d'un panneau, les utilisateurs avaient tendance à ne pointer qu'un sous-ensemble des cibles du plateau, créant ainsi de nouvelles densités. Afin d'éviter d'avoir un paramètre non contrôlé, nous avons décidé d'allumer toutes les cibles d'un panneau, même celles en-dehors du volume de sélection. Ce choix nous a permis d'ignorer la taille du cône de sélection et de contrôler le nombre de cibles pré-sélectionnées lors du pointage pendant l'expérience.

Au début de l'étape de désambigüisation, un retour d'information est fourni :

— dans le monde physique avec P2Roll et P2Slide. L'élément correspondant à l'angle du poignet avec P2Roll ou la position du doigt sur l'écran avec P2Slide devient bleu, indiquant quelle sera la sélection en cas de validation. Lorsque cet élément courant est sur la cible à sélectionner, celle-ci devient verte.

— sur l'écran du téléphone pour la liste. Un son additionnel signale à l'utilisateur que la liste de noms est apparue sur l'écran, ceci afin d'attirer son attention visuelle sur le téléphone [Talsma *et al.*, 2010]. Aucun retour d'information n'est fourni sur les objets physiques.

Une fois la désambigüisation validée (correcte ou non), un son indique que l'essai est terminé et la *cible de départ* redevient rouge pour relancer un nouvel essai.

5.3.6 Protocole expérimental

Nous avons utilisé un protocole à mesures répétées intra-participant. Les facteurs expérimentaux étaient la technique de désambigüisation T (List, P2Roll, P2Slide) et la densité D (2, 4, 8, 16).

L'expérience a duré environ 90 minutes par participant. Elle était divisé en trois parties correspondant chacune à une technique de désambigüisation. L'ordre de présentation des techniques était contrebalancé entre les participants. La position des panneaux était quant à elle contrebalancée entre les participants en utilisant un carré latin 4×4, de façon à ce que les quatre densités soient à toutes les distances possibles de la *cible de départ*. Entre chaque partie, les participants devaient remplir un questionnaire qualitatif à propos de la technique qu'ils venaient d'utiliser. Chaque partie était composée de 6 blocs : 3 blocs d'entraînement et 3 blocs de mesures de performance. Les 3 blocs d'entraînement visaient à (1) éviter l'effet "wow" se produisant lors de notre étude préliminaire lorsque les participants utilisaient nos techniques pour la première fois, (2) faire en sorte que tous les participants soient au même niveau d'expertise, sachant que quatre d'entre eux n'avaient pas l'habitude des écran tactiles de téléphones et (3), permettre aux utilisateurs d'apprendre les techniques, notamment les relations spatiales gauche/droite entre les cibles pour P2Roll et P2Slide arbitrairement décidées dans certains cas comme les cibles 10 et 13 à la Figure 5.8. Chaque bloc était composé de 40 sélections aléatoires (10 par panneaux / densités). Ce protocole a donc mené à 3 techniques $\times$ 3 blocs $\times$ 40 sélections $=$ 360 sélections par participants, soit 4320 sélections au total.

5.3.7 Résultats

Les principales mesures dépendantes sont le temps de sélection et le taux d'erreurs (pour H1a et H2). Afin d'obtenir une meilleure compréhension des différences entre nos techniques, nous avons aussi considéré leur empreinte spatiale. L'empreinte d'un dispositif correspond à la longueur totale du chemin parcouru pendant une sélection (pointage *et* désambigüisation). En considérant que le dispositif a été localisé à N positions 3D p_k pendant une sélection, son empreinte sera :

$$Empreinte = \sum_{k=1}^{k=N} ||p_k.p_{k-1}||$$

Cette mesure est couramment utilisée afin d'évaluer la fatigue [Grossman et Balakrishnan, 2006].

1,58% des données ont été retirées avant analyse car considérées comme des valeurs aberrantes (données s'écartant de la moyenne d'au moins trois fois l'écart type).

Nous avons analysé nos données avec des tests non-paramétriques puisque (1) un test de normalité de Shapiro-Wilk a montré une violation de la normalité supposée pour Anova [$p < 0.0001$] et (2), un test de Bartlett a montré une violation de l'homogénéité des variances pour tous les facteurs [$p < 0.0001$].

5.3.7.1 H1a : Garder l'attention de l'utilisateur dans le monde physique est plus efficace que de forcer l'utilisateur à porter son attention sur une représentation numérique

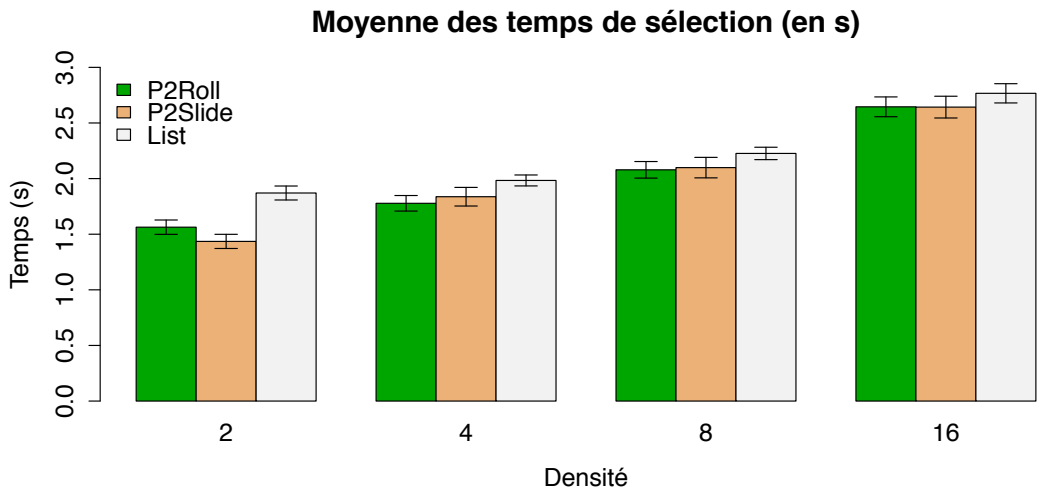


FIGURE 5.10 – Temps de sélection (en secondes). Les barres d'erreur représentent les intervalles de confiance à 95%.

Des tests de Friedman montrent des effets significatifs de T $[\chi^2(2) = 129.9, p < 0.0001]$ et D $[\chi^2(3) = 1166.1, p < 0.0001]$ sur le temps de sélection (Figure 5.10).
Des tests de Mann-Whitney deux-à-deux avec une correction de Bonferroni révèlent comme prévu que la densité D a un impact significatif sur les temps de sélection avec toutes les techniques $[p < 0.001]$. De plus, les techniques maintenant l'attention de l'utilisateur dans le monde physique (P2Roll et P2Slide) ont de meilleurs temps de sélection pour les densités 2, 4 et 8 $[p < 0.0001]$ que la liste. P2Roll et P2Slide offrent respectivement un gain de temps de 16% et 23% dans la densité 2, 10% et 7% dans la densité 4 et 6% et 7% dans la densité 8 par rapport à la liste. Nous n'avons pas trouvé de différence entre les techniques pour la densité 16.

Des tests du khi-deux révèlent que la densité D a un impact sur P2Roll $[\chi^2(3) = 39.1, p < 0.0001]$ et P2Slide $[\chi^2(3) = 55.7, p < 0.0001]$ seulement (Figure 5.11). Nous notons une hausse importante du taux d'erreurs pour P2Roll et P2Slide : d'environ 3% dans la densité 2 à environ 20% dans la densité 16. Au contraire, la densité n'impacte pas le taux d'erreurs obtenu avec la liste (entre 2% et 5%).

Nous affinons notre analyse du taux d'erreurs en prenant en compte le type d'erreurs.
Tout d'abord, les erreurs peuvent survenir pendant l'étape de *pointage*, c'est-à-dire lorsque la cible voulue n'est pas dans le volume de sélection (Figure 5.12 et Figure 5.13, valeurs *Volume de sélection erroné*). 16% des sélections incorrectes avec P2Roll et 22% avec P2Slide étaient de telles erreurs (contre 14% pour la liste). Deux observations au cours

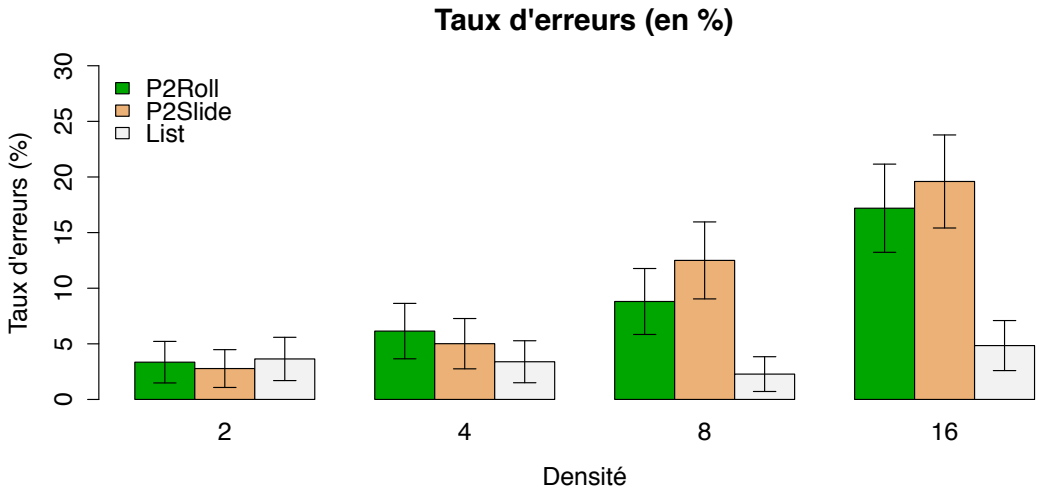


FIGURE 5.11 – Taux d'erreurs (en pourcentage) pour chaque technique en fonction de la densité. Les barres d'erreurs indiquent les intervalles de confiance à 95%.

de l'expérience peuvent expliquer ces erreurs :

— Certains participants avaient tendance à verrouiller le pointage trop tôt en utilisant les techniques laissant l'attention sur les objets, sélectionnant ainsi le mauvais panneau. Cette observation est confirmée par le nombre d'annulation effectuée avec chaque technique : 6,3% avec P2Roll, 2,7% avec P2Slide et 0,5% avec la liste.

— L'action pour annuler le verrouillage du pointage était plus compliquée à effectuer avec P2Roll et P2Slide (glisser le pouce vers le bas de l'écran puis relâcher le contact une fois le pouce dans la bonne zone) qu'avec la liste (simplement appuyer sur le bouton *retour* du téléphone).

Ainsi, la précipitation pour verrouiller le pointage et la difficulté pour annuler ce verrouillage ont conduit à plus d'erreurs avec P2Roll et P2Slide qu'avec la liste.

Les erreurs peuvent aussi survenir pendant l'étape de désambigüisation, c'est-à-dire quand la cible voulue est dans le volume de sélection (Figure 5.12 et Figure 5.13, valeurs *Sélection avant d'atteindre la cible* et *après avoir dépasser la cible*). P2Roll et P2Slide ont 60% de leurs erreurs survenant juste après avoir atteint la cible (*Sélection après avoir dépasser la cible*). La validation se passe en moyenne 0,51s après la perte de la cible pour P2Roll et 0,75s pour P2Slide.

Une explication possible de ces erreurs peut venir du lien étroit entre la validation (relâcher le doigt de l'écran) et la sélection (position angulaire du poignet et position du doigt sur l'écran) : relâcher le doigt peut en effet modifier la dernière position angulaire du poignet ou bien la dernière position du doigt sur l'écran créant menant ainsi à la sélection d'une cible adjacente à celle voulue.

Enfin, une dernière explication pour les taux d'erreurs plus élevés de P2Roll et P2Slide

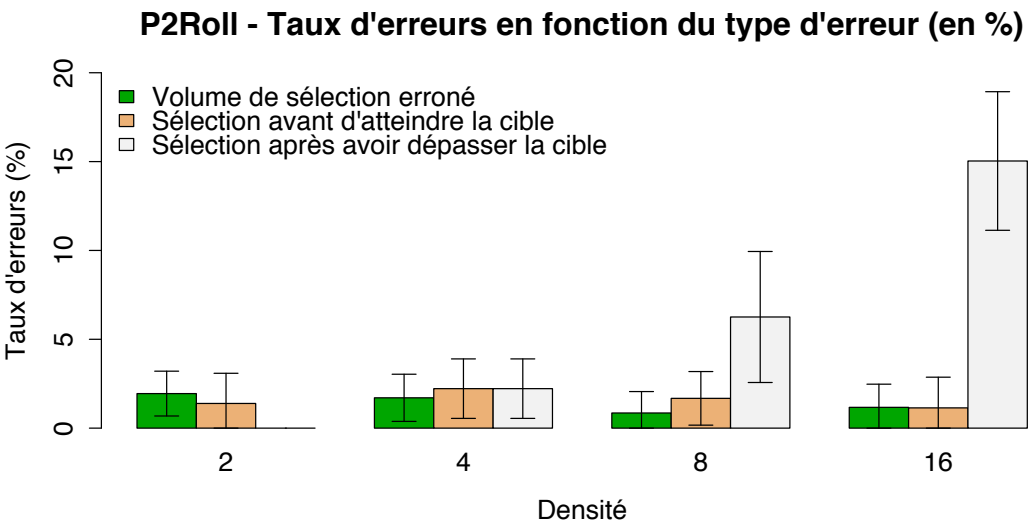


FIGURE 5.12 – Taux d'erreurs (en pourcentage) pour P2Roll en fonction de la densité et du type d'erreur.

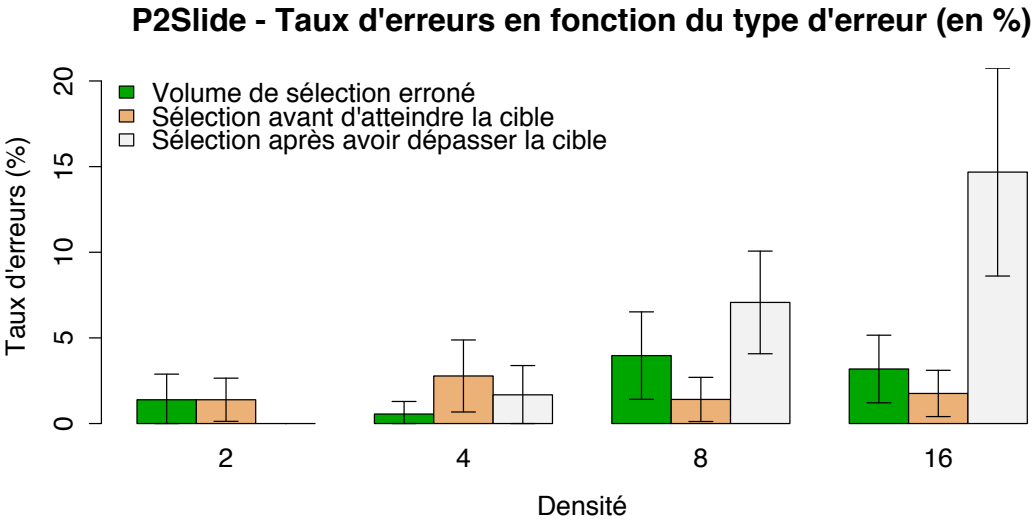


FIGURE 5.13 – Taux d'erreurs (en pourcentage) pour P2Slide en fonction de la densité et du type d'erreur.

peut venir de la fatigue lors d'une utilisation prolongée. Nous considérons l'empreinte comme définie précédemment :

$$Empreinte = \sum_{k=0}^{k=N-1} ||p_k.p_{k-1}||$$

Avec p_k la k-ième position 3D parmi les N positions enregistrées durant la sélection.

La liste a une plus petite empreinte (moyenne = 18.58 cm, écart type = 9.44 cm) que P2Roll (moyenne = 26.73 cm, écart type = 13.24 cm) et P2Slide (moyenne = 24.82 cm, écart type = 13.39 cm). Des tests de Friemdan révèlent un effet significatif de T $[\chi^2(2) = 217.9, p < 0.0001]$: les participants ont effectué de plus petits mouvements avec la liste qu'avec P2Roll et P2Slide. Cela pourrait induire de la fatigue et ainsi expliquer pourquoi les participants ont eu plus d'erreurs par la suite avec P2Roll et P2Slide.

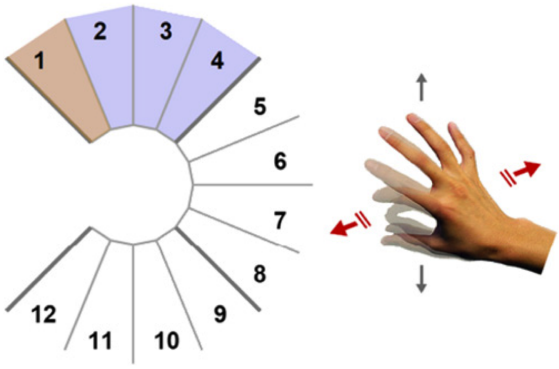


FIGURE 5.14 – Illustration de rapMenu de [Ni *et al.*, 2011]. Les quatre gestes de pincement (de l'index au petit doigt) sélectionnent les éléments 1-4, 5-8 et 9-12 lorsque l'angle du roulis du poignet se trouve dans une de ces trois zones.

Les pistes pour améliorer P2Roll et P2Slide au regard de ces résultats sont de trois ordres. Premièrement, l'action pour annuler le verrouillage du pointage devrait être étudiée plus en détail pour être simplifiée. Notre action était à la fois difficile à effectuer (relâcher le pouce alors que celui-ci est situé en bas de l'écran) et trop proche de la validation normale (relâcher le pouce). Nous pourrions proposer un geste découplé de la validation, comme une secousse du téléphone par exemple.
Deuxièmement, la précision de P2Roll dans les cas de fortes densités pourrait être améliorée en adaptant rapMenu [Ni *et al.*, 2011] : une zone angulaire serait reliée à plusieurs objets (et non un seul), et le pincement d'un doigt donné avec le pouce pourrait être relié à un objet de cette zone angulaire comme illustré par la Figure 5.14. Cette solution présente l'avantage de demander moins de précision quant au roulis du poignet.
Troisièmement, la validation pourrait être découplée de la sélection. Par exemple avec P2Slide, la validation pourrait être une inclinaison en avant comme pour TouchOver [Scoditti *et al.*, 2011]. Cela permettrait de réduire le nombre d'erreurs dues à des mouvements involontaires lors de la sélection.

Laisser l'attention de l'utilisateur dans le monde physique pendant l'étape de désambigüisation est plus performant en terme de temps. Pour diminuer le taux d'erreurs de P2Roll et P2Slide dans les cas de fortes densités d'objets physiques, nous prévoyons d'explorer de nouvelles manières d'annuler l'étape de pointage ainsi que des méthodes de validation découplées de la sélection. De tels mécanismes peuvent empêcher environ 70% des erreurs

obtenues avec P2Roll et P2Slide : plus de 10% d'erreurs dues à l'annulation et plus de 60% d'erreurs dues à la validation couplée à la sélection.

5.3.7.2 H1b : Les utilisateurs préfèrent garder leur attention dans le monde physique plutôt que de porter leur attention sur une représentation numérique

Notre seconde hypothèse était que les utilisateurs préféraient garder leur attention sur les objets physiques eux-mêmes plutôt que de se détourner vers une représentation numérique pendant l'étape de désambigüisation. Cette hypothèse est validée puisque 10 participants sur 12 ont préféré soit P2Roll (5 participants), soit P2Slide (5 participants) et classé la liste en dernière position. Seuls deux participants ont classé la liste en première position. Une analyse plus détaillée montre que même si les utilisateurs ont trouvé de nombreux avantages à la liste (par exemple, ils se sentaient confiants et ils ont trouvé que la technique était facile d'utilisation et d'apprentissage), ils ont également trouvé la technique moins agréable à utiliser.

Il ressort également de l'analyse des entretiens que 7 participants ont mentionné de manière négative le changement d'attention, en disant par exemple *"Ce n'est pas cool de regarder l'écran et de perdre de vue les balles"*. 4 participants ont mentionné de manière négative l'impression d'avoir des actions supplémentaires à effectuer (telles que la lecture ou le défilement de la liste), en disant par exemple *"Il est fastidieux de devoir lire..."*. Pourtant, certains participants étaient conscients des avantages de la liste : *"Je ne suis pas obligé de faire des mouvements précis"* montrant ainsi qu'ils étaient conscients des limites de précision avec P2Roll et P2Slide dans les cas de fortes densités.

Pour conclure sur H1b, les utilisateurs préfèrent garder leur attention sur les objets physiques pendant l'étape de désambigüisation, même s'ils sont conscients des limites de précision potentielles, notamment dans les cas de densités élevées d'objets physiques. Cela confirme qu'il est intéressant d'étudier ce type de techniques.

5.3.7.3 H2 : Les petits mouvements ne sont pas forcément plus efficaces

Notre hypothèse H2 était que les petits mouvements de P2Slide allaient mener à une sélection plus rapide que les mouvements plus amples de P2Roll. Un test de Mann-Whitney deux-à-deux avec une correction de Bonferroni révèle que P2Slide est plus rapide que P2Roll seulement dans le densité 2 [$p < 0.0001, r = .048$]. Les deux techniques sont équivalentes dans les autres densités. Nous expliquons ce résultat par le fait que les petits mouvements sont efficaces quand la tâche est facile. Mais lorsque la difficulté augmente, les petits mouvements doivent êtres exécutés avec précision, réduisant ainsi la différence de temps entre des mouvements petits et des mouvements plus larges.

En ce qui concerne les taux d'erreurs, un test khi-deux ne révèle aucune différence significative entre les deux techniques, quelle que soit la densité $[\chi^2(1) = 0.8491, p > 0.1]$. Pourtant, les mouvements du poignet de P2Roll semblent avoir été plus fatiguants : l'analyse des entretiens montre que la fatigue a principalement été rapportée pour P2Roll (7 participants pour P2Roll, 3 pour P2slide et 4 pour la liste).

Pour conclure sur H2, P2Slide mène à de meilleurs temps de sélection que P2Roll seulement pour la densité 2. En se basant sur une analyse à la fois quantitative et qualitative, nous pouvons considérer que P2Roll et P2Slide sont équivalentes pour de plus fortes densités (4, 8 et 16).

Résumé

- Les techniques laissant l'attention de l'utilisateur dans le monde physique sont plus rapides que la liste pour les densités 2, 4 et 8. En revanche, leur taux d'erreurs est plus important que la liste pour les densités 8 et 16.
- Les utilisateurs préfèrent désambigüiser un pointage en gardant leur attention sur les objets physiques.
- Il n'existe qu'une différence entre les techniques P2Roll et P2Slide utilisant deux référentiels différents et laissant l'attention de l'utilisateur dans le monde physique : P2Slide est plus rapide que P2Roll dans la densité 2.

5.4 Conclusion sur l'évaluation expérimentale

Dans ce chapitre, nous avons décrit notre étude portant sur les options de conception du groupe *Interaction* de notre espace de conception. Nous avons en effet voulu explorer les avantages et inconvénients de techniques de désambigüisation laissant l'attention de l'utilisateur sur les objets physiques précédemment pointés (option *physique* sur l'axe 1, espace de représentation). Pour cela, nous avons considéré deux alternatives pour l'interaction lors de la désambigüisation : un roulis du poignet avec P2Roll (option *relatif au corps* sur l'axe 2, espace de contrôle) et un glissement du pouce sur un écran tactile avec P2Slide (option *relatif au dispositif* sur l'axe 2, espace de contrôle). Nous avons évalué ces deux techniques en les comparant à une technique de référence présentant le nom des objets sur une liste affichée sur l'écran d'un téléphone.

Nous avons montré via une expérience contrôlée en laboratoire qu'une désambigüisation laissant l'attention de l'utilisateur dans le monde physique est performante et préférée par les utilisateurs comparée à une technique les obligeant à se concentrer sur une représentation numérique des objets. Ce gain de performance (temps de sélection et taux d'erreurs) est valide jusqu'à une désambigüisation de 8 objets. Pour de plus fortes densités (16 objets dans le volume de sélection), toutes les techniques mènent à des temps de sélection équivalents. Nous avons pu identifier cette convergence dans la densité 16 en

considérant une densité plus élevée que les études précédentes (4 objets pour [Rukzio *et al.*, 2007] [Välkkynen *et al.*, 2006] et 5 objets pour [Swindells *et al.*, 2002]). Il faut toutefois noter que les erreurs de sélection sont importantes pour les techniques de désambigüisation dans le monde physique dans les cas de densités élevées (8 et 16). Partant de ce constat, de nouvelles études sont envisagées.

5.5 Travaux futurs à partir de l'évaluation expérimentale

Trois pistes de recherche sont identifiées faisant suite à cette étude expérimentale.

Premièrement, nous envisageons l'amélioration de P2Roll et de P2Slide au regard de la précision dans les cas de densités élevées. Nous avons déjà identifié deux améliorations possibles : une annulation du pointage simplifiée et un découplage entre la validation (lever le doigt de l'écran) et la sélection (roulis du poignet ou position du pouce sur l'écran).
D'autres pistes restent à explorer, telles que l'utilisation de gestes discrets plutôt que continus, c'est-à-dire faire glisser le doigt plusieurs fois sur l'écran au lieu de faire un glissement continu pour P2Slide par exemple. Une autre solution serait l'ajout de degré de liberté en entrée. En effet, puisque la technique de référence, à savoir la liste, était en 1D, nous avons évalué des technique 1D. Nous envisageons l'extension de nos techniques en 2D. Par exemple, lors d'une étude pilote, les participants ont fourni des retours positifs quant à une version de P2Slide en 2D utilisant des glissements horizontaux et verticaux sur l'écran.

Deuxièmement, nous avons l'intention de fournir une meilleure compréhension du temps de sélection en étendant la loi de Fitts à nos techniques de désambigüisation. Puisque l'étape de pointage dans le monde physique peut être modélisé par la loi de Fitts ainsi que le roulis du poignet par exemple, nous aimerions analyser leur association, en particulier dans le cas d'actions séquentielles pouvant être effectuées en parallèle à mesure que l'utilisateur devient expert, comme dans le cas de RakeCursor [Blanch et Ortega, 2009]. RakeCursor est une technique de pointage pour des interfaces graphiques permettant à l'utilisateur de sélectionner le curseur à manipuler grâce au regard, puis de manipuler ce curseur grâce à la souris. Ces deux modalités peuvent être effectuées en parallèle. Les auteurs proposent une interprétation possible du gain de performance en se basant sur la théorie de l'information de la loi de Fitts : les bandes passantes de chaque modalité (pointage et regard) s'additionnent grâce au parallélisme des deux modalités. Une modélisation de nos techniques en deux temps permettrait une meilleure compréhension de ce phénomène, à savoir le temps qu'il faut pour que le parallélisme soit optimal, les indices de difficultés (du pointage et de la désambigüisation) permettant ce parallélisme, etc.
Cette étude nécessite une expérimentation autre que celle effectuée dans nos travaux. En effet, l'étude présentée ici visait à savoir si les utilisateurs sont performants avec nos techniques, sans contrainte sur leur manière de procéder. Au contraire, dans une

expérience pour l'étude de la modélisation de nos techniques, il conviendra de contrôler certains paramètres (comme forcer les utilisateurs à être à une certaine position angulaire en sélectionnant la *cible de départ*) et ainsi contraindre les utilisateurs à utiliser des mouvements qu'ils n'auraient peut-être pas effectués en conditions réelles.

Enfin, nous envisageons d'étendre nos travaux à des scénarios se rapprochant du monde réel :

— L'étude de l'adaptation des utilisateurs aux fortes densités dans la vie réelle. Par exemple, à partir de quelle densité préfèrent-t-ils se déplacer pour réduire le nombre d'objets dans le volume de sélection ?

— L'étude de la désambigüisation d'objets ne pouvant fournir de retour d'information visuel et ainsi l'exploration d'alternatives de conception telles qu'un retour d'information audio.

— L'étude de la partie **contrôle** qui suit la partie **sélection**. Après avoir sélectionné un objet physique via un geste de pointage (*geste déictique*, paragraphe 2.2), avoir éventuellement effectué une désambigüisation, nous aimerions explorer le déclenchement d'une commande (comme allumer / éteindre) grâce à l'interaction gestuelle, et donc des *gestes symboliques dynamiques* (paragraphe 2.2). Ceci fait l'objet de la partie II de ce manuscrit.

6 Conclusion de la partie sélection

Sommaire

6.1 Conclusion sur la désambigüisation **71**

6.2 Travaux futurs sur la désambigüisation **72**

6.1 Conclusion sur la désambigüisation

Nous avons présenté des travaux sur la sélection d'objets physiques utilisant un pointage volumique (paragraphe 3.2). Plus précisément, nous avons étudié l'étape de désambigüisation interactive (paragraphe 3.3) nécessaire lorsque plusieurs objets sont pré-sélectionnés par le volume.

Nous avons décrit un espace de conception (chapitre 4) pour caractériser les techniques de désambigüisation en considérant les points de vue de l'*Interaction* (paragraphe 4.2.1) et du *Système de désambigüisation* (paragraphe 4.2.2). Ce dernier permet de définir les connaissances dont a besoin le système sur l'identité des objets physiques et/ou leur emplacement dans le monde physique.
Nous avons illustré le pouvoir descriptif de notre espace de conception en montrant que toutes les techniques de désambigüisation existantes peuvent être décrites par les 11 axes de notre espace de manière distincte (paragraphe 4.3).

Nous avons montré le pouvoir génératif de notre espace de conception avec la création et l'évaluation de deux nouvelles techniques de désambigüisation : P2Roll (Physical Pointing Roll) et P2Slide (Physical Pointing Slide) (chapitre 5). Nous avons comparé ces techniques pour connaître les avantages et inconvénients des techniques laissant l'attention de l'utilisateur dans le monde physique, sur les objets eux-mêmes (axe *Espace de représentation* de notre espace de conception, paragraphe 4.2.1.1). Nos deux techniques permettent également d'explorer deux options de conception pour l'entrée de la désambigüisation

(axe *Espace de contrôle* de notre espace de conception, paragraphe 4.2.1.2).
L'évaluation expérimentale montre que (paragraphe 5.3) :

— P2Roll et P2Slide permettent une désambigüisation plus rapide qu'une liste affichant le nom des objets physiques, forçant ainsi l'utilisateur à changer d'attention visuelle en passant des objets physiques à une représentation numérique des objets sur un écran,

— P2Roll et P2Slide demandent plus de précision que la liste dans les fortes densités, et

— les utilisateurs préfèrent utiliser P2Roll et P2Slide plutôt que la liste.
⇒ Laisser l'attention de l'utilisateur dans le monde *physique* est donc plus efficace que de proposer une représentation *numérique* pour désambigüiser jusqu'à 8 cibles. C'est aussi l'option préférée des utilisateurs.

— P2Slide est plus rapide que P2Roll seulement dans la densité la plus faible (c'est-à-dire deux objets dans le volume de sélection).
⇒ Les options *relatif au corps* et *relatif à un dispositif* de l'axe *espace de contrôle* peuvent donc être considérées comme équivalentes lorsque l'attention de l'utilisateur est dans le monde physique.

Cette évaluation expérimentale a aussi permis d'identifier des pistes de travaux futurs.

6.2 Travaux futurs sur la désambigüisation

Nous envisageons d'effectuer d'autres itérations entre l'étude de l'espace de conception et l'étude de techniques de désambigüisation. En effet, la création de nouvelles techniques améliorant les performances de l'utilisateur pendant la désambigüisation peut amener à l'ajout d'options de conception dans l'espace pouvant elles-mêmes mener à la création de nouvelles techniques.

L'évaluation expérimentale a révélé plusieurs solutions à explorer pour améliorer P2Roll et P2Slide, notamment au regard de la précision nécessaire dans les cas de densités d'objets élevées. Ces solutions portent sur l'annulation du pointage, l'ajout de degré de liberté ainsi que l'étude de l'action de validation afin d'éviter des mouvements involontaires lors de la sélection.

Enfin, les techniques de désambigüisation peuvent bénéficier d'analyses détaillées pour fournir une compréhension approfondie sur :

— la relation entre les deux modalités (pour le pointage et la désambigüisation) grâce à une modélisation via la loi de Fitts.

— les limites entre un pointage basé sur un laser et un pointage basé sur un volume suivi d'une désambigüisation. Il serait intéressant de trouver les limites de chaque technique en termes de taille des objets, de densité, etc. Cette analyse pourrait se faire en condition réelle afin d'aussi identifier le comportement de l'utilisateur en contexte (choix d'une technique plutôt qu'une autre, choix de se déplacer pour

surmonter des difficultés de précision, etc.)

Une autre piste est, après avoir considéré la **sélection** d'objets physiques augmentés (geste *déictique*), d'étudier le **contrôle** de ces objets. Le déclenchement de commandes telles que *allumer* ou *éteindre* peut se faire via des *gestes dynamiques symboliques* (*cf.* paragraphe 2.2). Ceci constitue le sujet de la partie II du manuscrit.

Deuxième partie

Contrôle d'un objet physique

Spécification de commandes par des gestes 3D : Nécessité de guider

Table des matières

7 Introduction **79**

8 Espace de Conception **83**

8.1 Espace de Conception 85

8.1.1 Structure générale 85

8.1.2 Axes 87

8.1.3 Positionnement : espaces de conception 103

8.1.4 Positionnement : techniques de guidage 106

8.2 Evaluation 108

8.2.1 Outil interactif pour faciliter l'utilisation des pouvoirs de notre espace de conception 109

8.2.2 Scénarios d'utilisation 115

8.3 Conclusion 118

9 Système de Guidage : OctoPocus3D **121**

9.1 Introduction 122

9.2 Choix de conception 124

9.2.1 Recherche d'une solution grâce à notre espace de conception 124

9.2.2 Description d'OctoPocus 124

9.2.3 Alternatives de conception : exemples 129

9.3 Conception d'OctoPocus3D 135

9.3.1 Affichage des gestes 3D 135

9.3.2 Retour d'information rétroactif : terminal et/ou continu 144

9.3.3 Retour d'information prédictif : origine du guide et portions des gestes affichées 145

9.4 Évaluations expérimentales 148

9.4.1 Ressources 149

9.4.2 Expérience 1 : retour d'information rétroactif 151

9.4.3 Expérience 2 : retour d'information prédictif 156
9.5 Déploiement *in situ* 160
9.6 Conclusion 161
9.7 Travaux futurs 162

7 Introduction

Dans cette partie, nous présentons nos travaux sur le **contrôle** des objets physiques grâce à des gestes *symboliques dynamiques* (paragraphe 2.2). Plus particulièrement, nous nous intéressons au guidage de ces gestes. Comme pour notre étude des gestes *déictiques* pour la **sélection** des objets physiques via un pointage (partie I), nos résultats sont de deux ordres :

— conceptuel avec la création d'un espace de conception (chapitre 8) et
— pratique avec la conception et l'évaluation d'une technique de guidage notée OctoPocus3D (chapitre 9).

Nous motivons tout d'abord notre choix de nous intéresser au guidage de gestes symboliques dynamiques.

Récemment, le domaine de l'interaction gestuelle 3D a motivé de nombreuses études que nous classons à différents niveaux d'abstraction (Figure 7.1), que ce soit pour des mécanismes de capture et de suivi de gestes [Colaço *et al.*, 2013] (Figure 7.1, niveau 1), des algorithmes de reconnaissance de gestes [Kratz et Rohs, 2010] (Figure 7.1, niveau 2), des taxonomies de gestes [Karam et Schraefel, 2005] ou encore des applications utilisant des gestes 3D comme modalité d'interaction [Kela *et al.*, 2006] (Figure 7.1, niveau 3). Il est à noter que de plus en plus de systèmes commerciaux s'intéressent à l'interaction gestuelle 3D. On peut citer en exemples la Kinect de Microsoft [1], la télécommande Wiimote [2], ou bien plus récemment Leap Motion [3] ou encore la plateforme Atheer Air [4] présentée lors du Consumer Electronic Show (CES) 2015.

Malgré ces différentes avancées, l'adoption de l'interaction gestuelle 3D comme modalité en entrée par un large public est encore modérée. Une exception est l'industrie du jeu vidéo, où l'interaction gestuelle a été adoptée. Nous pensons que l'un des freins principaux à cette adoption dans les autres domaines d'application vient du fait que les commandes

1. http://kinectforwindows.org (lien valide en septembre 2015)
2. http://www.wii.com (lien valide en septembre 2015)
3. https://www.leapmotion.com (lien valide en septembre 2015)
4. https://www.atheerlabs.com (lien valide en septembre 2015)

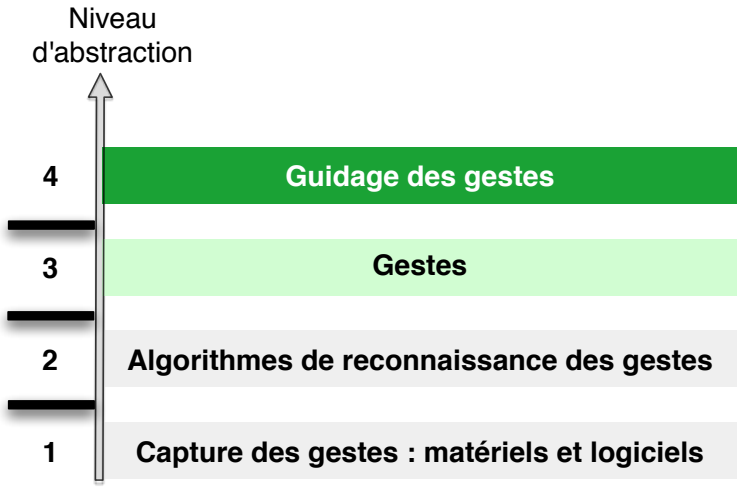


FIGURE 7.1 – Différents niveaux d'abstraction pour étudier l'interaction gestuelle.

gestuelles ne se révèlent pas d'elles-mêmes : les utilisateurs ne savent pas (1) quelles commandes sont disponibles et (2), quels sont les gestes associés à ces commandes.
Pour répondre à ce problème de révélation, deux approchent complémentaires existent (Figure 7.1, niveaux 3 et 4) :

1. La première solution consiste à s'intéresser à la conception des jeux de gestes (Figure 7.1, niveau 3) utilisés dans un contexte bien précis [Kühnel *et al.*, 2011, Vatavu, 2013](e.g., contrôle d'objets du quotidien tels que les rideaux ou les lumières). Ces travaux étudient la correspondance entre une commande (e.g. "allumer la lumière") et un geste (e.g., lever la main). Ces études peuvent inclure des gestes définis par des experts [Nacenta *et al.*, 2013] ou par les utilisateurs eux-mêmes [Kühnel *et al.*, 2011]. Dans les deux cas, il s'avère qu'il est difficile d'obtenir un consensus entre les utilisateurs quant aux choix des gestes ainsi qu'à leurs associations avec des commandes [Vatavu, 2013]. Une solution serait alors de proposer des systèmes interactifs qui ré-emploient les gestes préalablement définis par l'utilisateur [Vatavu, 2012]. Avec cette solution, l'utilisateur enregistre des associations commandes/gestes par exemple sur son téléphone. Par la suite, lorsqu'il s'approche d'un système utilisant l'interaction gestuelle, le système récupère automatiquement les associations définies sur le téléphone de l'utilisateur. Bien que le concept soit intéressant, cette solution nécessiterait une architecture commune entre toutes les applications utilisant l'interaction gestuelle - solution peu réalisable aujourd'hui.
2. La seconde solution consiste à fournir une technique de guidage à l'utilisateur [Anderson *et al.*, 2013], [Bau et Mackay, 2008], [Bragdon *et al.*, 2009], [Freeman *et al.*, 2009], [Kurtenbach *et al.*, 1994] (Figure 7.1, niveau 4). Ces techniques de guidage permettent de *révéler* les commandes disponibles ainsi que de *guider* l'utilisateur pendant qu'il exécute le geste désiré.

Vu les limitations de la première démarche - s'intéresser aux jeux de gestes - , les travaux présentés dans la suite du manuscrit se concentrent sur la seconde approche : le guidage de gestes 3D. En effet, nous pensons que cette approche peut aider l'interaction gestuelle à être adoptée dans des domaines autres que celui du jeu vidéo.

De nombreux guides existent afin d'aider les utilisateurs lors de la phase de continuation, c'est-à-dire pendant l'exécution du geste (cf. paragraphe 2.2.1.2). Malheureusement, ces guides sont souvent conçus de manière *ad hoc*. Nos travaux s'articulent donc autour de la création et l'utilisation d'un espace de conception dédié aux techniques de guidage de gestes.

8 Espace de Conception

Sommaire

8.1 Espace de Conception 85
8.1.1 Structure générale 85
8.1.1.1 Catégories d'axes : retours d'information rétroactif et prédictif 85
8.1.1.2 Groupes d'axes *Quand*, *Quoi*, *Comment* et *Où* 86
8.1.2 Axes 87
8.1.2.1 Groupe d'axes : *Quand* 87
8.1.2.1.1 Déclenchement du guidage 88
8.1.2.1.2 Exécution du geste 92
8.1.2.1.3 Mode de guidage 93
8.1.2.2 Groupe d'axes : *Quoi* 94
8.1.2.2.1 Geste(s) affiché(s) 94
8.1.2.2.2 Valeur de performance pour le retour d'information rétroactif 98
8.1.2.2.3 Impact du retour d'information rétroactif sur le retour d'information prédictif 98
8.1.2.3 Groupe d'axes : *Comment* 99
8.1.2.4 Groupe d'axes : *Où* 102
8.1.3 Positionnement : espaces de conception 103
8.1.3.1 Retour d'information rétroactif 104
8.1.3.2 Retour d'information prédictif 105
8.1.4 Positionnement : techniques de guidage 106
8.2 Evaluation 108
8.2.1 Outil interactif pour faciliter l'utilisation des pouvoirs de notre espace de conception 109
8.2.1.1 Fonction : description détaillée d'une technique de guidage 109
8.2.1.2 Fonction : comparaison entre les techniques de guidage . 111
8.2.1.3 Fonction : rechercher des zones inexplorées dans l'espace de conception 113

8.2.1.4 Fonction : concevoir une nouvelle technique de guidage . 114
8.2.1.5 Fonctions secondaires liées à la base de connaissance de l'outil 115
8.2.2 Scénarios d'utilisation 115
8.2.2.1 Trouver une solution d'interaction 115
8.2.2.2 Trouver un problème de recherche 117
8.3 Conclusion 118

Notre espace de conception s'appuie sur des concepts bien connus dans le domaine de l'IHM comme la notion de *mode* [Hinckley *et al.*, 2006] ou encore les *relations spatiales* entre différentes modalités [Vernier et Nigay, 2001]. Ces notions n'ayant pas vraiment de relations les unes avec les autres dans la littérature, notre espace de conception les unifie afin de spécifier et d'organiser des options de conception pour la création de technique de guidage. La structure de notre espace s'organise autour de différents axes, chacun relié à un problème de conception et possédant diverses options pour répondre à ce problème.
Le champ d'application de cet espace de conception concerne le comportement des techniques de guidage plutôt que leur forme concrète (i.e. la représentation qu'elles utilisent afin de montrer les gestes à l'utilisateur). Nous cherchons à définir le comportement attendu du guide avant de nous intéresser à sa manière de rendre perceptible les informations de guidage.
Nous considérons les travaux effectués au regard du guidage de n'importe quel type de gestes (paragraphe 2.2) afin de prendre en compte un large ensemble d'options de conception. En effet, puisque nous nous concentrons sur le comportement des guides et non sur ce qu'ils présentent concrètement à l'utilisateur, les nombreux travaux effectués pour guider des gestes sont pertinents et indispensables à notre étude concernant le comportement d'une technique de guidage de gestes.

Publications liées à ce chapitre

1. William Delamare, Céline Coutrix et Laurence Nigay. "Designing Guiding Systems for Gesture-Based Interaction", in proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2015). Duisburg, Germany, 2015, p. 44-53.
2. William Delamare, Céline Coutrix et Laurence Nigay. "A Tool for Optimizing the Use of a Large Design Space for Gesture Guiding Systems", in proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2015), Duisburg, Germany, 2015, p. 238-241.

Ce chapitre est consacré à un espace de conception. Ce dernier organise les choix de conception pour une technique de guidage de gestes (paragraphe 8.1). Après cette description, nous présentons l'évaluation de notre espace de conception via l'étude d'un outil interactif (paragraphe 8.2).

8.1 Espace de Conception

Nous commençons par présenter la structure de notre espace (paragraphe 8.1.1) - deux catégories (paragraphe 8.1.1.1) et quatres groupes orthogonaux (paragraphe 8.1.1.2). Puis nous décrivons chaque axe de notre espace (paragraphe 8.1.2). Nous continuons la description de notre espace de conception en présentant les espaces de conception (paragraphe 8.1.3) et techniques de guidage (paragraphe 8.1.4) existants. Présenter l'existant après la présentation de notre espace permet de positionner nos travaux quant aux axes et options de conception que nous avons choisis d'inclure ou non dans notre espace.

8.1.1 Structure générale

Nous présentons les catégories et groupes d'axes structurant notre espace de conception. Cette structure est composée de deux catégories d'information, chacune composée de quatre groupes d'axes.

8.1.1.1 Catégories d'axes : retours d'information rétroactif et prédictif

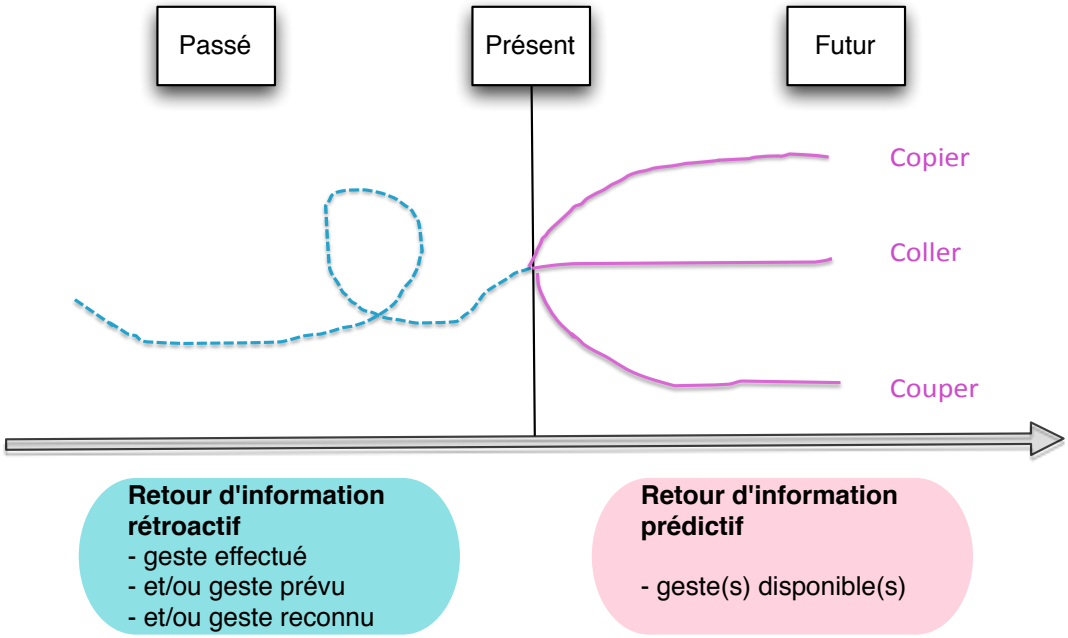


FIGURE 8.1 – Distinction entre retour d'information rétroactif et retour d'information prédictif pour une technique de guidage de gestes : affichage d'un retour d'information rétroactif (i.e. gestes effectué/prévu/reconnu dans le passé) et d'un retour d'information prédictif (i.e. gestes disponibles dans le futur) pendant que l'utilisateur exécute un geste.

Notre espace de conception est composé de deux grandes catégories d'axes, chaque catégorie étant reliée à un aspect fondamental du guidage : le mécanisme en charge du retour d'information rétroactif ("feedback") et le mécanisme en charge du retour d'information prédictif ("feedforward") (Figure 8.1).
Le retour d'information rétroactif fournit l'information concernant les actions déjà effectuées par l'utilisateur (i.e. dans le passé). Pour une technique de guidage, ce mécanisme peut donner des informations à propos du geste effectué, telles que sa qualité d'exécution par rapport au geste initialement prévu et/ou comment le système interprète ou reconnait ce geste. Sur l'exemple de la Figure 8.1, le guide 2D dessine en pointillés la portion de geste effectuée depuis le début, c'est-à-dire le geste effectué jusqu'ici.
Le retour d'information prédictif fournit l'information avant toute action (i.e. pour de futures actions). Pour une technique de guidage, le mécanisme de retour d'information prédictif devrait présenter les commandes disponibles ainsi que les gestes qui permettent de les déclencher. Sur l'exemple de la Figure 8.1, le retour d'information prédictif montre trois portions restantes de gestes que l'utilisateur peut compléter afin de déclencher une commande parmi 'Copier', 'Coller' et 'Couper'.
La distinction entre retour d'information rétroactif et retour d'information prédictif a été faite dans plusieurs travaux [Vermeulen *et al.*, 2013, Bau et Mackay, 2008]. Toutefois, la définition du retour d'information prédictif est loin d'être consensuelle. Les travaux de [Vermeulen *et al.*, 2013] visent à unifier les différentes définitions du retour d'information prédictif qui existent dans la littérature. Pour cela, il faut distinguer l'action et son résultat. Selon leur définition, le retour d'information prédictif ne révèle que le résultat d'une action, le but auquel l'utilisateur souhaite arriver. Les actions nécessaires pour arriver à ce but sont quant à elles révélées par les *affordances perçues*. Ces affordances perçues permettent à l'utilisateur de comprendre les actions possibles grâce à des affordances sensorielles couplées à des affordances physiques [Vermeulen *et al.*, 2013]. Par exemple, lorsqu'un utilisateur souhaite débloquer l'écran de son téléphone (Figure 8.2), le bouton à glisser avec une icône de flèche ainsi que l'animation lumineuse sur le texte font partie intégrante des affordances perçues et indiquent à l'utilisateur l'action qu'il doit faire afin d'atteindre son but. Le retour d'information prédictif relayé par l'étiquette "slide to unlock" indique la fonctionnalité de cette action. Par soucis de simplification, nous faisons l'amalgame entre ces deux types d'information. En effet, bien que l'on puisse ultérieurement les distinguer, nous considérons qu'une technique de guidage présente les deux informations (i.e. les gestes et les commandes associées). De ce fait, nous appelons dans la suite de ce document le mécanisme en charge de représenter ces deux informations le mécanisme de retour d'information prédictif, afin de souligner son opposition avec le mécanisme de retour d'information rétroactif.

8.1.1.2 Groupes d'axes *Quand*, *Quoi*, *Comment* et *Où*

Les deux catégories composant notre espace de conception sont elles-mêmes décomposées en quatre groupes d'axes. Ces groupes mettent en évidence différents aspects du

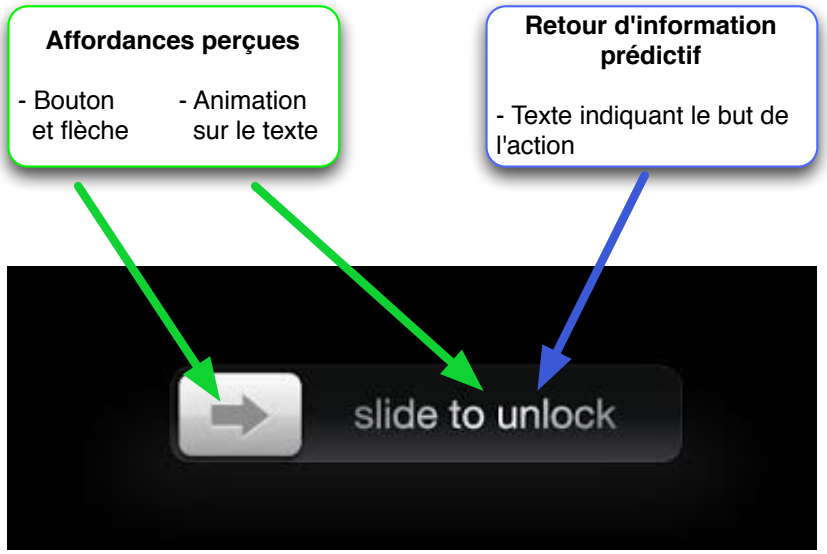


FIGURE 8.2 – Exemple illustrant les concepts d'affordances perçues et de retour d'information prédictif selon [Vermeulen *et al.*, 2013].

comportement de la technique de guidage suivant quatre questions : *Quand* pour les caractéristiques temporelles, *Quoi* pour les caractéristiques du contenu, *Comment* pour les caractéristiques de la modalité, et *Où* pour les caractéristiques spatiales.

L'espace de conception possède un total de 35 axes, proposant au total 115 options ainsi que leurs différentes combinaisons. La Figure 8.3 montre la structure de notre espace de conception. Pour visualiser le détail des axes et des options de conception, le lecteur peut se référer à la Figure 8.4 pour les axes de la catégorie du retour d'information rétroactif et la Figure 8.5 pour ceux de la catégorie du retour d'information prédictif.

8.1.2 Axes

Nous présentons les axes qui composent notre espace de conception. Cette présentation suit la structure générale de notre espace de conception et les quatre groupes d'axes *Quand*, *Quoi*, *Comment* et *Où*.

8.1.2.1 Groupe d'axes : *Quand*

Ce groupe d'axes décrit les caractéristiques temporelles du retour d'information rétroactif et du retour d'information prédictif (Figure 8.6). Nous considérons l'exécution d'un geste comme unité de temps. Avec cette référence temporelle, nous distinguons le début (i.e. le *déclenchement*), le milieu (i.e. l'*exécution* du geste) et la fin (i.e. la sortie du *mode* de guidage). Ce schéma est similaire à celui des étapes composant l'interaction gestuelle (i.e.

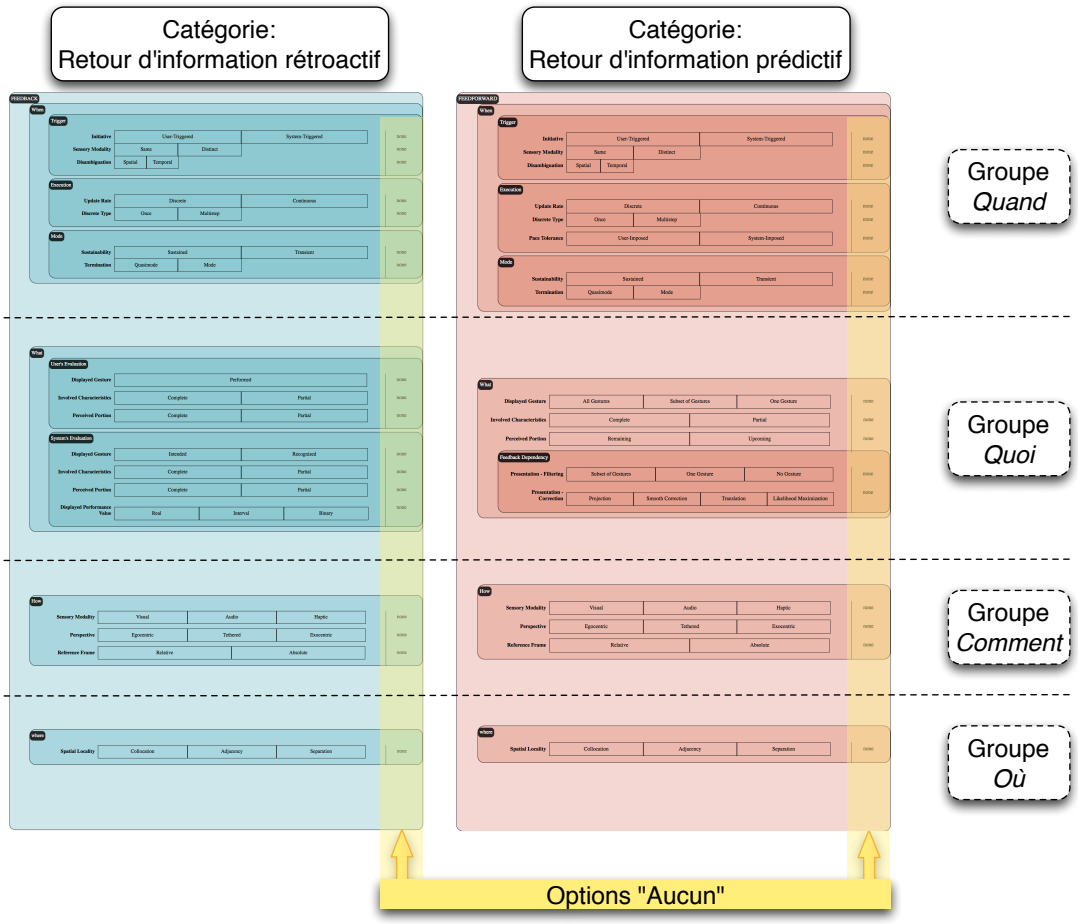


FIGURE 8.3 – Aperçu de l'espace de conception avec 35 axes organisés selon deux catégories - retour d'information rétroactif et retour d'information prédictif. Chaque catégorie est composée de quatre groupes - Quand (caractéristiques temporelles), Quoi (caractéristiques du contenu), Comment (caractéristiques de la modalité) et Où (caractéristiques spatiales). Chaque axe possède des options de conception ainsi qu'une option "aucun".

initialisation, continuation et terminaison du geste). Nous faisons néanmoins la distinction entre les deux schémas temporels, c'est-à-dire les étapes de l'interaction gestuelle et les étapes du guidage (Figure 8.7) puisque l'initialisation du geste peut correspondre à un instant t différent de l'enclenchement du guide par exemple. Ici, nous décrivons les étapes correspondant au guidage d'un geste.

8.1.2.1.1 Déclenchement du guidage Nous définissons trois axes pour décrire l'évènement déclencheur :

1) *Initiative* : le mécanisme peut être déclenché par le *système* ou par l'*utilisateur*. Par exemple, un mécanisme de retour d'information prédictif déclenché par l'utilisateur peut afficher les gestes après un temps de pause sur une surface tactile

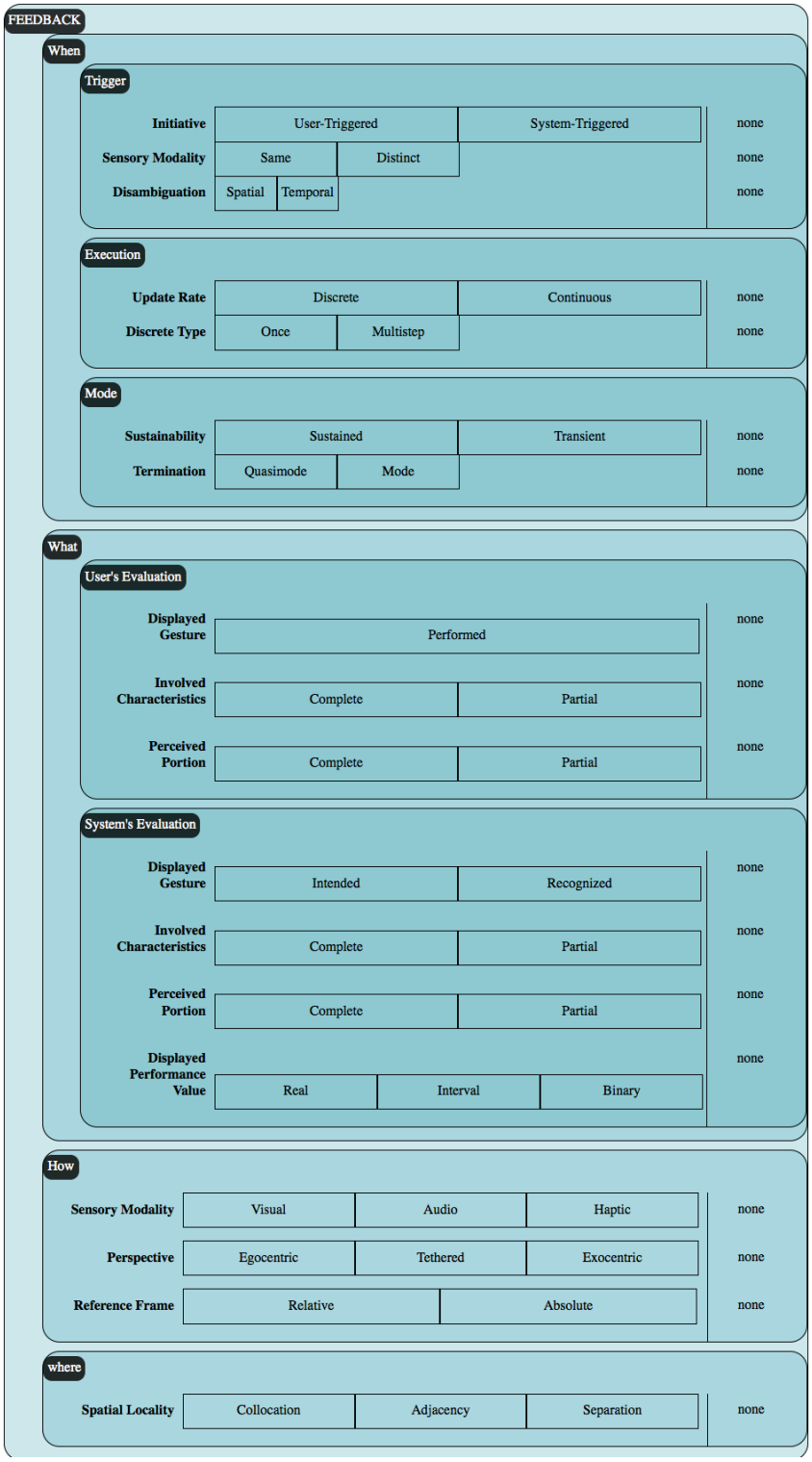


FIGURE 8.4 – Catégorie : retour d'information rétroactif.

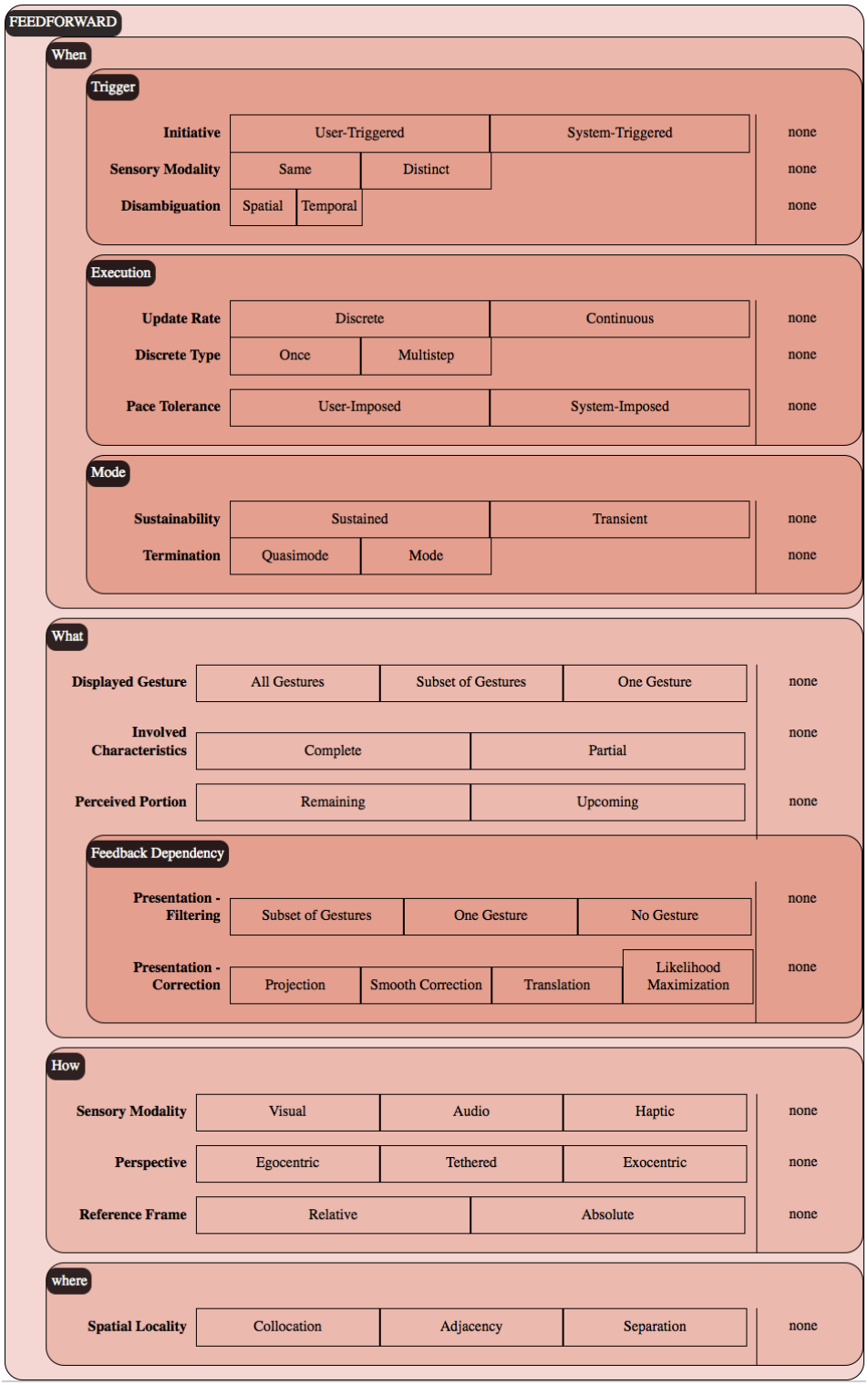


FIGURE 8.5 – Catégorie : retour d'information prédictif.

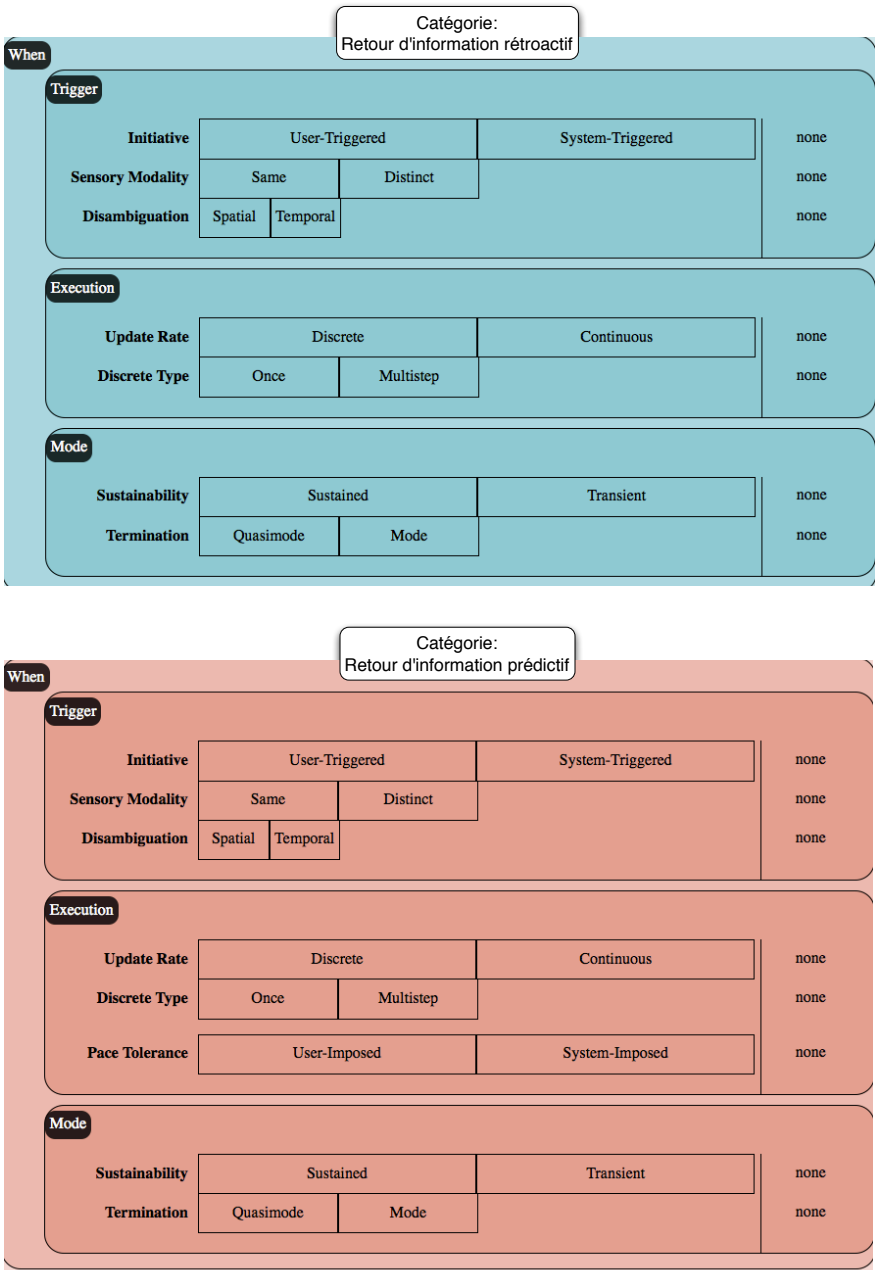


FIGURE 8.6 – Groupes *Quand* du retour d'information rétroactif (haut) et du retour d'information prédictif (bas).

[Kurtenbach *et al.*, 1994, Freeman *et al.*, 2009]. Un mécanisme de retour d'information prédictif peut aussi être déclenché par le système et afficher automatiquement plus d'information si celui-ci juge que l'utilisateur est en difficulté [Anderson *et al.*, 2013].

Un guide permet aux utilisateurs d'exécuter un geste en utilisant une modalité d'interaction donnée. Cela peut être des gestes 2D effectués avec le doigt sur une surface tactile ou des

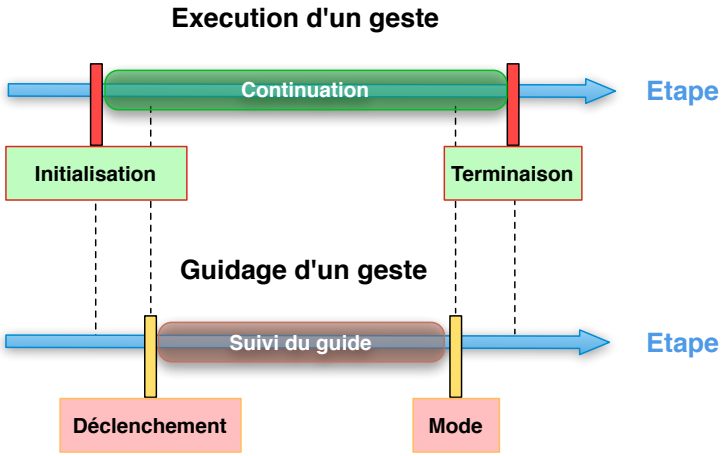


FIGURE 8.7 – Distinction entre les étapes de l'interaction gestuelle (initialisation, continuation et terminaison d'un geste) et les étapes du guidage (enclenchement du guide, comportement pendant l'éxecution et sortie du mode de guidage).

gestes 3D effectués dans l'air avec la main par exemple. Nous nommons cette modalité guidée "modalité principale". Avec cette définition, nous affinons le cas d'un déclenchement effectué par l'utilisateur avec deux axes :

2) *Modalité* : l'utilisateur peut déclencher le mécanisme voulu en utilisant la modalité principale (i.e. la *même*) ou une modalité *différente*.
Avec une application visant à guider des gestes sur un écran tactile par exemple, toucher l'écran [Bragdon *et al.*, 2009] correspond au cas où le déclenchement du guide est effectué par une action selon la même modalité que celle attitrée pour effectuer les gestes. Lorsque l'utilisateur appuie sur un bouton d'une télécommande wiimote afin d'obtenir un retour d'information rétroactif, il utilise une modalité différente [Schwaller *et al.*, 2014].

Dans le cas d'un guide activé par l'utilisateur grâce à la modalité principale, un mécanisme de désambigüisation est nécessaire. Il faut en effet distinguer une action réalisée dans le contexte de l'application et l'action réalisée pour déclencher le guide.

3) *Désambigüisation* : le système peut détecter l'évènement déclencheur grâce à un critère *temporel* (e.g, un temps de pause [Kurtenbach *et al.*, 1993]) ou *spatial* (e.g., faire des gestes dans un espace réservé [Bragdon *et al.*, 2009] ou bien appuyer sur un bouton [Ghomi *et al.*, 2013]).

8.1.2.1.2 Exécution du geste Une fois que l'utilisateur ou le système a déclenché le mécanisme de retour d'information rétroactif ou le mécanisme de retour d'information prédictif, nous décrivons les caractéristiques temporelles de la technique de guidage pendant l'exécution d'un geste. Deux axes sont communs au retour d'information rétroactif et au retour d'information prédictif. Un troisième axe est spécifique au retour d'information

prédictif.

1) *Mise à jour* : l'information présentée par le système peut être mis à jour de manière *continue* ou *discrète*.
   Représenter en temps réel le chemin 2D effectué par l'utilisateur est un exemple de mécanisme de retour d'information rétroactif mise à jour de façon continue [Anderson et Bischof, 2013]. Cela permet à l'utilisateur de juger son geste 2D de manière continue tout au long de son exécution. Un autre exemple de mécanisme de retour d'information prédictif avec une mise à jour continue est celui utilisé par OctoPocus [Bau et Mackay, 2008] ou ShadowGuides [Freeman *et al.*, 2009]. Ces techniques ne représentent que les gestes les plus susceptibles d'être reconnus en prenant en compte le début du geste réalisé par l'utilisateur. Le jeu de gestes affiché est donc continuellement mis à jour.
   Une mise à jour discrète peut être détaillée selon le nombre d'occurrences :

2) *Nombre d'occurrences* : le système peut afficher l'information *une* ou *plusieurs* fois.
   Un guide peut n'afficher les gestes qu'au début (i.e. retour d'information prédictif), comme avec une Crib-Sheet [Bau et Mackay, 2008] ou un Marking Menu [Kurtenbach *et al.*, 1993] ou bien ne montrer le résultat qu'après l'exécution du geste (i.e. retour d'information rétroactif) [Freeman *et al.*, 2009]. En revanche, un Hierarchical Marking Menu [Kurtenbach *et al.*, 1994] met à jour les commandes disponibles à chaque fois que l'utilisateur passe un niveau du menu - donc en plusieurs étapes, de façon discrète (axe *mise à jour*).

Un dernier axe décrit les contraintes temporelles imposées sur l'exécution du geste. Cet axe est donc spécifique au retour d'information prédictif :

3) *Retour d'information prédictif - Rythme* : le système peut présenter les gestes à effectuer avec un rythme choisi par *l'utilisateur* ou bien imposé par le *système*.
   Par exemple, la condition 3D Self-Guided Arrow de LightGuide [Sodhi *et al.*, 2012] projette une flèche 3D sur la main de l'utilisateur et laisse ce dernier choisir à quel rythme il veut exécuter son geste. Au contraire, un avatar que l'utilisateur doit imiter en temps réel impose une vitesse d'exécution [Kamal *et al.*, 2014].
   Cette dimension ne fait pas référence à la vitesse à laquelle le geste sera présenté à l'utilisateur (comme avec une vidéo [Kamal *et al.*, 2014]), mais fait référence à la contrainte imposée quant à la vitesse d'exécution du geste. L'option "imposée par le système" est principalement utilisé lorsque la vitesse du mouvement est pertinente, comme lorsque l'on cherche à guider des mouvements de danse par exemple [Anderson *et al.*, 2013].

8.1.2.1.3 Mode de guidage Nous définissons la façon dont les mécanismes de retour d'information rétroactif et de retour d'information prédictif se terminent après avoir été activés. Deux axes caractérisent la fin du mode de guidage :

1) *Maintien du mode* : le mécanisme peut être *transitoire* ou *soutenu*.
Un mécanisme transitoire se termine automatiquement, dès lors qu'un geste est reconnu [Bau et Mackay, 2008] ou bien si le système surveille les performances de l'utilisateur et juge que ce dernier n'a plus besoin des informations de guidage [Anderson *et al.*, 2013]. Dans le cas d'un mode soutenu, nous détaillons son interruption :

2) *Interruption du mode* : le mode peut être soutenu à la manière d'un *quasimode* ou peut être *persistant*.
Un exemple de quasimode est le relâchement d'un bouton dédié [Schwaller *et al.*, 2014]. Quitter un mode persistant peut se faire en appuyant sur un bouton étiqueté "fermer" affiché à l'écran [Bragdon *et al.*, 2009].

8.1.2.2 Groupe d'axes : *Quoi*

Ces axes caractérisent l'information communiquée par les mécanismes de retour d'information rétroactif et de retour d'information prédictif (Figure 8.8). Le mécanisme de retour d'information rétroactif fournit deux sources d'information : les informations sur les actions passées effectuées par l'utilisateur (Figure 8.8, A) et celles concernant le traitement de ces dernières actions par le système (Figure 8.8, B). Le mécanisme de retour d'information prédictif fournit des informations sur les actions futures seulement (Figure 8.8, C). Ces trois sources d'information ont trois axes en commun pouvant néanmoins avoir des options de conception différentes : le(s) geste(s) affiché(s) (Figure 8.8, axes A1 , B1 et C1), les caractéristiques impliquées dans cette représentation telle que la forme géométrique du geste ou bien la vitesse d'exécution par exemple ((Figure 8.8, axes A2 , B2 et C2) et la portion perçue grâce à cette représentation (Figure 8.8, axes A3 , B3 et C3).

8.1.2.2.1 Geste(s) affiché(s) (axes A1 , B1 et C1)

Une technique de guidage fournit des informations concernant les gestes passés et/ou futurs. Les axes suivants caractérisent ces informations à propos : des gestes effectués (retour d'information rétroactif, 1), des gestes prévus et/ou reconnus (retour d'information rétroactif, 2), et des gestes gérés par le système (retour d'information prédictif, 3).

1. Axe A1 . *Retour d'information rétroactif - Évaluation par l'utilisateur, Geste affiché* : cet axe caractérise la possibilité qu'offre le système à l'utilisateur de se corriger en jugeant lui-même le geste effectué et ne comporte qu'une seule option. Afin que l'utilisateur puisse évaluer son geste, le système peut afficher le geste *effectué*, ou pas.

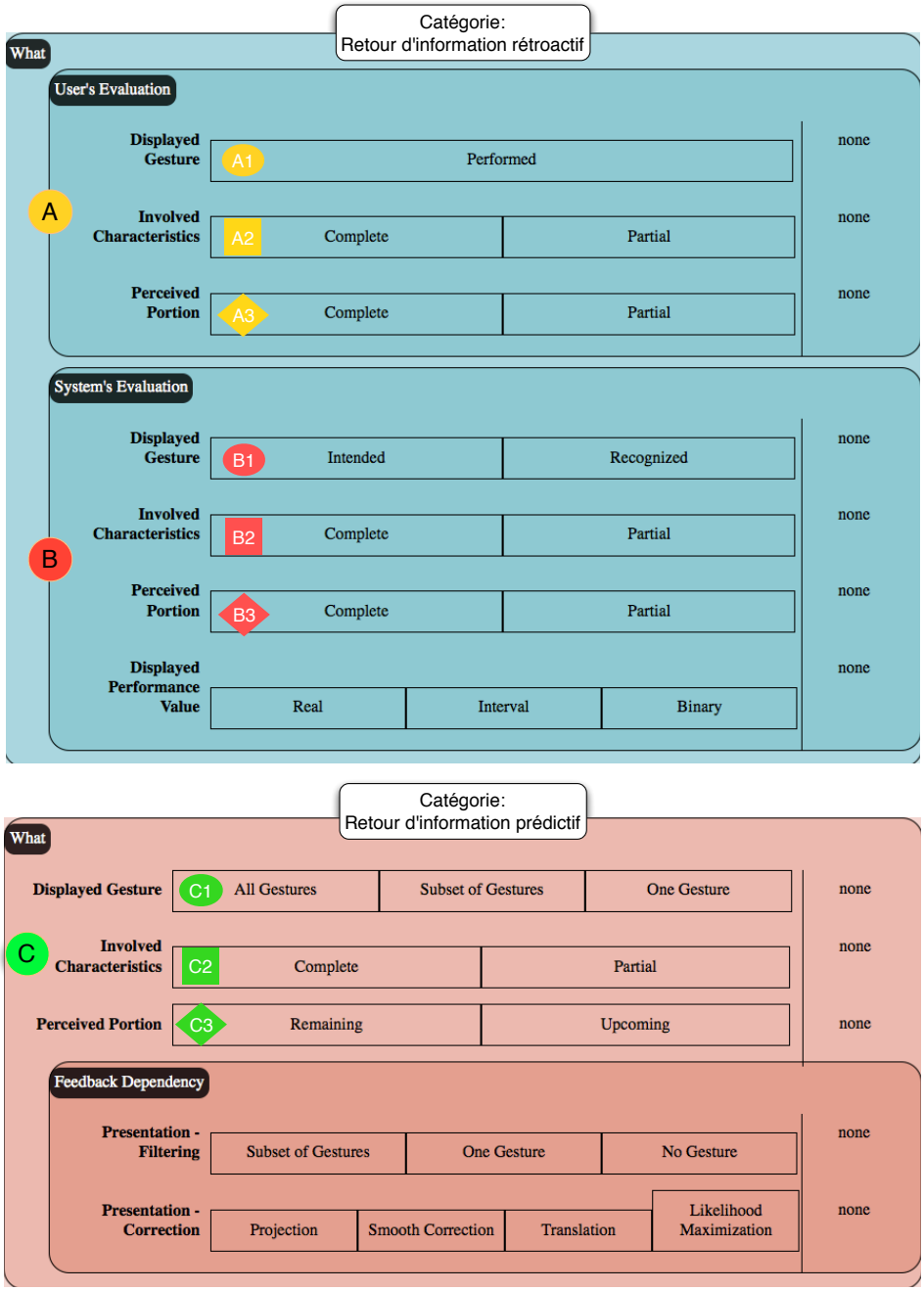


FIGURE 8.8 – Groupes *Quoi* du retour d'information rétroactif (haut) et du retour d'information prédictif (bas). Le retour d'information rétroactif fournit de l'information quant aux actions précédemment effectuées (A) ainsi qu'au traitement de ces actions par le système (B). Le retour d'information prédictif fournit de l'information sur les actions futures (C). Ces sous-groupes ont trois axes similaires concernant les gestes affichés (1), les caractéristiques impliquées (2) et les portions affichées (3).

C'est le cas de systèmes affichant le chemin déjà effectué avec un tracé depuis son point de départ jusqu'à la position courante du curseur par exemple [Anderson et Bischof, 2013]. Au contraire, un guidage via une vidéo de démonstration [Kamal *et al.*, 2014] n'affiche pas le geste effectué par l'utilisateur.

2. Axe B1 . *Retour d'information rétroactif - Évaluation par le système, Geste affiché* : le système peut représenter le geste *prévu* et/ou le geste *reconnu*.
   Ainsi, l'utilisateur peut comparer le geste effectué avec le geste prévu tout en sachant ce que le système a reconnu. Un avatar 3D montrant le mouvement que l'utilisateur aurait dû reproduire illustre un affichage du geste prévu [Kamal *et al.*, 2014]. SimpleFlow [Bennett *et al.*, 2011] affiche le geste qui sera reconnu si l'utilisateur relâche le bouton de souris.
3. Axe C1 . *Retour d'information prédictif - Geste affiché* : le système de guidage peut révéler *tous* les gestes disponibles, un *sous-ensemble* de gestes ou *un seul* geste.
   La première étape de ShadowGuides [Freeman *et al.*, 2009] ou le 3D Marking Menu [Ren et O'Neill, 2012] révèlent tous les gestes disponibles. Au contraire, la seconde étape de ShadowGuides [Freeman *et al.*, 2009] guide l'exécution d'un *sous-ensemble* de gestes, c'est-à-dire ceux qui sont disponibles à partir de la posture de la main effectuée par l'utilisateur lors de la première étape. Enfin, GestureBar [Bragdon *et al.*, 2009] propose une barre d'outil ne représentant que le geste associé à la commande désirée.

Ces axes caractérisent le(s) geste(s) affiché(s) (tous les gestes, un sous-ensemble ou un seul geste). Les gestes affichés par le guide peuvent transmettre plusieurs informations à l'utilisateur. Lorsque l'on parle de gestes, on pense souvent en premier lieu à leur forme géométrique (e.g., dessiner un *carré* ou un *cercle*). Il y a cependant un large éventail de caractéristiques liées aux gestes, autres que la forme. Nous nous concentrons maintenant sur les caractéristiques du geste rendues perceptibles par le guide.
La pertinence des caractéristiques dépend fortement de l'algorithme de reconnaissance utilisé par le système ainsi que du contexte dans lequel l'utilisateur sera guidé. Dans un contexte d'interaction gestuelle 3D dans l'air, effectuer un grand carré horizontal pourrait déclencher une commande différente d'un petit carré horizontal. La technique de guidage peut offrir de l'information quant à l'ampleur du geste, son orientation, mais aussi sa localisation (par exemple, effectuer un geste "effacer" doit être effectué sur un objet contextuel), la force (par exemple, un carré 2D peut être dessiné avec des lignes plus épaisses aux endroits où l'utilisateur a appliqué plus de pression sur la surface), etc. On peut aussi mentionner les caractéristiques temporelles d'un geste. Ces caractéristiques peuvent être utiles lors du guidage de gestes tels que des mouvements de danse par exemple, à savoir, la durée totale du mouvement, les vitesses et accélérations locales, etc. Certaines techniques telles que YouMove considèrent des caractéristiques comme la régularité du mouvement ou bien la stabilité de la posture de l'utilisateur [Anderson *et al.*, 2013]. Au contraire, certains algorithmes de reconnaissance de geste fournissent un moyen de ne pas tenir compte de certaines caractéristiques telles que l'échelle ou l'orientation du

geste en entrée [Kratz et Rohs, 2010].
Le but n'est pas de faire une liste exhaustive de toutes les caractéristiques. Nous nous concentrons sur une description à plus haut niveau du guide. Pour cela, nous définissons deux axes : l'un concernant les caractéristiques impliquées dans la représentation des gestes et un autre axe mettant l'accent sur la quantité d'information fournie à l'utilisateur :

1) Axes A2 , B2 et C2 . *Caractéristiques impliquées* : le système peut fournir des informations concernant l'ensemble *complet* des caractéristiques (par exemple la forme, la force, la position, etc.) ou un ensemble *partiel*. Il est à noter qu'un ensemble de caractéristiques complet ne se réfère pas à une liste exhaustive des caractéristiques du geste, mais uniquement aux caractéristiques pertinentes pour le domaine d'application.
Ainsi, si un guide vise à gérer seulement la forme du geste et la pression appliquée pendant l'exécution par exemple, son mécanisme de retour d'information rétroactif impliquera l'ensemble complet des caractéristiques si il fournit de l'information quant à la forme et à la pression. Les techniques existantes cherchent en majorité à guider l'utilisateur quant à la forme géométrique des gestes [Bau et Mackay, 2008] [Ghomi *et al.*, 2013] [Rovelo *et al.*,]. De fait, fournir un retour d'information rétroactif ou un retour d'information prédictif à propos d'une seule caractéristique revient à considérer que ces guides considèrent l'ensemble complet des caractéristiques. YouMove [Anderson *et al.*, 2013] est un exemple de technique utilisant un sous-ensemble des caractéristiques de geste à guider. Cette technique propose un guidage comportant différentes étapes. Ces étapes impliquent chacune différentes caractéristiques, telles que la stabilité lors du maintien d'une posture ou bien la dextérité durant l'exécution d'un geste. Ainsi, chaque étape propose un ensemble partiel de caractéristiques que le système propose de guider.

2) Axes A3 , B3 et C3 . *Portion perçue* : le mécanisme de retour d'information rétroactif (resp. retour d'information prédictif) peut représenter les gestes de façon à ce que l'utilisateur perçoive l'information complète, c'est-à-dire la portion *complète* du geste (resp. la portion totale *restante* du geste) ou bien une portion *partielle* du geste (resp. une portion *imminente*) (Figure 8.9). Cet axe décrit à quel point l'utilisateur aura connaissance du geste. Cette décision de conception pourrait être motivée par le fait de vouloir minimiser la quantité d'information présentée à l'utilisateur.
Un tracé du geste effectué depuis son point de départ fournit une représentation complète comme retour d'information rétroactif [Anderson et Bischof, 2013]. Au contraire, Gestu-Wan n'affiche qu'une portion partielle du geste effectué en ne montrant que la dernière posture 3D reconnue par le système [Rovelo *et al.*,]. De manière similaire, un utilisateur guidé par une Crib-Sheet perçoit la totalité restante des gestes [Bau et Mackay, 2008] grâce au retour d'information prédictif, alors qu'avec un Hierarchical Marking Menu [Kurtenbach *et al.*, 1994], l'utilisateur ne perçoit que la partie imminente des gestes, à savoir la direction des éléments

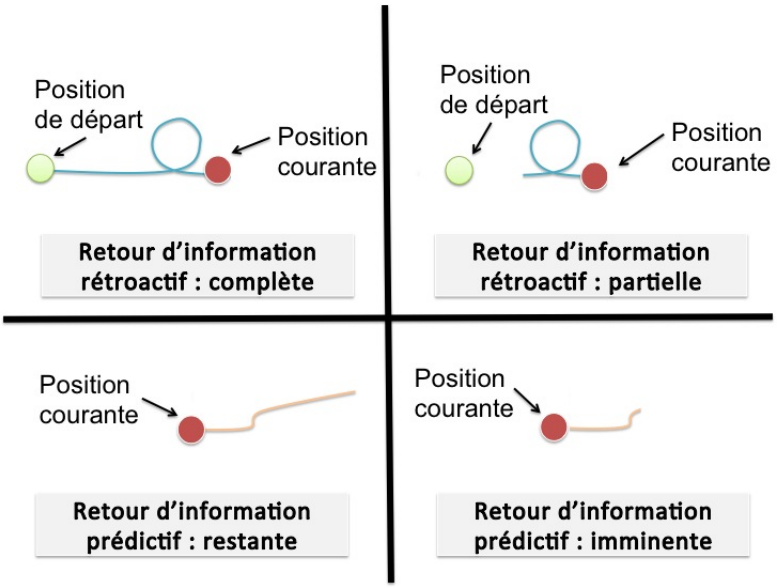


FIGURE 8.9 – Illustration de l'axe *Portion perçue*. Ligne du haut : le retour d'information rétroactif représente la portion de geste effectué complète (gauche) ou une portion partielle (droite). Ligne du bas : le retour d'information prédictif représente la portion restante du geste à effectuer (gauche) ou seulement la partie imminente (droite).

du niveau courant du menu.

8.1.2.2.2 Valeur de performance pour le retour d'information rétroactif La technique de guidage peut évaluer l'exécution du geste par l'utilisateur. Dans le cadre de l'information fournie par le mécanisme de retour d'information rétroactif, nous considérons l'affichage d'une valeur de performance résultant d'une comparaison entre le geste effectué et le geste attendu ou reconnu :

Valeur de performance affichée : le système peut fournir une valeur *binaire*, un *intervalle* ou bien une valeur *réelle*. Cette valeur doit ensuite être interprétée par l'utilisateur.
Une valeur binaire peut être le résultat d'un système de reconnaissance, comme "reconnu" ou "non reconnu" [Kurtenbach *et al.*, 1993]. Le système de reconnaissance peut aussi renvoyer le score d'un geste donné sous forme de valeur réelle [Kamal *et al.*, 2014] (comme un pourcentage par exemple). Le système peut aussi utiliser une métrique personnalisée telle que la distance entre la position actuelle des mains et celle prévue [Portillo-Rodriguez *et al.*, 2008].

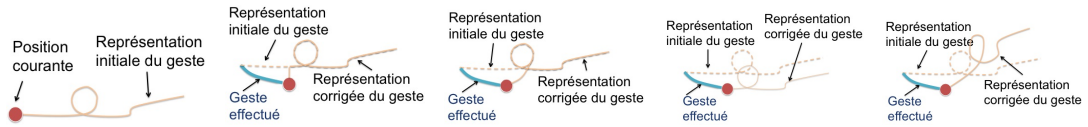
8.1.2.2.3 Impact du retour d'information rétroactif sur le retour d'information prédictif Deux axes dédiés au retour d'information prédictif caractérisent comment son contenu informationnel est impacté par les actions passées, et donc par le mécanisme de retour d'information rétroactif. Nous considérons deux impacts potentiels :

1) *Présentation - Filtrage* : en se basant sur ce que l'utilisateur a déjà effectué, le système peut retirer des gestes parmi ceux révélés au début du guidage. Pendant l'exécution d'un geste, le filtrage peut être *partiel* et redéfinir un sous-ensemble de geste ou *un seul* geste à représenter. Le filtrage peut aussi être complet et ainsi ne présenter *aucun* geste à l'utilisateur.
OctoPocus révèle tous les gestes lorsque le guidage est enclenché [Bau et Mackay, 2008]. Par la suite, pendant l'exécution d'un geste, ceux ayant un score en-dessous d'un seuil limite ne sont plus présentés à l'utilisateur. Il s'agit donc d'un filtrage partiel. On peut imaginer une version d'OctoPocus qui ne laisserait qu'un seul geste - celui avec le plus haut score de reconnaissance. L'Adaptive Guide dispose d'un filtrage complet puisqu'il retire le guidage pendant l'exécution d'un geste, laissant ainsi l'utilisateur finir l'exécution du geste par lui-même [Anderson et Bischof, 2013].

2) *Présentation - Correction* : en se basant sur ce que l'utilisateur a déjà effectué, la technique de guidage peut modifier le contenu du retour d'information prédictif. Nous listons cinq corrections possibles : une *projection*, une *correction lisse*, une *re-localisation* et une *maximisation du score*.
Si la trajectoire du geste effectué par l'utilisateur s'écarte du geste prévu, la technique de guidage peut modifier la représentation du geste pour faire en sorte que la position actuelle de l'élément à guider (doigt, curseur de souris, main, etc.) soit redirigée vers le point le plus proche sur le geste prévu, c'est-à-dire grâce à une *projection* (Figure 8.10b). Le guide peut aussi corriger la trajectoire de manière plus *lisse* en ramenant progressivement l'utilisateur sur la trajectoire du geste à effectuer (Figure 8.10c). Le guide peut également modifier la position du tracé du geste et *re-localiser* ce tracé pour le relier à la position courante (Figure 8.10d). Enfin, le guide peut modifier la fin du tracé afin de faire en sorte que la combinaison "partie effectuée" + "partie restante" *maximise le score* de reconnaissance du geste si exécutée (Figure 8.10e). Par exemple, si l'utilisateur s'écarte de la trajectoire prévue en étant guidé par Follow Spot [Sodhi *et al.*, 2012], alors le système ramènera la main sur la trajectoire initiale grâce à une projection. Le guide 3D Arrow [Sodhi *et al.*, 2012] ou le mécanisme de correction du système de guidage pour des mouvements du corps [Schönauer *et al.*, 2012] permettent une correction lisse de la main. Enfin, OctoPocus [Bau et Mackay, 2008] re-localise la représentation des gestes sous le curseur.

8.1.2.3 Groupe d'axes : *Comment*

Ce groupe d'axes décrit comment les informations des mécanismes de retour d'information rétroactif et de retour d'information prédictif transmettent les informations de guidage (Figure 8.11). Nous mettons l'accent sur les modalités de sortie, et non sur la forme perceptible finale des guides. En premier lieu, nous considérons la modalité sensorielle humaine utilisée pour percevoir les informations de guidage.



(a) Début (b) Projection (c) Correction lisse (d) Re-localisation (e) Maximisation du score

FIGURE 8.10 – Illustration de l'axe *Présentation - Correction*. (a) Etat initial. (b) Projection : le geste représenté comprend un segment supplémentaire pour ramener l'utilisateur sur le point le plus proche du geste à effectuer. (c) Correction lisse : le geste représenté comprend un segment supplémentaire pour progressivement ramener l'utilisateur sur le geste initial. (d) Re-localisation : le geste est représenté selon la position actuelle de l'utilisateur. (e) Maximisation du score : la combinaison du geste représenté et de la partie du geste déjà effectuée maximise le score de reconnaissance du geste.

How				
Sensory Modality	Visual	Audio	Haptic	none
Perspective	Egocentric	Tethered	Exocentric	none
Reference Frame	Relative	Absolute		none

FIGURE 8.11 – Groupe *Comment* commun au retour d'information rétroactif et au retour d'information prédictif.

Modalité sensorielle : le guide peut transmettre de l'information via une modalité sensorielle visuelle, auditive ou tactile.
En effet, le guide peut afficher la forme des gestes [Bragdon *et al.*, 2009], utiliser des vibrations pour transmettre l'échelle de l'écart entre le geste prévu et celui réalisé [Portillo-Rodriguez *et al.*, 2008], ou bien utiliser un son afin d'indiquer le succès ou l'échec d'un geste [Kamal *et al.*, 2014].

Les axes suivants décrivent plus en détail la modalité visuelle. C'est en effet la modalité la plus exploitée dans la littérature. Nous considérons qu'une modalité est définie comme un couplage entre un langage d'interaction et un dispositif physique [Vernier et Nigay, 2001]. Plutôt que de détailler les différents moyens de représenter visuellement un geste, nous adoptons une approche avec un plus haut niveau d'abstraction en considérant les options de conception du langage de la modalité visuelle. Nous définissons deux axes : la *perspective* et le *référentiel*.

1) *Perspective* : l'affichage des gestes peut être *égocentrique* (c'est-à-dire du point de vue de l'utilisateur [Sodhi *et al.*, 2012]) ou *exocentrique* (c'est-à-dire du point de vue d'une troisième personne [Portillo-Rodriguez *et al.*, 2008]). Il existe entre ces deux options des perspectives dites *couplées*, c'est-à-dire du point de vue d'une troisième personne, mais tout de même reliées au point de vue de l'utilisateur, comme avec un miroir augmenté muni d'un mécanisme de suivi de tête [Anderson

et al., 2013].
Pour simplifier la compréhension de cet axe, nous pouvons faire un parallèle avec les jeux vidéos de courses de voiture (Figure 8.12). La perspective dans le cockpit est *égocentrique* (a). La perspective à l'extérieur du véhicule mais suivant la voiture est *couplée* (b). Enfin, la perspective rendue par la carte dans le coin bas-gauche de l'image est *exocentrique* (c).

(a) (b) (c)

FIGURE 8.12 – Illustrations [1] de l'axe *Perspective* : parallèle avec un jeu vidéo de courses. (a) Perspective égocentrique. (b) Perspective couplée. (c) Perspective exocentrique. Ces perspectives sont aussi présentes lorsqu'une technique de guidage affiche des informations visuelles.

2) *Référentiel* : l'information peut être affichée de manière *relative* par rapport à la position du membre de l'utilisateur à guider, ou bien dans un référentiel *absolu*. Plusieurs techniques de guidage affichent les gestes centrés sur la position du curseur [Kurtenbach *et al.*, 1993, Bau et Mackay, 2008]. Au contraire, la première étape de l'Interactive Crib-Sheet affiche les gestes à un emplacement prédéfini, indépendamment de la position courante du curseur [Kurtenbach *et al.*, 1994].

Nous soulignons que le *Référentiel* et la *Perspective* de la modalité visuelle sont deux axes indépendants. Le Tableau 8.1 illustre avec des exemples que chaque combinaison d'options est possible.

1. Sources des images : http://blog.caranddriver.com/test-drive-ferrari-racing-legends-video-game-review/ (lien valide en septembre 2015) et http://www.gameinformer.com/games/nascar_the_game_2011/b/xbox360/archive/2011/03/29/nascar-review.aspx (lien valide en septembre 2015)

Perspective	Référentiel	
	Relatif	Absolu
Egocentrique	L'option de guidage *Ghosted Hints* [White *et al.*, 2007] affiche le geste à effectuer grâce une série d'images semi-transparentes à travers des lunettes de Réalité Augmentée (égocentrique) à partir de l'objet tenu en main par l'utilisateur (relatif).	LightGuide [Sodhi *et al.*, 2012] affiche le geste à effectuer grâce à des indices visuels projetés sur la main de l'utilisateur (égocentrique). La projection ne suit pas la main de l'utilisateur, mais est ancrée à un emplacement spécifique dans le monde physique et son référentiel (absolu).
Couplée	YouMove [Anderson *et al.*, 2013] affiche sur un miroir (couplage) le squelette d'un enseignant aligné sur la réflexion de l'utilisateur (relatif).	Exemple fictif : l'avatar 3D d'un enseignant vu de face exécutant un mouvement. L'utilisateur peut bouger sa tête et ainsi modifier son point de vue grâce à un mécanisme de couplage entre la tête et la caméra de la scène 3D (couplage). Le mouvement est exécuté sans considérer la position ou bien la posture de l'utilisateur (absolu).
Exocentrique	YouMove [Anderson *et al.*, 2013] affiche une vue de côté (exocentrique) du squelette de l'utilisateur aligné à celui de l'enseignant (relatif).	YouMove [Anderson *et al.*, 2013] affiche une vidéo d'un enseignant exécutant le mouvement de face (exocentrique). La vidéo est enregistrée, donc sans lien avec l'utilisateur (absolu).

TABLE 8.1 – Exemples montrant l'orthogonalité des axe *Perspective* et *Référentiel*.

8.1.2.4 Groupe d'axes : *Où*

Ce groupe décrit la relation spatiale entre l'emplacement sur lequel l'utilisateur doit se concentrer pour percevoir les informations de guidage et celui sur lequel il doit se concentrer pour réaliser l'action (Figure 8.13). Avec cette définition, la localisation spatiale d'une technique de guidage pour des gestes 2D effectués avec une souris se définit par

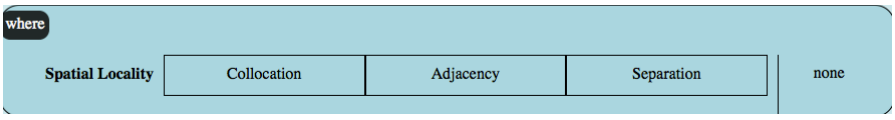


FIGURE 8.13 – Groupe *Où* commun au retour d'information rétroactif et au retour d'information prédictif.

rapport à l'emplacement du curseur, et non par rapport à celui de la souris. Cet axe impacte le choix des dispositifs physiques qui serviront à transmettre les informations de guidage [Vernier et Nigay, 2001].

Localisation spatiale : l'emplacement pour percevoir les information de guidage et celui pour réaliser le geste peuvent être *colocalisés*, *adjacents* ou *séparés*. LightGuide projette des informations visuelles colocalisées sur la main de l'utilisateur à guider [Sodhi *et al.*, 2012]. ShadowGuides [Freeman *et al.*, 2009] affiche les gestes dans une zone adjacente à celle réservée à l'exécution des gestes, ceci afin d'éviter de potentiels problèmes d'occultation. Pour des raisons techniques, les techniques de guidage destinées aux gestes 3D utilisent souvent un affichage séparé afin d'éviter d'instrumenter l'utilisateur [Anderson *et al.*, 2013, Kamal *et al.*, 2014, Portillo-Rodriguez *et al.*, 2008].

Nous avons présenté les axes de notre espace de conception. Certains axes sont inspirés d'axes d'espaces de conception existants. Nous décrivons ces axes dans le paragraphe 8.1.3). Nous décrivons ensuite les techniques de guidage qui nous ont amené à une nouvelle option sur un axe ou bien à un nouvel axe complet dans le paragraphe 8.1.4.

8.1.3 Positionnement : espaces de conception

Certains axes de notre espace sont inspirés d'espaces de conception existants. Nous ne considérons que les options de conception, et non les recommandations de conception, c'est-à-dire ce que le système peut faire plutôt que ce qu'il devrait faire. En effet, certains travaux recommandent de respecter certains principes de conception. Par exemple, [Kurtenbach *et al.*, 1994] définissent des principes de haut niveau afin de faciliter l'exécution ainsi que l'apprentissage des gestes. [Bragdon *et al.*, 2009] étendent ces principes en mettant l'accent sur l'accessibilité des gestes. Enfin, [Anderson *et al.*, 2013] définissent des lignes directrices pour aider le guidage et l'apprentissage de mouvements corporels. On peut toutefois noter qu'il y a peu de recommandations puisque peu d'options de conception ont été identifiées. En effet, il convient d'identifier ces options de conception avant de pouvoir identifier leur impact sur le guidage.

Pour décrire les options de conception identifiées dans la littérature, nous considérons les deux aspects fondamentaux du guidage qui structurent notre espace de conception : le retour d'information rétroactif et le retour d'information prédictif. Bien que certaines études considèrent les deux aspects [Bau et Mackay, 2008], certaines ne se concentrent

que sur l'un d'entre eux, c'est-à-dire le retour d'information rétroactif [Schönauer *et al.*, 2012] ou le retour d'information prédictif [Sodhi *et al.*, 2012].

8.1.3.1 Retour d'information rétroactif

[Bau et Mackay, 2008] définissent quatre dimensions pour caractériser le mécanisme de retour d'information rétroactif dédié à l'interaction gestuelle :

1. Tout d'abord, ils définissent la *valeur de reconnaissance* que présente le système à l'utilisateur. Cette valeur peut être binaire (comme "oui" ou "non"), un intervalle ou bien une valeur réelle (comme le score de reconnaissance par exemple).
   Notre espace de conception intègre cet axe.
2. En second lieu, ils définissent le *filtrage*, qui décrit comment le retour d'information rétroactif modifie l'ensemble des gestes affichés. Après ce filtrage, le guide peut afficher tous les gestes, une sous-partie ou la meilleure option.
   Nous intégrons cet axe dans notre espace de conception et ajoutons une option : "aucun". Cette option est illustrée par Adaptive Guide qui, après le filtrage, n'affiche aucun geste à l'utilisateur [Anderson et Bischof, 2013]. En effet, Adaptive Guide supprime la représentation de plus en plus tôt pendant l'exécution du geste au fur et à mesure que le résultat s'améliore. Cette solution empêche l'utilisateur de trop se reposer sur le guide. Cette option vise à améliorer l'apprentissage des gestes. Plutôt que d'inclure cet axe dans la partie "retour d'information rétroactif", nous avons créé un sous-groupe spécifique concernant l'impact du retour d'information rétroactif dans la partie "retour d'information prédictif".
3. Troisièmement, ils définissent la *fréquence de mise à jour*. Le système peut fournir un retour d'information rétroactif en continu pendant l'exécution du geste, en plusieurs étapes ou une seule fois.
   Dans notre espace de conception, cet axe est divisé en deux afin de souligner la différence entre continu et discret (plusieurs étapes ou une seule fois).
4. Enfin, ils définissent l'axe *représentation*, qui décrit la façon dont le retour d'information rétroactif est présenté à l'utilisateur en s'appuyant sur des caractéristiques visuelles.
   Puisque nous nous concentrons sur le comportement des guides et non sur leur forme concrète, nous n'intégrons pas cet axe à notre espace de conception. Ce type de description est utile lors de l'étape suivante dans un processus de conception de technique de guidage : la définition de la forme concrète du guide, après avoir défini son comportement.

[Schönauer *et al.*, 2012] effectuent une analyse en profondeur du retour d'information rétroactif multimodal pour des mouvements du corps. Ils traitent neuf aspects du guidage, dont six ne présentent aucune option de conception relative au comportement du guide. Trois aspects sont donc pertinents pour la conception du comportement d'une technique de guidage.

1. Tout d'abord, ils examinent les modalités envisageables (visuelle, auditive ou tactile par exemple).
   Notre espace de conception intègre cet axe dans nos deux catégories : le retour d'information rétroactif et le retour d'information prédictif.
2. En second lieu, ils examinent la localisation spatiale du retour d'information rétroactif car celui-ci pourrait entrainer un changement d'attention ou générer des interférences entre stimuli (comme avec deux retours d'information rétroactifs tactiles trop proches l'un de l'autre par exemple). En revanche, les auteurs ne proposent aucune option de conception.
   Nous affinons cet axe avec trois relations de Allen appliquées aux relations spatiales [Vernier et Nigay, 2001] : la colocalisation, l'adjacence et la séparation. Cela nous permet de distinguer des techniques à priori similaires telles qu'un guide statique positionné sous le curseur ou dans un coin de l'écran [Anderson et Bischof, 2013].
3. Enfin, ils définissent un comportement du système afin de corriger le geste de l'utilisateur en se basant sur la synchronisation et la rapidité du mouvement : utiliser une prédiction de mouvements afin de progressivement ramener l'utilisateur vers le geste correct. [Sodhi *et al.*, 2012] proposent également cette correction ainsi que d'autres options : la "projection" et la "re-localisation".
   Notre espace de correction contient ces trois options sur l'axe *Présentation - Correction* Figure 8.10. Cet axe contient une option supplémentaire, à savoir la "maximisation du score". Cette option définit une nouvelle correction afin de maximiser le score de reconnaissance du geste en cours. Comme pour l'axe de filtrage, l'axe *Présentation - Correction* appartient au sous-groupe qui caractérise l'impact du retour d'information rétroactif sur le retour d'information prédictif.

8.1.3.2 Retour d'information prédictif

[Bau et Mackay, 2008] définissent deux dimensions pour caractériser le mécanisme de retour d'information prédictif.

1. Tout d'abord, ils définissent le niveau de détail fourni par le mécanisme de retour d'information prédictif. Le guide peut représenter tout le geste, seulement une portion ou simplement fournir un indice minimaliste quant à la direction à suivre.
   Notre espace de conception contient cet axe et le considère à la fois pour le retour d'information rétroactif et pour le retour d'information prédictif. Pour ce faire, nous utilisons des termes plus génériques tels que "portion imminente" et "portion restante" pour le retour d'information prédictif et "portion complète" et "portion partielle" pour le retour d'information rétroactif. Nos définitions permettent alors de caractériser le niveau de connaissance qu'aura l'utilisateur des gestes passés et futurs.
2. En second lieu, ils définissent la fréquence de mise à jour du mécanisme de retour d'information prédictif (une fois, en plusieurs étapes ou en continu).

Dans notre espace de conception, la fréquence de mise à jour définit deux axes afin de faire la distinction entre une fréquence continue et discrète (premier axe) qui peut elle-même apparaître une fois ou plusieurs fois (second axe).

[Sodhi *et al.*, 2012] se concentrent sur la conception d'indices visuels projetés sur la main de l'utilisateur dans un contexte de guidage de mouvement pour de la ré-éducation musculaire. Bien qu'ils distinguent le retour d'information rétroactif du retour d'information prédictif, leur intérêt se porte principalement sur le contenu du retour d'information prédictif ainsi que sa représentation. Néanmoins, les auteurs définissent deux axes décrivant la comportement du guide : le rythme et la perspective.

1. Le rythme fait référence à la vitesse à laquelle l'utilisateur pourra effectuer son geste, et donc à la vitesse à laquelle le retour d'information prédictif présentera les informations de guidage.
   Notre espace de conception contient cet axe.
2. La perspective fait référence au point de vue à partir duquel l'utilisateur verra le geste. Les auteurs considèrent les points de vue égocentrique et exocentrique.
   Notre espace de conception inclut cet axe en ajoutant une option supplémentaire : le point de vue couplé issu de [Milgram et Colquhoun, 1999].

Les études précédemment citées définissent des axes et des options de conception de notre espace de conception. La contribution de celui-ci réside principalement dans le travail effectué afin de recueillir, d'unifier, d'organiser et d'affiner les options pertinentes dans un espace de conception compréhensible et utilisable. Ainsi, ce dernier (1) intègre les axes de conception existants, (2) les unifie dans une structure commune et cohérente et (3), les étend afin de fournir un espace de conception aussi approfondi que possible pour déterminer le comportement des techniques de guidage (Quand, Quoi, Comment, Où). La distinction entre retour d'information rétroactif et retour d'information prédictif souligne le fait que les concepteurs devraient envisager ces options de conception pour ces deux aspects.

8.1.4 Positionnement : techniques de guidage

Afin de créer notre espace de conception, nous avons recensé 46 techniques de guidage, qui, en prenant en compte les techniques ayant plusieurs étapes, mènent à 60 descriptions. Les techniques de guidage peuvent en effet comporter plusieurs étapes ayant des descriptions différentes. Nous avons souvent cité dans ce chapitre des étapes de techniques existantes. Par exemple, nous avons cité les deux étapes de ShadowGuides [Freeman *et al.*, 2009] afin d'illustrer l'axe *Retour d'information prédictif - Geste affiché* (item 3) du groupe *Quoi* dans le paragraphe 8.1.2.2 (page 96). Afin d'éviter une liste exhaustive des 60 descriptions à effectuer au regard de 35 axes, nous décrivons partiellement les guides ayant mené aux nouveaux axes ainsi qu'à de nouvelles options de notre espace de conception. Cette liste exhaustive peut néanmoins se trouver via un outil interactif décrit paragraphe 8.2 et accessible via l'adresse http://iihm.imag.fr/delamare/guidance (lien valide en septembre

2015).

Les techniques de guidage existantes proposent différentes caractéristiques temporelles. Par exemple, ShadowGuides, (Hierarchical) Marking Menus, OctoPocus et YouMove utilisent un temps de pause afin de déclencher le guidage [Anderson *et al.*, 2013] [Bau et Mackay, 2008] [Freeman *et al.*, 2009] [Kurtenbach *et al.*, 1993] [Kurtenbach *et al.*, 1994]. Les concepteurs justifient cette méthode par le fait que le guide peut apparaitre lorsque l'utilisateur devient hésitant. En revanche, GestureBar [Bragdon *et al.*, 2009] et les guides développés lors de l'étude de [Anderson et Bischof, 2013] apparaissent lorsque l'utilisateur clique dans une zone dédiée (comme un bouton par exemple). Nous avons donc ajouté trois axes - à savoir Initiative, Modalité et Désambigüisation - afin de caractériser le *déclenchement du guide* dans le groupe *Quand*. Par conséquent, nous pouvons distinguer les techniques de guidage qui apparaissent automatiquement (comme ShadowGuides [Freeman *et al.*, 2009]) des techniques qui nécessitent une action supplémentaire de la part de l'utilisateur (comme GestureBar [Bragdon *et al.*, 2009]).
De manière similaire, nous avons voulu faire la distinction entre différentes sorties du mode de guidage. Certains guides tels que ShadowGuides [Freeman *et al.*, 2009] restent affichés en permanence. D'autres disparaissent sans action supplémentaire de la part de l'utilisateur, après avoir terminé l'exécution d'un geste avec un Marking Menu [Kurtenbach *et al.*, 1994] par exemple. Enfin, certains guides nécessitent une action explicite (appuyer sur un bouton par exemple [Bragdon *et al.*, 2009]) afin de sortir du mode de guidage. Nous avons donc ajouté deux axes - à savoir Maintien et Interruption du mode - afin de caractériser le *mode du guide* dans le groupe *Quand*.

Nous avons considéré tous les types de guides dans notre étude, c'est-à-dire ceux visant à guider des gestes afin d'interagir avec un système [Freeman *et al.*, 2009] (avec des gestes symboliques (paragraphe 2.2)) et ceux visant à guider des mouvements de danse par exemple [Anderson *et al.*, 2013], donc sans sémantique. Ces différents types de guides sont utilisés dans différents contextes et ont ainsi accès à différentes informations. Par exemple, un guide pour des mouvements de danse connait le geste en cours d'exécution alors qu'un guide pour des gestes interactifs essaiera au contraire de le deviner grâce à des scores de reconnaissance. Il en est de même pour les caractéristiques impliquées durant le guidage du geste : certains guident ne considèrent que la forme du geste [Sodhi *et al.*, 2012] alors que d'autres cherchent à guider plusieurs aspects d'un mouvement, tels que la précision ou bien le rythme [Anderson *et al.*, 2013]. Les axes *geste(s) affiché(s)* et *caractéristiques impliquées* dans les groupes *Quoi* du retour d'information rétroactif et du retour d'information prédictif permettent de capturer ces différences en terme de comportement des guides sans avoir à créer une liste exhaustive de ces caractéristiques.

Enfin, les techniques utilisant la modalité visuelle peuvent afficher les gestes de deux manières différentes : bien que YouMove [Anderson *et al.*, 2013] et le système utilisé par [Kamal *et al.*, 2014] présentent le geste via l'animation d'un squelette, l'un relie ce squelette à la posture de l'utilisateur (relatif), l'autre affiche un squelette déconnecté

de l'utilisateur (absolu). Afin de distinguer ces techniques, nous avons introduit l'axe *Référentiel* dans le groupe *Comment*.

Résumé

Notre espace de conception considère deux aspects primordiaux du comportement d'une technique de guidage de geste : les mécanismes de retour d'information rétroactif et de retour d'information prédictif. Ces deux catégories sont composées de quatre groupes répondant à quatre questions : Quand (caractéristiques temporelles), Quoi (caractéristiques du contenu), Comment (caractéristiques de média) et Où (caractéristiques spatiales).

Nous montrons que les options existantes pertinentes pour caractériser le comportement d'une technique de guidage sont bien prises en compte dans notre espace de conception. De plus, les techniques elles-mêmes ont mené à la définition et à l'ajout de nouvelles options afin que notre espace de conception puisse capturer les spécificités de tous les guides existants. L'intégration de toutes ces options comprend un travail d'unification et d'organisation afin de créer un espace de conception cohérent.

8.2 Evaluation

Comme pour l'espace de conception concernant l'étape de désambigüisation du pointage, nous caractérisons celui dédié aux techniques de guidage de gestes selon trois dimensions [Beaudouin-Lafon, 2004] :

- *Pouvoir descriptif* : capacité à décrire un large éventail de techniques.
- *Pouvoir évaluatif* : capacité à définir quelles options de conception sont les meilleures.
- *Pouvoir génératif* : capacité à faciliter la création de nouvelles techniques.

Le fait que notre espace de conception possède 35 axes décrivant 46 techniques menant à 60 descriptions différentes d'étapes de guidage rend son utilisation difficile pour décrire, évaluer et générer des guides. En effet, classifier et décrire un grand nombre de techniques existantes parmi un grand nombre d'options de conception est une tâche relativement complexe. De plus, il peut être difficile de trouver des zones inexplorées dans un tel espace de conception (aspect génératif).
Afin de faciliter son utilisation, nous fournissons un outil interactif accessible en ligne : http://iihm.imag.fr/delamare/guidance/ (lien valide en septembre 2015). Nous décrivons les fonctionnalités offertes par notre outil tout en analysant les différents pouvoirs de notre espace de conception. Nous terminons cette analyse en présentant différents scénarios illustrant comment notre outil peut être utilisé par d'autres chercheurs ou concepteurs.

8.2.1 Outil interactif pour faciliter l'utilisation des pouvoirs de notre espace de conception

FIGURE 8.14 – Onglets de notre outil en ligne. Ces onglets mènent à différentes fonctionnalités proposées afin de faciliter l'usage de notre espace de conception : une présentation de ce dernier (Overview), une description des techniques de guidage existantes (One by One), une comparaison de ces techniques (Compare), une exploration interactive des options utilisées (Interactive Design Space) et la création de nouveaux guides (Create), une bibliographie ainsi qu'une page de contact.

Notre outil offre des fonctionnalités visant à aider les utilisateurs (i.e. concepteurs / chercheurs / ingénieurs) à trouver des techniques existantes ou bien à concevoir de nouvelles techniques de guidage. Ces fonctionnalités sont :

1. une description détaillée des techniques existantes une par une (Figure 8.14, onglet *One by One*),
2. une comparaison haut-niveau ou bien détaillée entre les techniques (Figure 8.14, onglet *Compare*),
3. une exploration interactive de la fréquence d'utilisation des options de conception par les techniques existantes (Figure 8.14, onglet *Interactive Design Space*) et,
4. une conception personnalisée de sa propre technique de guidage afin de le comparer à celles existantes (Figure 8.14, onglet *Create*).

8.2.1.1 Fonction : description détaillée d'une technique de guidage

Notre outil permet à l'utilisateur de voir la description détaillée d'une technique en particulier (http://iihm.imag.fr/delamare/guidance/desc.html (lien valide en septembre 2015)). Pour cela, l'outil affiche la liste de toutes les techniques référencées (Figure 8.15, B). Une fois que l'utilisateur a cliqué sur le nom de la technique désirée, l'outil affiche :

1. la référence de la technique concernée (Figure 8.15, C1),
2. le nombre d'étapes composant cette technique de guidage (Figure 8.15, C2),
3. une liste de liens permettant d'accéder de manière rapide à la description de ces différentes étapes (Figure 8.15, C2),
4. la description détaillée de ces différentes étapes (Figure 8.15, A). Pour chacune de ces étapes, l'outil fournit une description en colorant en vert directement dans l'espace de conception, les options utilisées. Les options affichant un astérisque informent l'utilisateur qu'il peut accéder à des informations supplémentaires quant à cette option grâce à une info-bulle lors du passage du curseur (Figure 8.15, D).

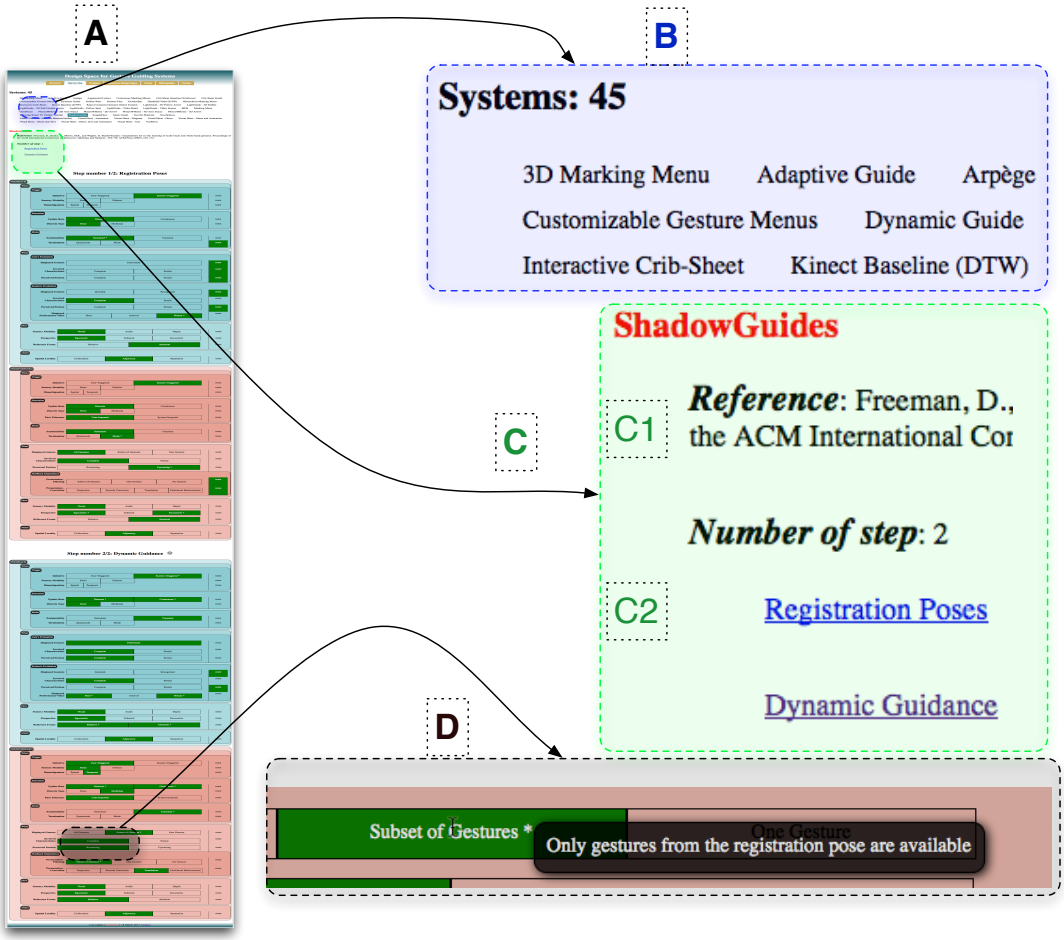


FIGURE 8.15 – Description de toutes les techniques de guidage une par une (A). Tout d'abord, l'utilisateur choisit une technique dans une liste de tous les guides inclus dans notre outil (B). L'outil affiche ensuite la référence de la technique choisie (C1) ainsi que des liens vers la description de chaque étape composant le guide (C2). Enfin, l'outil affiche pour chaque étape les options utilisées en les colorant en vert dans l'espace de conception (A et D). Certaines options indiquent via un astérisque la possibilité d'accéder à une info-bulle affichant les raisons menant au choix de cette option selon notre analyse (D).

La référence est utile car elle représente la source d'information de base ayant servie à classer la(s) technique(s) concernée(s). L'utilisateur peut ainsi avoir accès à la référence du guide afin d'obtenir de plus amples informations à propos de sa forme concrète, c'est-à-dire la représentation des gestes, des résultats des expériences utilisateurs, etc. Cette description des techniques peut s'avérer utile si l'utilisateur connait déjà une technique de guidage et aimerait se remémorer certains aspects de son comportement sans relire le(s) papier(s) de référence.

Cette fonctionnalité démontre le **pouvoir descriptif** de notre espace de conception. En effet, bien que la complétude d'un espace de conception ne puisse logiquement pas être prouvée une fois pour toute, notre espace caractérise toutes les techniques de guidage existantes tout en soulignant leurs différences.

8.2.1.2 Fonction : comparaison entre les techniques de guidage

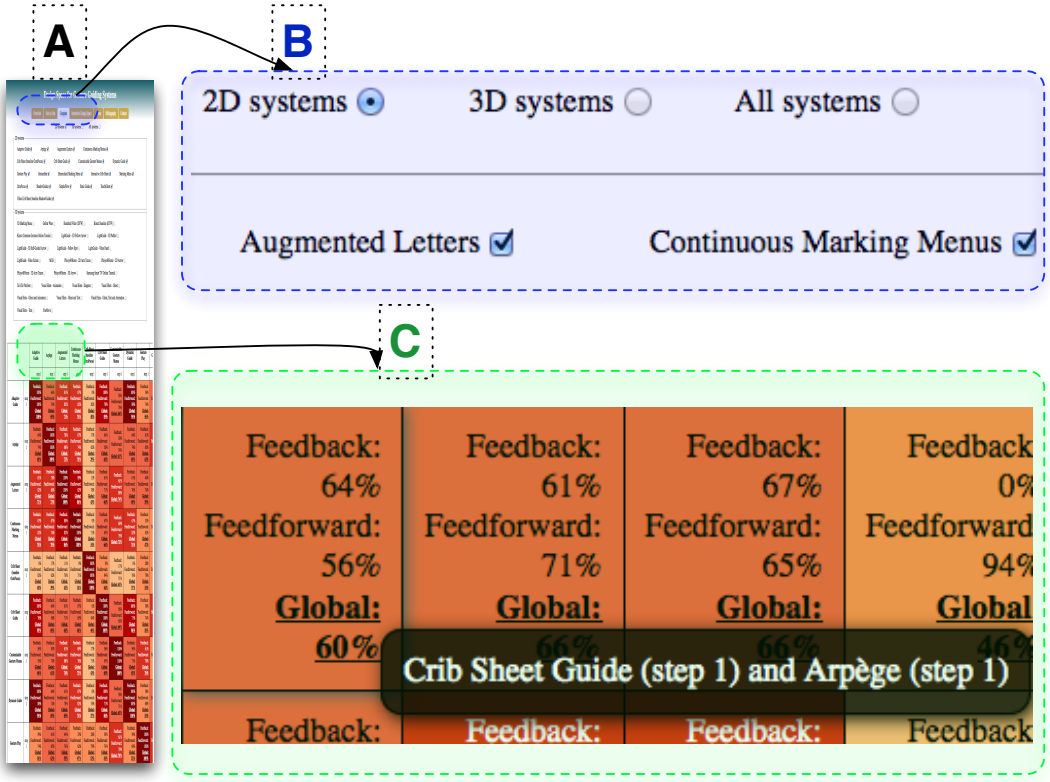


FIGURE 8.16 – Comparaison globale (A). Sélection des techniques à comparer (B). Tableau comparatif présentant dans chaque cellule trois scores de similarités - retour d'information rétroactif, retour d'information prédictif et global (C). Une info-bulle rappelle les techniques correspondant à la cellule survolée par le curseur (C).

L'outil fournit une fonction de comparaison des techniques entre elles (http://iihm.imag.fr/delamare/guidance/compare.html (lien valide en septembre 2015))(Figure 8.16). L'utilisateur peut rapidement choisir les techniques à comparer (2D, 3D ou toutes) tout en pouvant affiner son choix grâce à des cases à cocher (Figure 8.16, B). Une fois ce choix effectué, l'outil affiche un tableau présentant la similarité entre chaque étape des techniques sélectionnées (Figure 8.16, C). Cette similarité est présentée sous deux formes. Tout d'abord, un code couleur permet à l'utilisateur d'avoir un aperçu des similarités du premier coup d'œil. Nous proposons deux échelles de couleurs pour cette représentation visuelle de la similarité. Ensuite, chaque cellule affiche trois scores de similarité : un score

pour le mécanisme de retour d'information rétroactif, un score pour le mécanisme de retour d'information prédictif, et un score global. Ces trois scores, affichés en pourcentage, représentent le nombre d'options de conception partagées par les techniques comparées. En fonction du nombre de guides à comparer, le tableau peut être relativement grand (en hauteur et en largeur). Pour palier à ce problème lors de l'exploration du tableau, une info-bulle rappelle à l'utilisateur les techniques correspondant à la cellule survolée par le curseur (Figure 8.16, C).

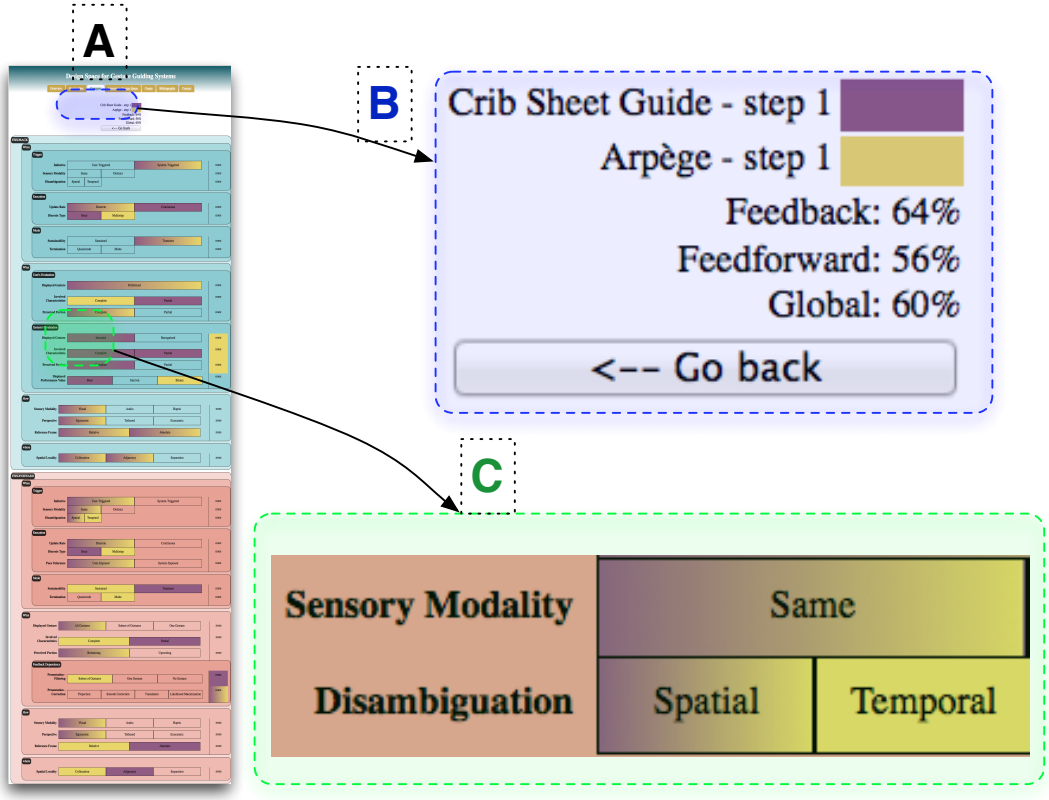


FIGURE 8.17 – Comparaison deux à deux (A). Résumé de la comparaison (B). Coloration de l'espace de conception en fonction des deux techniques impliquées (C). Les options partagées par les deux techniques sont colorées avec un dégradé entre les deux couleurs (C).

Pour obtenir plus de détails sur les scores de similarité, l'utilisateur peut cliquer sur une cellule afin de visuellement comparer les options de conception directement dans l'espace de conception (Figure 8.17, A). L'outil affiche tout d'abord le résumé de cette comparaison détaillée, c'est-à-dire les noms et étapes des guides impliqués, les trois scores de similarité ainsi que la légende utilisée pour colorer les options dans l'espace de conception (Figure 8.17, B). Une couleur est attribuée à chaque guide afin de visualiser en un coup d'oeil les similarités et les différences entre leurs options de conception. Si les deux techniques partagent une même option, alors celle-ci sera colorée avec un dégradé

entre les des deux couleurs.

Ces deux comparaisons permettent à l'utilisateur d'exploiter les **pouvoirs descriptif et évaluatif** de notre espace de conception. En effet, la comparaison à haut-niveau fournit un aperçu visuel des techniques existantes en aidant à rapidement identifier celles qui sont similaires entre elles ou au contraire celles qui sont uniques. Un examen plus détaillé de ces comparaisons permet à l'utilisateur d'identifier les différences clés entre deux guides. Ce dernier point est important lorsque l'on désire comprendre les raisons de différences de performances par exemple. En effet, pour contribuer au pouvoir évaluatif d'un espace de conception, il convient de comparer des techniques qui ne diffèrent que selon un axe de conception afin de capturer les effets des options de cet axe sur les performances.

8.2.1.3 Fonction : rechercher des zones inexplorées dans l'espace de conception

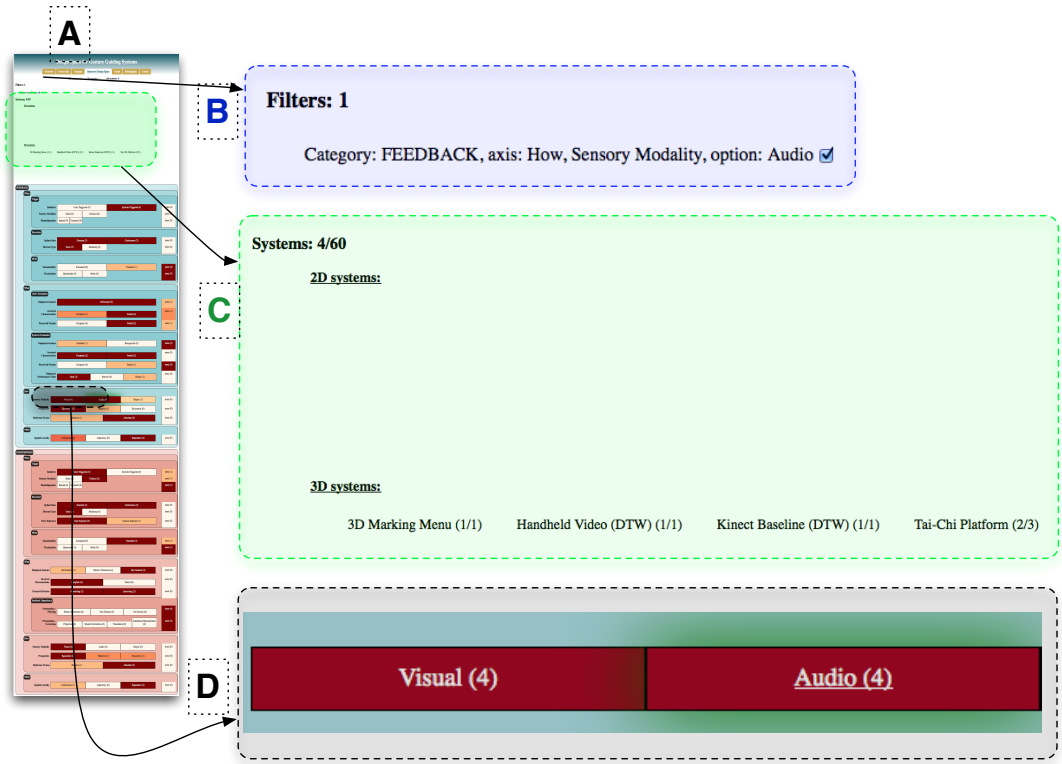


FIGURE 8.18 – Espace de conception interactif (A). Dans cet exemple, l'état courant contient un filtre (B). L'outil n'affiche donc que les quatre guides implémentant cette option (C). L'utilisateur peut alors visualiser combien de techniques implémentent les options composant l'espace de conception (D).

L'outil offre la possibilité d'explorer la fréquence d'utilisation des options de conception http://iihm.imag.fr/delamare/guidance/interactive-designspace.html (lien valide en sep-

tembre 2015) (Figure 8.18, A). Pour cela, l'espace de conception est affiché avec trois informations supplémentaires :

1. La couleur d'une option correspond au nombre de guides (ou l'une de ses étapes) implémentant cette option. Ainsi, plus une option est implémentée par un grand nombre de techniques, plus sa couleur sera proche du rouge.
2. Le nombre de techniques implémentant l'option est explicitement affiché à côté de son nom (Figure 8.18, D).
3. Une info-bulle liste les noms et les étapes des guides lorsque le curseur survole une option.

L'utilisateur peut cliquer sur une option afin de ne prendre en compte que les guides implémentant cette option. L'ensemble de la page se met à jour, c'est-à-dire la liste des filtres courants (Figure 8.18, B), les techniques restantes (Figure 8.18, C) ainsi que les couleurs des options de l'espace de conception. Les options servant de filtres sont mises en évidence grâce à un halo vert pour que l'utilisateur puisse s'en souvenir lorsqu'il navigue dans la page. A tout moment, une option peut être retirée en cliquant à nouveau dessus. Dans l'exemple de la Figure 8.18, l'utilisateur a cliqué sur l'option "Audio" de l'axe "Modalité" du groupe "Comment" du mécanisme de retour d'information rétroactif (Figure 8.18, B). Seulement quatre paires distinctes <technique / étape> implémentent cette option (Figure 8.18, C).

8.2.1.4 Fonction : concevoir une nouvelle technique de guidage

L'outil offre la possibilité de concevoir de nouvelles techniques de guidage http://iihm.imag.fr/delamare/guidance/create.html (lien valide en septembre 2015). Pour cela, l'utilisateur peut :

1. Entrer le nombre d'étapes envisagées. Après validation, l'outil affichera l'espace de conception autant de fois que nécessaire, c'est-à-dire une fois pour chaque étape.
2. Cliquer sur les options désirées. Une fois activée, l'option affiche une couleur verte. L'utilisateur peut à tout moment désactiver une option en cliquant dessus à nouveau.
3. Cliquer sur le bouton "Valider" pour ensuite comparer sa technique à celles existantes. Cette fonctionnalité est la même que celle présentée précédemment lorsque l'utilisateur désire comparer toutes les techniques existantes. Le nombre de lignes correspond au nombre d'étape(s) du guide nouvellement créé.

Cette fonctionnalité est utile dans deux scénarios.
Tout d'abord, elle peut être utilisée pour trouver un guide existant qui répondrait aux critères spécifiés par l'utilisateur. En effet, grâce aux scores de similarité, l'utilisateur peut trouver un guide similaire à celui qu'il avait en tête (score de similarité global de 100%) ou bien adapter celui qui aura le score le plus élevé.
Enfin, cette fonctionnalité augmente le **pouvoir évaluatif**. En effet, pour construire

le pouvoir évaluatif d'un espace de conception, il faut systématiquement effectuer des expériences en faisant varier les options le long d'un axe seulement. Or, il est difficile de réutiliser les résultats des expériences précédentes car elles diffèrent souvent selon plusieurs options d'un ou plusieurs axes. Par exemple, la Crib Sheet et le Static Guide [Anderson et Bischof, 2013] diffèrent selon deux axes, ce qui nous empêche de tirer des conclusions sur les différentes options impliquées. Désormais, les chercheurs peuvent utiliser notre outil afin de trouver des guides existants similaires à une nouvelle technique. En portant une attention particulière aux différences, ils peuvent choisir le guide auquel se comparer expérimentalement en minimisant le nombre d'axes impliqués dans ces différences, et ainsi contribuer au pouvoir évaluatif de l'espace de conception. Ils pourront alors conclure sur l'impact d'une option sur les performances de l'utilisateur.

8.2.1.5 Fonctions secondaires liées à la base de connaissance de l'outil

Cet outil sert aussi de support et de base de connaissances. Nous avons donc implémenté des fonctionnalités secondaires qui viennent s'ajouter aux fonctionnalités principales listées précédemment.

— Il y a notamment une page d'accueil sur laquelle l'utilisateur pourra trouver le *résumé* du papier de référence de notre espace de conception ainsi qu'une *représentation visuelle* des groupes, sous-groupes et axes qui le composent.
— Il y a aussi un onglet permettant d'accéder à la *bibliographie* complète des techniques de guidage référencées par l'outil. Cette bibliographie prend en compte les guides décrits dans des articles de recherche mais aussi ceux existants et exploités dans le milieu industriel.
— Il est important de pouvoir maintenir cette base de connaissance (actuellement composée de 46 techniques). Pour cela, une page permet de nous *contacter* afin de nous signaler de nouveaux guides à ajouter dans notre outil.

8.2.2 Scénarios d'utilisation

Nous illustrons comment notre outil peut être utile pour deux types d'utilisateurs cibles : des utilisateurs qui veulent trouver rapidement une solution à un problème (tels que les ingénieurs), et des utilisateurs qui veulent trouver des axes de recherche inexplorés - tels que des chercheurs en IHM.

8.2.2.1 Trouver une solution d'interaction

Trouver une technique existante.
Bill est un ingénieur en charge du développement d'une application destinée au contrôle d'appareils dans un habitat intelligent, tels que les lumières, la musique, les appareils

ménagers, etc. Bill a déjà une solution pour la phase de sélection d'un objet physique. Cette solution s'appuie sur un geste de pointage à distance suivi d'une étape de désambigüisation si plusieurs objets sont sélectionnés. Il cherche donc une solution pour contrôler ces objets sélectionnés (comme augmenter le volume ou l'intensité lumineuse par exemple) via des interactions gestuelles. Malheureusement, Bill n'a qu'une courte période de temps avant la démonstration du premier prototype. Il ne peut donc pas se permettre de passer trop de temps sur l'implémentation du système de reconnaissance de gestes, la définition du jeu de gestes ou la technique de guidage. Pourtant, Bill et son supérieur sont tous deux conscients qu'une technique de guidage est nécessaire afin de montrer aux utilisateurs les commandes disponibles ainsi que les gestes associés. Notre outil permet à Bill de trouver rapidement une technique de guidage satisfaisant tous ses critères, sans avoir à passer du temps à faire l'état de l'art complet : puisqu'il a des compétences en développement sur smartphone, il aimerait rapidement mettre en oeuvre un prototype (i) d'interaction gestuelle 3D en utilisant les capteurs du téléphone pour détecter les gestes en entrée. Un premier sondage auprès des utilisateurs finaux a montré que ces derniers préféreraient (ii) une modalité visuelle. Pour cette première itération, Bill veut se concentrer uniquement sur le retour d'information prédictif, en montrant (iii) tous les gestes (iv) une seule fois. En seulement quatre clics, il peut voir qu'un guide est conforme à ces critères : 3D Marking Menu [Ren et O'Neill, 2012]. Cette technique sert à guider des gestes de la main à effectuer en faisant des mouvements linéaires en l'air. Les gestes sont affichés sur un écran situé en face de l'utilisateur. Notre outil permet à Bill de rapidement trouver le papier de référence afin qu'il puisse bénéficier de l'étude déjà effectuée et obtenir des informations quant aux difficultés d'implémentation, des détails bas-niveau de la conception, des résultats des expériences, etc. Notre outil permet donc à Bill d'optimiser son temps sur cette partie du projet, sans avoir à concevoir et à développer un nouveau guide à partir de rien. Pour cette première itération, Bill décide donc de mettre en oeuvre une version de 3D Marking Menu [Ren et O'Neill, 2012].

S'inspirer et adapter des techniques existantes

Après la première itération, une extension des spécifications externes empêche Bill d'utiliser un écran additionnel pour présenter les gestes comme proposé dans la version originale de 3D Marking Menu [Ren et O'Neill, 2012]. En effet, les utilisateurs doivent avoir accès au guide dans toutes les pièces de l'habitat intelligent, mais il n'est pas envisageable d'intégrer un écran dans chacune de ces pièces. Bill choisit donc d'afficher les gestes sur l'écran du téléphone qu'il utilise déjà pour capturer les gestes en entrée. Il conclut néanmoins sur la nécessité d'ajouter un retour tactile à cause de la place limitée disponible pour l'affichage. Notre outil révèle rapidement que seulement trois techniques utilisent un mécanisme de retour d'information rétroactif avec une modalité tactile [Portillo-Rodriguez *et al.*, 2008] [Schönauer *et al.*, 2012] [Schwaller *et al.*, 2014]. Malheureusement, après avoir cliqué sur l'option "tactile" du retour d'information rétroactif, il s'aperçoit que ces techniques sont soit dédiées au guidage de gestes un par un soit utilisent la modalité tactile pour

uniquement signaler la reconnaissance ou non du geste effectué. Bill doit donc trouver un moyen d'adapter ces mécanismes tactiles dans le cas d'un guidage de plusieurs gestes en même temps. Bill supprime le filtre "tactile" du retour d'information rétroactif en re-cliquant dessus et peut voir que SimpleFlow [Bennett *et al.*, 2011] permet de filtrer les gestes présentés en n'affichant qu'un seul geste. Il utilise donc le mécanisme de prédiction de SimpleFlow afin de déterminer le geste en cours d'exécution. Il peut ensuite adapter le motif vibrotactile de [Schönauer *et al.*, 2012] afin de rendre perceptible la distance entre la position courante du smartphone et la limite de la zone 3D de la commande associée : plus le téléphone se rapproche de cette limite, plus la vibration est forte. L'utilisateur a donc un retour tactile quant aux zones spatiales situées en face de lui. Enfin, Bill ajoute une hystérésis pour éviter des vibrations lorsque le téléphone franchit une limite.

8.2.2.2 Trouver un problème de recherche

Rapide aperçu des techniques existantes
John, doctorant en IHM, aimerait avoir une vue d'ensemble des techniques de guidage existantes. Alors que certains domaines, tels que le pointage, bénéficient d'articles de recherche faisant l'état de l'art [Argelaguet et Andujar, 2013], aucun n'existe en ce qui concerne le guidage de gestes. De plus, ces études peuvent devenir obsolètes après un certain temps. Grâce à notre outil, John peut avoir une liste continuellement mise à jour des techniques de guidage. Après avoir lu toutes ces références, John peut avoir oublié certains détails ou bien confondre des options de conception entre deux techniques. Notre outil lui permet de se rappeler ces détails via les descriptions de chaque guide référencé. John aimerait aussi rapidement comparer certaines techniques, ne serait-ce que pour savoir lesquelles sont uniques et lesquelles ont inspiré d'autres travaux. Par la suite, il s'aperçoit que deux guides tirés d'une même étude sont très similaires, Adaptive Guide et Dynamic Guide [Anderson et Bischof, 2013], avec un score global de 94%. Il clique sur la case correspondante du tableau comparatif pour visualiser les deux chemins colorés dans l'espace de conception. D'un coup d'oeil, il voit que (1) le filtrage d'Adaptive Guide est complet (il supprime tous les gestes affichés) alors que Dynamic Guide affiche une sous-partie des gestes et (2), Dynamic Guide fournit une correction en re-localisant les gestes affichés pendant l'exécution d'un geste.

Création d'une nouvelle technique de guidage
Jane, chercheuse en IHM, aimerait aborder le problème de guidage de gestes. En utilisant notre outil, elle peut facilement trouver des zones inexplorées de l'espace de conception. Elle espère pouvoir trouver de telles zones qui définissent des axes de recherche intéressants. Notre outil lui révèle que la majorité des guides déclenchent automatiquement le mécanisme de retour d'information rétroactif. Or, il s'avère que si l'utilisateur a l'information pendant le guidage, il aura tendance à trop s'appuyer dessus et cela aura des répercussions sur l'apprentissage moteur des gestes [Park *et al.*, 2000]. Jane pense qu'il y a des possibilités

d'amélioration sur ce point. Elle fait la supposition que si l'utilisateur doit effectuer une action supplémentaire pour activer le mécanisme de retour d'information rétroactif, alors il est possible que l'utilisateur évite d'effectuer ce déclenchement lorsqu'il estime que cela n'est pas vraiment nécessaire. Jane désire étudier cette option de conception afin de vérifier son effet sur l'apprentissage moteur de gestes.
Jane peut également filtrer les techniques en cliquant sur les options de l'espace de conception. Par exemple, si elle veut étudier (i) les guides pour des gestes 3D qui (ii) affichent de manière continue (iii) la partie restante (iv) de tous les gestes, elle remarquera qu'aucune technique n'existe pour le moment. Avec seulement quatre clics, elle a pu trouver une étude de cas qui n'avait pas encore été proposée dans la littérature. Un peu étonnée et curieuse, elle retire le filtre "3D" afin de vérifier ce qui existe en 2D. Elle s'aperçoit alors qu'une technique de guidage pour des gestes 2D satisfait ces critères : OctoPocus [Bau et Mackay, 2008]. Finalement, elle identifie des défis intéressants dans l'adaptation de ce guide pour passer des gestes 2D aux gestes 3D.

Résumé

Notre espace de conception possède un large choix d'options et classe un grand nombre de techniques existantes. Afin de faciliter son utilisation, nous fournissons un outil accessible en ligne proposant quatre fonctionnalités : (1) une description détaillée des techniques existantes, (2) une comparaison haut-niveau entre les techniques affichant trois scores de similarité (retour d'information rétroactif, retour d'information prédictif et global) en fonction du nombre d'options de conception qu'ont en commun les techniques, mais aussi une comparaison visuelle détaillée de techniques deux à deux, (3) une exploration interactive de la fréquence d'utilisation des options de conception par les guides existants et (4), une conception interactive de nouvelles techniques de guidage et la comparaison à celles existantes.

Grâce à ces fonctionnalités, nous illustrons via différents scénarios comment l'outil facilite l'utilisation de notre espace de conception afin que l'utilisateur puisse bénéficier de ses pouvoirs descriptif, évaluatif et génératif.

La base de connaissance utilisée pour la création de notre espace de conception et référencée par notre outil est aussi une contribution de ces travaux.

8.3 Conclusion

Afin de contribuer aux travaux dédiés au guidage de gestes, nous fournissons un espace de conception visant à capturer les spécificités du comportement des techniques de guidage. Pour cela, nous distinguons les mécanisme de *retour d'information rétroactif* et de *retour*

d'information prédictif qui fournissent respectivement de l'information quant aux actions effectuées et aux actions possibles dans le futur. Ces deux catégories sont elle-mêmes décrites en quatre groupes en lien avec quatre questions : *Quand*, *Quoi*, *Comment* et *Où*. Cet espace de conception propose un total de 35 axes, avec 115 options de conception.
Ce large choix d'options combiné au grand nombre de techniques de guidage proposées dans la littérature (46 guides) rend l'utilisation de l'espace de conception difficile. Afin de faciliter son utilisation, notamment afin de décrire les techniques existantes, de les comparer, d'évaluer les avantages et inconvénients des options de conception, ou encore de prospecter des zones inexplorées de l'espace de conception, nous proposons un outil en ligne offrant différentes fonctionnalités. Nous montrons via des scénarios comment cet outil permet à différents types d'utilisateurs d'exploiter au mieux notre espace de conception afin de trouver des solutions ou des problèmes de recherche.
Dans la partie suivante, nous utilisons nous-même l'outil en ligne et notre espace de conception afin de créer une technique de guidage pour des gestes 3D dans un contexte d'habitat intelligent et contrôler les objets physiques augmentés le composant.

9 Système de Guidage : OctoPocus3D

Sommaire

9.1 Introduction 122

9.2 Choix de conception 124

9.2.1 Recherche d'une solution grâce à notre espace de conception 124

9.2.2 Description d'OctoPocus 124

9.2.2.1 Retour d'information rétroactif d'OctoPocus et d'OctoPocus3D 125

9.2.2.1.1 Groupe *Quand* 125

9.2.2.1.2 Groupe *Quoi* 125

9.2.2.1.3 Groupe *Comment* 126

9.2.2.1.4 Groupe *Où* 126

9.2.2.2 Retour d'information prédictif 127

9.2.2.2.1 Groupe *Quand* 127

9.2.2.2.2 Groupe *Quoi* 127

9.2.2.2.3 Groupe *Comment* 128

9.2.2.2.4 Groupe *Où* 128

9.2.2.3 Vue d'ensemble 129

9.2.3 Alternatives de conception : exemples 129

9.2.3.1 Fréquence de mise à jour du retour d'information rétroactif 130

9.2.3.2 Position du guide définie par le retour d'information prédictif 132

9.2.3.3 Portion de gestes affichée par le retour d'information prédictif 134

9.3 Conception d'OctoPocus3D 135

9.3.1 Affichage des gestes 3D 135

9.3.1.1 Solutions retenues pour la version finale d'OctoPocus3D . 140

9.3.1.1.1 Occultation et transparence 140

9.3.1.1.2 Perspective linéaire 141

9.3.1.1.3 Éclairage et nuances d'ombres 141

9.3.1.1.4 Profondeur cinétique et couplage tête - affichage 141

9.3.1.2 Solutions non retenues pour la version finale d'OctoPocus3D 142

9.3.1.2.1 Ajouts d'éléments visuels 142
9.3.1.2.2 Exploitation des propriétés graphiques des tubes 143
9.3.2 Retour d'information rétroactif : terminal et/ou continu 144
9.3.3 Retour d'information prédictif : origine du guide et portions des gestes affichées 145
9.4 Évaluations expérimentales 148
9.4.1 Ressources 149
9.4.1.1 Conception des jeux de gestes pour les évaluations 149
9.4.1.2 Équipement et configuration 149
9.4.1.3 Etude préliminaire : OctoPocus3D *vs* vidéos de démonstration 150
9.4.2 Expérience 1 : retour d'information rétroactif 151
9.4.2.1 Participants et plan d'expérience 151
9.4.2.2 Résultats 153
9.4.2.2.1 Taux de reconnaissance 153
9.4.2.2.2 Temps d'exécution 155
9.4.2.2.3 Discussion 155
9.4.3 Expérience 2 : retour d'information prédictif 156
9.4.3.1 Participants et plan d'expérience 156
9.4.3.2 Résultats 158
9.4.3.2.1 Taux de reconnaissance 158
9.4.3.2.2 Temps d'exécution 158
9.4.3.2.3 Discussion 159
9.5 Déploiement *in situ* 160
9.6 Conclusion 161
9.7 Travaux futurs 162

9.1 Introduction

Nous cherchons à exploiter le pouvoir génératif de notre espace de conception afin de créer une technique de guidage de gestes 3D pour contrôler des objets physiques augmentés.

Nous présentons tout d'abord comment notre espace de conception nous a menés à l'exploration de plusieurs extensions d'OctoPocus [Bau et Mackay, 2008], un guide de gestes 2D, adapté à des gestes 3D : OctoPocus3D. Nous faisons ensuite un rappel des différentes options de conception existantes (paragraphe 9.2.3) avant de détailler la conception d'OctoPocus3D (paragraphe 9.3). Nous présentons ensuite deux évaluations expérimentales d'OctoPocus3D (paragraphe 9.4) avant de conclure (paragraphe 9.6) et d'annoncer des pistes de travaux futurs (paragraphe 9.7).

Pour construire notre espace de conception, nous nous sommes inspirés de travaux traitant de gestes différents les uns des autres : des gestes 2D ou 3D, des mouvements de danse [Anderson *et al.*, 2013], de ré-éducation musculaire [Sodhi *et al.*, 2012], ou bien pour

exécuter des commandes et ainsi interagir avec un système [Rovelo *et al.*,]. Cette diversité de gestes est à l'origine d'une grande variété de techniques de guidage, que nous avons décrites au sein de notre espace de conception. Ce faisant, nous avons pu capturer les spécificités du comportement du plus grand nombre de guides possible. Toutefois, dans le cadre de notre projet DELight (paragraphe 2.1.1), notre objectif est d'interagir à distance avec des objets physiques dans un contexte d'intelligence ambiante (paragraphe 2.1.2), donc avec des gestes *symboliques dynamiques* (paragraphe 2.2). Ces gestes permettent en effet de définir un vocabulaire d'entrée permettant de relier un geste à une commande. Ce vocabulaire nécessite cependant que l'utilisateur puisse connaître (1) les commandes disponibles et (2) les gestes associés, motivant donc nos travaux sur le guidage de gestes 3D (chapitre 7).

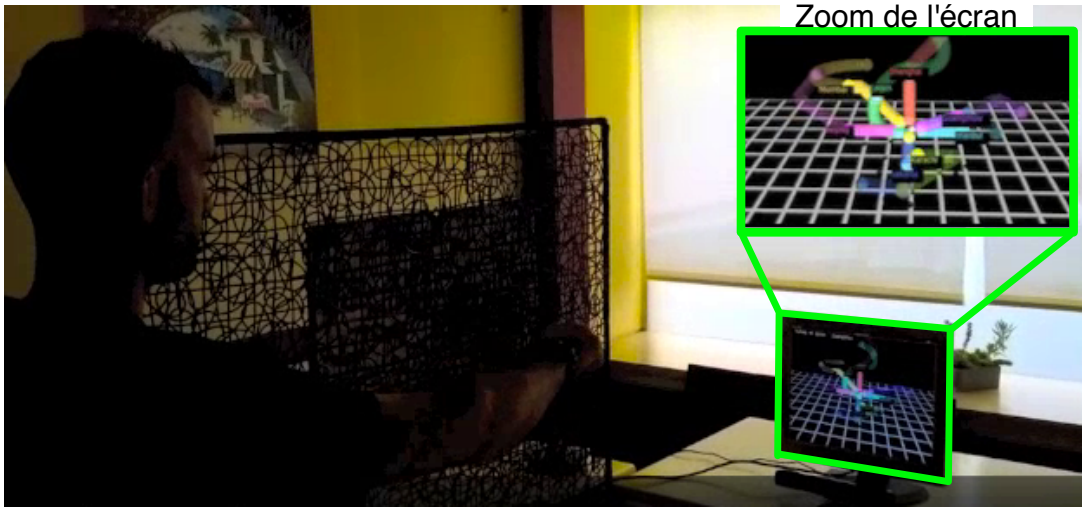


FIGURE 9.1 – Utilisateur effectuant un geste avec notre prototype à l'entrée d'un restaurant. Le guide affiche les gestes à effectuer sous forme de tubes 3D sur un écran distant placé en face de l'utilisateur. Coin haut-droit : zoom de l'écran.

Dans ce chapitre, nous présentons un guide de gestes 3D s'affichant sur un écran distant (Figure 9.1). En effet, contrairement au guidage d'un ensemble de gestes 2D [Bau et Mackay, 2008] ou au guidage d'un unique geste 3D [Sodhi *et al.*, 2012], l'affichage de plusieurs gestes 3D peut difficilement être co-localisé avec la main de l'utilisateur. Il faudrait pour cela instrumenter ce dernier afin qu'il puisse visualiser les gestes. Nous étudions une solution complémentaire qui consiste à utiliser un écran distant afin de visuellement représenter le jeu de gestes 3D. Le guide révèle à l'utilisateur les commandes qui lui sont mises à disposition sous forme d'étiquettes 2D. Les gestes associés à ces commandes sont représentés par des tubes 3D reliés aux commandes grâce à un code couleur et à leur position sur l'écran.

9.2 Choix de conception

Nous détaillons (1) comment notre espace de conception a révélé qu'OctoPocus était un bon candidat pour être adapté aux gestes symboliques dynamiques 3D (paragraphe 9.2.1), (2) les options de conception d'OctoPocus en mettant l'accent sur certaines options intéressantes à explorer et (3) les différentes options existantes pour ces options de conception.

9.2.1 Recherche d'une solution grâce à notre espace de conception

Nous reprenons en partie le scénario de Jane (paragraphe 8.2.2.2) : nous aimerions créer un guidage présentant (i) tous les gestes (ii) de manière continue afin d'aider l'utilisateur pendant l'exécution d'un geste. En deux clics, l'outil révèle que seulement deux techniques ont le comportement que nous recherchons :

1. un guide de mouvements de Tai Chi [Portillo-Rodriguez *et al.*, 2008], plus précisément sa première étape du guidage qui consiste à montrer plusieurs avatars exécutants chacun un mouvement spécifique que l'utilisateur peut imiter afin d'accéder à un guidage plus détaillé, et
2. un guide pour des gestes 2D, à savoir OctoPocus [Bau et Mackay, 2008], qui présente les gestes grâce à des lignes 2D affichées sur un écran.

Le guide OctoPocus semble être un meilleur candidat que le guide de Tai Chi pour une adaptation à notre cas d'étude. En effet, ce dernier n'a le comportement recherché que dans sa première étape de guidage : la révélation des gestes. Au contraire, OctoPocus propose la révélation des gestes ainsi que le guidage en une seule étape.
Il est toutefois intéressant de noter des différences entre ces deux guides. Par exemple, OctoPocus affiche la partie restante complète des gestes à effectuer alors que le guide de Tai Chi n'affiche que la partie imminente à venir via des animations. Une autre différence marquante est que l'un (OctoPocus) propose une correction du retour d'information prédictif en re-localisant l'affichage du guide alors que l'autre (guide de Tai Chi) ne propose aucune correction. Dans la partie suivante, nous discutons plus en détail des options de conception d'OctoPocus.

9.2.2 Description d'OctoPocus

OctoPocus [Bau et Mackay, 2008] guide les utilisateurs pendant l'exécution de gestes 2D. Le guide affiche les gestes par des traits reliés à des noms de commandes via leur couleur et leur position sur l'écran.
Nous décrivons OctoPocus à partir des axes de notre espace de conception. Nous mettons en avant les options de conception que nous jugeons intéressantes à explorer en tenant compte du fait que nous cherchons à guider des gestes 3D. Ces options et leurs alternatives seront discutées dans la section suivante (paragraphe 9.2.3).

9.2.2.1 Retour d'information rétroactif d'OctoPocus et d'OctoPocus3D

9.2.2.1.1 Groupe *Quand* (Figure 9.2)

L'initiative pour le déclenchement du retour d'information rétroactif est laissé au système. La fréquence de mise à jour du retour d'information rétroactif est à la fois discrète - en présentant le nom du geste reconnu à la fin de l'exécution du geste - et continue. En effet, le guide transmet les scores de reconnaissance de chaque geste continument via l'épaisseur des traits : moins le geste est susceptible d'être reconnu par le système, plus le trait correspondant est fin.
Enfin, le retour d'information rétroactif est transitoire : il disparait automatiquement à la fin de l'exécution d'un geste.

Le retour d'information rétroactif continu est une caractéristique clé d'OctoPocus. Il est supposé aider les utilisateurs à se corriger pendant l'exécution d'un geste. Nous voulons étudier cet aspect du guide lors de la conception d'OctoPocus3D, technique de guidage pour des gestes 3D effectué avec la main dans le vide. Nous discutons de cette option de conception et de ces alternatives dans la section suivante (paragraphe 9.2.3).

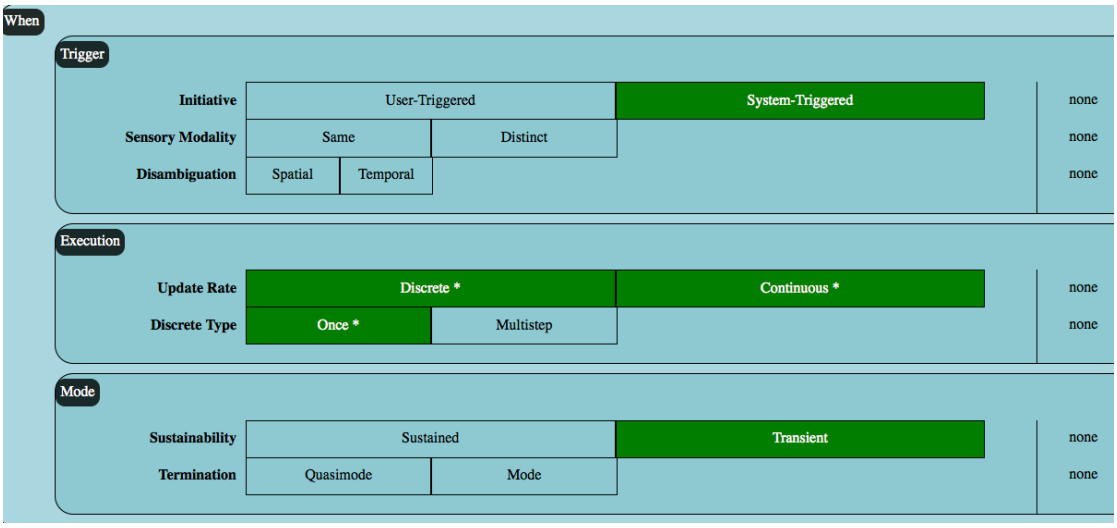


FIGURE 9.2 – Groupe d'axes *Quand* du retour d'information rétroactif d'OctoPocus.

9.2.2.1.2 Groupe *Quoi* (Figure 9.3)

OctoPocus affiche le geste effectué depuis le début du guidage grâce à un trait noir. Etant donné que le guide ne considère que la forme des gestes, cette représentation du geste effectué inclut la totalité des caractéristiques impliquées.
En revanche, le guide n'affiche ni le geste reconnu, ni le geste que l'utilisateur voulait effectuer.
OctoPocus fournit des valeurs de performances : une valeur binaire (le nom de la commande reconnue) et une valeur réelle (via l'épaisseur des traits).

Le fait que le guide n'affiche aucun geste excepté celui effectué est lié au fait qu'OctoPocus fourni un retour d'information rétroactif continu. La stratégie est de laisser l'utilisateur se corriger *pendant* l'exécution. Nous discutons des alternatives de conception dans la section suivante (paragraphe 9.2.3).

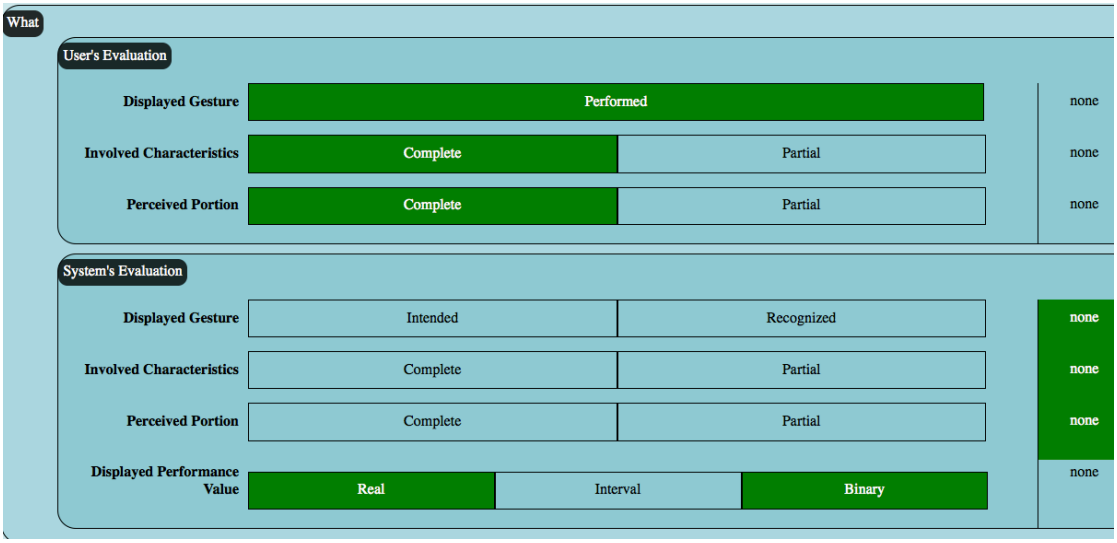


FIGURE 9.3 – Groupe d'axes *Quoi* du retour d'information rétroactif d'OctoPocus.

9.2.2.1.3 Groupe *Comment* (Figure 9.4)

Le retour d'information rétroactif d'OctoPocus utilise une modalité visuelle en affichant des informations avec une perspective égocentrique dans un référentiel relatif (pour les gestes et l'épaisseur des traits) et absolu (pour le nom de la commande reconnue).

OctoPocus3D utilise les mêmes options qu'OctoPocus afin d'afficher les informations du retour d'information rétroactif.

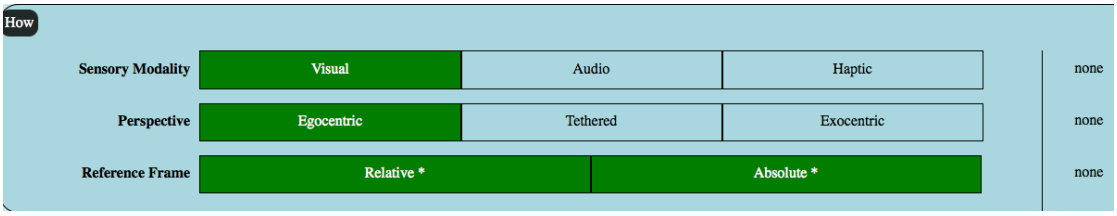


FIGURE 9.4 – Groupe d'axes *Comment* du retour d'information rétroactif d'OctoPocus.

9.2.2.1.4 Groupe *Où* (Figure 9.5)

Le retour d'information rétroactif d'OctoPocus affiche des informations co-localisées (pour les traits et leur épaisseur) et adjacente à l'emplacement pour effectuer les gestes (pour le nom de la commande reconnue affichée en haut de l'écran).

Au contraire, OctoPocus3D utilisant un écran distant pour afficher le guide, la localisation du guide est *séparée* de l'emplacement prévu pour effectuer les gestes.

FIGURE 9.5 – Groupe d'axes *Où* du retour d'information rétroactif d'OctoPocus.

9.2.2.2 Retour d'information prédictif

9.2.2.2.1 Groupe *Quand* (Figure 9.6)

L'initiative pour le déclenchement du retour d'information prédictif est laissé à l'utilisateur. Celui-ci peut faire apparaitre le guide en utilisant la modalité principale et en effectuant un temps de pause (critère temporel).
Le retour d'information prédictif est continument mis à jour et laisse l'utilisateur libre de choisir la vitesse à laquelle il souhaite exécuter les gestes.
Enfin, le mécanisme de retour d'information prédictif est transitoire car il disparait dès la fin d'exécution d'un geste.

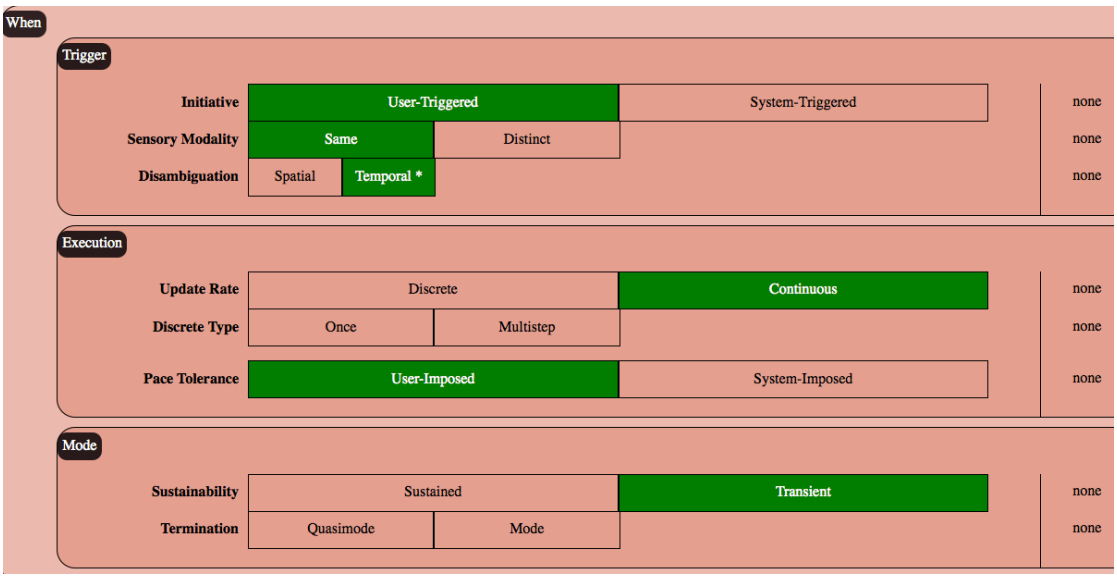


FIGURE 9.6 – Groupe d'axes *Où* du retour d'information prédictif d'OctoPocus.

9.2.2.2.2 Groupe *Quoi* (Figure 9.7)

OctoPocus affiche tous les gestes en considérant toutes les caractéristiques impliquées dans le guidage, à savoir la forme géométrique des gestes. Les gestes sont totalement affichés, c'est-à-dire que les traits représentent la portion complète à venir des gestes.
Le retour d'information rétroactif a deux impacts sur le retour d'information prédictif :

1. Le retour d'information prédictif n'affiche pas les gestes ayant un score de reconnaissance en-dessous d'une valeur limite prédéfinie (filtrage de la présentation).
2. Le retour d'information prédictif affiche les gestes sous le curseur. Il y a donc une re-localisation de l'affichage sous la position courante du curseur.

Pour OctoPocus3D, l'un des défis majeurs est l'affichage et la compréhension d'une scène 3D représentant de nombreux gestes par des tubes. Ce point est discuté dans le paragraphe 9.3.1. Les options de conception du groupe d'axes *Quoi* du retour d'information prédictif ont deux impacts sur la représentation graphique d'une scène 3D :

1. la *complexité visuelle* de la représentation visuelle via l'axe *Portion perçue* des gestes.
   Il y a un compromis entre la possibilité d'appréhender la globalité du geste et une potentielle surcharge d'information augmentant la complexité visuelle de la scène dans laquelle sont représentés les gestes.
2. la *stabilité visuelle* de la représentation visuelle via l'axe *Présentation-correction*. La re-localisation de l'affichage du retour d'information prédictif induit par le retour d'information rétroactif peut en effet apporter des problèmes de compréhension d'une scène 3D déjà encombrée.

Ces options de conception ainsi que leurs alternatives sont discutées dans le paragraphe 9.2.3.

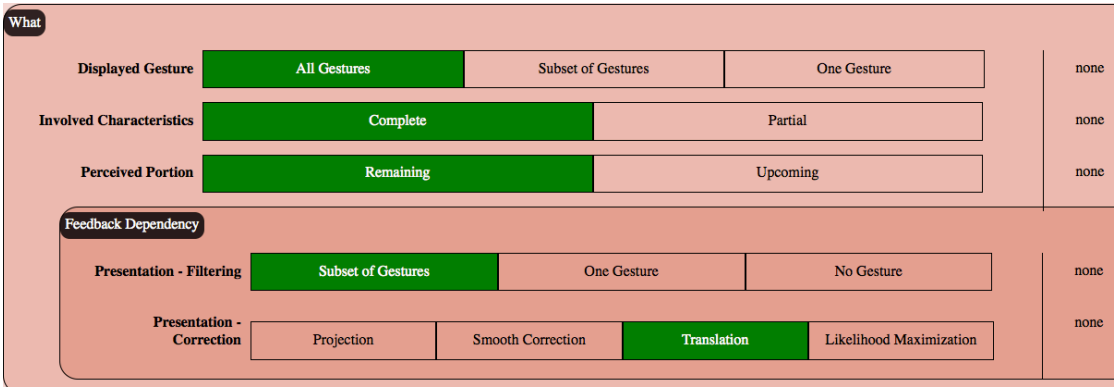


FIGURE 9.7 – Groupe d'axes *Quoi* du retour d'information prédictif d'OctoPocus.

9.2.2.2.3 Groupe *Comment* (Figure 9.8)

Le retour d'information prédictif d'OctoPocus utilise une modalité visuelle en affichant des informations via une perspective égocentrique dans un référentiel relatif (les traits et les étiquettes affichant le noms des commandes).

9.2.2.2.4 Groupe *Où* (Figure 9.9)

Le retour d'information prédictif d'OctoPocus affiche des informations (les traits et

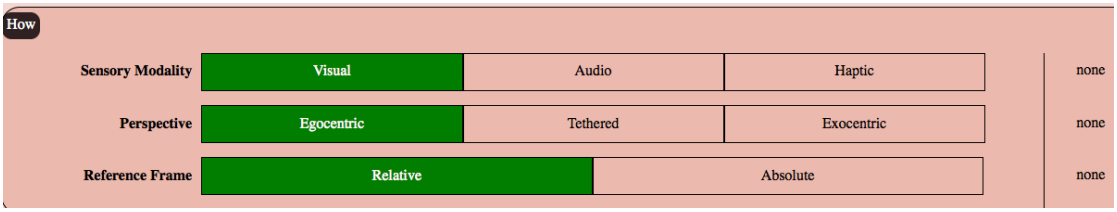


FIGURE 9.8 – Groupe d'axes *Comment* du retour d'information prédictif d'OctoPocus.

les étiquettes affichant le noms des commandes) co-localisées avec l'emplacement pour effectuer les gestes.

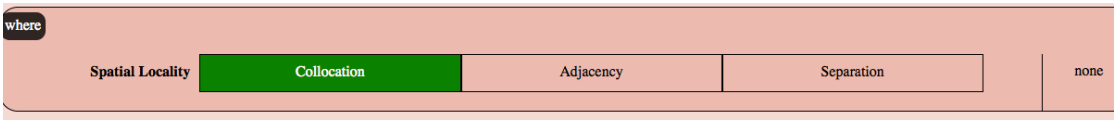


FIGURE 9.9 – Groupe d'axes *Où* du retour d'information prédictif d'OctoPocus.

9.2.2.3 Vue d'ensemble

Nous avons identifié des options de conception clés d'OctoPocus pour lesquelles nous voulons explorer des alternatives lors de la création d'OctoPocus3D (Figure 9.10). OctoPocus3D représente les gestes 3D par des tubes 3D affichés sur un écran distant de l'utilisateur. Nous considérons cette représentation visuelle comme une modalité de sortie composée d'un couple dispositif/langage $< d, L >$ [Vernier et Nigay, 2001]. Le dispositif de sortie d est l'écran distant. Le langage d'interaction L est la représentation visuelle des gestes, c'est-à-dire les tubes graphiques 3D. Les options de conception que l'on cherche à évaluer ont toutes un impact sur le langage d'interaction de la modalité visuelle utilisée par OctoPocus3D.

Dans ce chapitre, après une revue de l'état de l'art quant aux aspects des techniques de guidage liés à notre étude (paragraphe 9.2.3), nous décrivons les différentes alternatives de conception d'OctoPocus3D utilisant des tubes 3D comme représentation des gestes (paragraphe 9.3). Nous rapportons dans la suite les résultats de deux expériences contrôlées en laboratoire explorant l'impact de ces options de conception sur le temps d'exécution des gestes ainsi que sur le taux de reconnaissance obtenus en utilisant les différentes versions d'OctoPocus3D (paragraphe 9.4).

9.2.3 Alternatives de conception : exemples

Nous passons en revue les options de conception du retour d'information rétroactif et du retour d'information prédictif précédemment citées qui impactent notre modalité de sortie visuelle. Nous illustrons ces options avec un sous-ensemble de guides existants, que

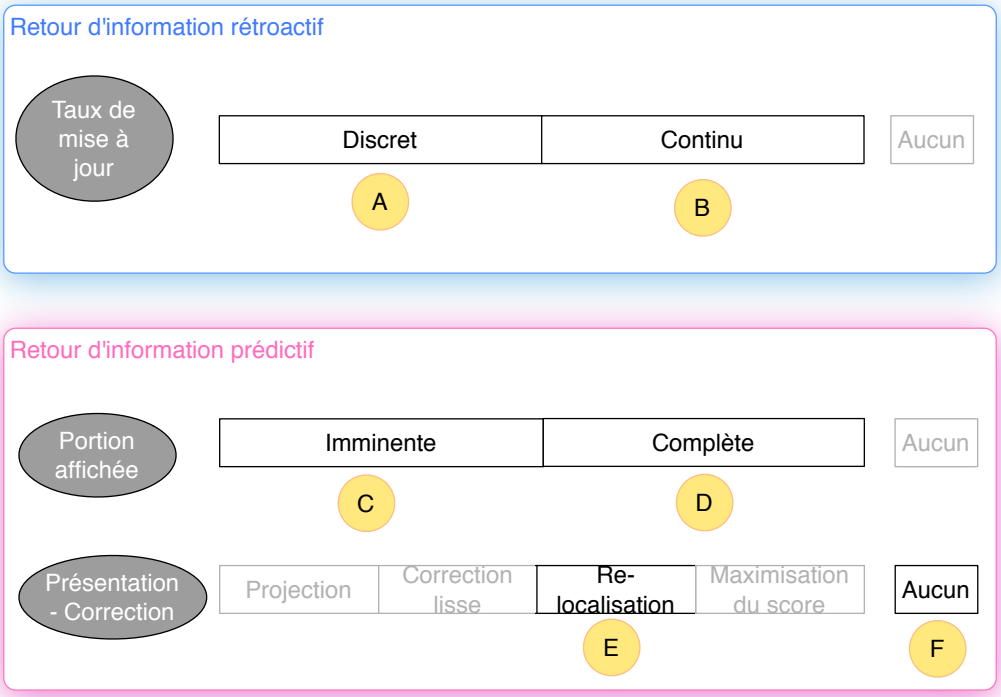


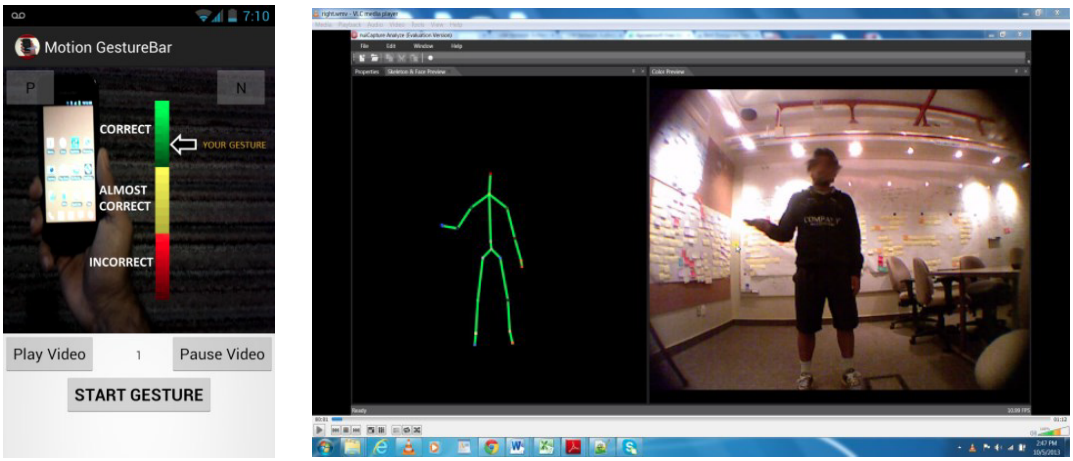
FIGURE 9.10 – Résumé des options de conception explorées lors de l'adaptation d'OctoPocus aux gestes 3D. Pour le retour d'information rétroactif, comparaison entre une stratégie d'apprentissage consistant à utiliser une mise à jour discrète (A) ou continue (B). Pour le retour d'information prédictif, exploration d'options impactant la complexité visuelle de la scène 3D selon la portion de geste affichée (C et D) ou la stabilité de la représentation visuelle selon la correction de la présentation (E et F).

ce soit pour des gestes 2D ou bien 3D.

9.2.3.1 Fréquence de mise à jour du retour d'information rétroactif

Le retour d'information rétroactif peut avoir une fréquence de mise à jour continue ou bien discrète. Un retour d'information rétroactif continu est appelé *concurrent*. Dans le cas d'une mise à jour discrète, le retour d'information rétroactif apparait généralement à la fin de l'exécution d'un geste. Il est donc appelé *terminal*

Un retour d'information rétroactif terminal peut être visuellement présenté de différentes manières. GestureBar [Bragdon *et al.*, 2009] fournit un tel retour d'information rétroactif après une phase d'entrainement dans une zone dédiée en affichant le message textuel "Bien joué" ou "Incorrect" à l'utilisateur. Un autre exemple [Kamal *et al.*, 2014] propose de représenter une flèche indiquant le score obtenu sur une échelle graphique avec trois



(a) (b)

FIGURE 9.11 – Illustrations de différents retours d'information rétroactifs terminaux. (a) Flèche sur une échelle de couleur. (b) Rediffusion des mouvements de l'enseignant et de l'utilisateur. Les deux techniques viennent de [Kamal *et al.*, 2014].

niveaux de couleur (Figure 9.11a). Enfin, d'autres techniques affichent simplement le nom de la commande reconnue [Ren et O'Neill, 2012].
Une autre solution pour fournir un retour d'information rétroactif terminal consiste à fournir un moyen de comparer le geste effectué avec le geste prévu. Il est important de noter que cette solution implique que le système connaisse le geste que l'utilisateur prévoit d'effectuer, comme lors d'une phase d'entrainement par exemple. Pour cela, certaines techniques proposent de représenter visuellement deux avatars : un pour l'utilisateur et un jouant le rôle de l'enseignant [Kamal *et al.*, 2014]. Avec cette solution, le retour d'information rétroactif terminal consiste à rediffuser simultanément le geste exécuté ainsi que le geste montré par l'avatar de l'enseignant (Figure 9.11b). Une autre technique propose cette solution pour des mouvements de Tai Chi [Portillo-Rodriguez *et al.*, 2008], mais enrichie grâce à des informations statistiques complémentaires.

La modalité visuelle de la technique de guidage pour les mouvements de Tai Chi [Portillo-Rodriguez *et al.*, 2008] fournit également des informations de manière continue, c'est-à-dire un retour d'information rétroactif concurrent. En effet, le guide présente de façon continue l'erreur mesurée entre le mouvement à effectuer et celui exécuté par l'utilisateur sous la forme d'une intensité lumineuse d'une ligne qui est affichée entre les mains de l'avatar de l'utilisateur. En utilisant aussi des avatars, la solution de [Schönauer *et al.*, 2012] utilise des flèches afin d'indiquer une déviation entre l'avatar de l'utilisateur et celui de l'enseignant.

D'autres techniques de guidage fournissent un retour d'information rétroactif concurrent en se basant sur une modalité visuelle représentant les gestes eux-mêmes. Ces guides

sont souvent qualifiés de "dynamiques" [Bau et Mackay, 2008]. Par exemple, OctoPocus [Bau et Mackay, 2008] et ShadowGuides [Freeman *et al.*, 2009] affichent des gestes 2D grâce à des lignes 2D colorées et étiquetées avec le nom des commandes associées. Dès lors que l'utilisateur exécute un geste, les scores de reconnaissance des gestes (retour d'information rétroactif) sont affichés via l'épaisseur des lignes correspondantes : moins le geste a de chance d'être reconnu par le système, plus la ligne le représentant sera fine. Si ce score est inférieur à un seuil limite prédéfini, le geste n'est plus affiché. Ce type de retour d'information rétroactif présente l'avantage de simplifier l'affichage à l'écran pendant l'exécution du geste.

Enfin, certaines techniques ne fournissent aucun retour d'information rétroactif. C'est le cas de techniques telles que YouMove [Anderson *et al.*, 2013] visant à guider des mouvements du corps, pendant l'étape de "Guidage de Mouvement" par exemple. C'est aussi le cas pour des guides visant simplement à collecter des données pour la conception de jeux de gestes [Kühnel *et al.*, 2011] ou pour la reconnaissance de gestes [Fothergill *et al.*, 2012] [Hoffman *et al.*, 2010], ou bien visant à guider des mouvements pour de la ré-éducation musculaire [Sodhi *et al.*, 2012].

Des études ont également montré que le retour d'information rétroactif concurrent pouvait avoir un effet négatif sur l'apprentissage moteur de gestes, les utilisateurs se reposant trop sur les informations de guidage pendant l'entrainement [Anderson et Bischof, 2013] [Schmidt et Wulf, 1997]. Il existe cependant des solutions visant à réduire cet effet négatif. Par exemple, l'Adaptive Guide [Anderson et Bischof, 2013] ne souffre pas de cet effet puisque le retour d'information rétroactif disparait avec le retour d'information prédictif en fonction de l'expérience de l'utilisateur. Cela empêche ce dernier de trop se reposer sur le guide pendant la phase d'entrainement. Ainsi, puisque (1) nous nous concentrons sur les utilisateurs novices et (2), des solutions d'adaptation existent afin de réduire l'impact négatif d'un retour d'information rétroactif concurrent sur l'apprentissage moteur de gestes, nous continuons de considérer cette option dans notre étude.

9.2.3.2 Position du guide définie par le retour d'information prédictif

Dans la littérature, certains guides choisissent d'afficher les gestes à chaque instant sous l'élément à guider (comme le curseur de la souris pour des gestes 2D ou bien une représentation numérique de la main pour des gestes 3D). Ces guides ont donc fait le choix de corriger l'affichage du retour d'information prédictif en le re-localisant en fonction de la position courante de l'utilisateur. Nous pouvons citer les exemples du guide dynamique de [Anderson et Bischof, 2013], d'OctoPocus [Bau et Mackay, 2008] et de YouMove [Anderson et Bischof, 2013]. Pour ce dernier, les guides sont rattachés à des articulations du corps. LightGuide fournit des informations de guidage projetées directement sur la main de l'utilisateur [Sodhi *et al.*, 2012]. Mais dans ce cas, le nombre de gestes affichés est limité à un seul.

D'autres guides choisissent de ne pas re-localiser l'affichage des gestes et laisse donc le guide à une position ancrée dans l'espace, indépendamment de la position courante de l'utilisateur. C'est le cas du guide static de [Anderson et Bischof, 2013] par exemple. Un autre exemple est une Crib-Sheet statique, qui répertorie généralement le nom des commandes et illustre les gestes avec des dessins. On peut aussi citer les guides présentant les gestes via des vidéos [Fothergill *et al.*, 2012], [Hoffman *et al.*, 2010] ou des animations [Sodhi *et al.*, 2012]. C'est également le cas du guide pour les mouvements de Tai Chi utilisant l'avatar d'un enseignant vu à la troisième personne [Portillo-Rodriguez *et al.*, 2008].

Même si ces deux options (re-localisation de l'affichage en fonction de la position courante de l'utilisateur ou non) ont été considérées dans de précédents travaux, leur impact sur la qualité du guidage n'a pas été évalué.
L'étude de [Anderson et Bischof, 2013] ne sépare pas les effets de la re-localisation et du retour d'information rétroactif concurrent, décrit précédemment. [Anderson et Bischof, 2013] ont évalué l'impact de quatre guides sur l'apprentissage de gestes 2D :

1. une Crib-Sheet statique affichée dans le coin supérieur gauche de l'écran,

2. un guide dynamique reliant la transparence des gestes affichés à un score durant l'exécution,

3. un guide statique, qui, comme le guide dynamique, affiche les gestes grâce à des lignes sous le curseur de la souris. Toutefois, (1) le retour d'information prédictif n'est pas impacté par le retour d'information rétroactif (c'est-à-dire que le guide affiche le geste complet jusqu'à la fin) et (2), l'affichage reste à son emplacement initial pendant toute la durée du guidage.

4. un guide adaptatif : l'Adaptive Guide fournit le même retour d'information prédictif que le guide statique, mais disparait de plus en plus tôt au fur et à mesure que l'utilisateur s'améliore.

Si nous nous concentrons sur les résultats de cette expérience pendant la phase d'entrainement, c'est-à-dire quand les utilisateurs sont encore novices, le guide statique permet aux utilisateurs d'être plus précis que le guide dynamique notamment. Malheureusement, on ne peut dire avec certitude quel paramètre est à l'origine de cet effet : l'absence d'effet du retour d'information rétroactif sur le retour d'information prédictif (c'est-à-dire la transparence se mettant à jour durant l'exécution du geste) ou bien le point d'ancrage statique du guide. Ces deux paramètres restent donc à évaluer - pour des gestes 2D ainsi que 3D.
Il est donc difficile de dire si la stabilité de la scène (lorsque le guide n'est pas re-localisé) est plus, moins, ou tout aussi importante qu'un emplacement unique pour l'attention de l'utilisateur (lorsque le guide est affiché à tout moment au même endroit que sa main).

9.2.3.3 Portion de gestes affichée par le retour d'information prédictif

La modalité visuelle peut fournir un retour d'information prédictif représentant les portions complètes restantes des gestes à effectuer. C'est le cas d'OctoPocus [Bau et Mackay, 2008], d'une Crib-Sheet [Anderson et Bischof, 2013] ou bien d'un Marking Menu [Kurtenbach *et al.*, 1993]. C'est aussi le cas de certains guides n'affichant complètement qu'un geste à la fois, que ce soit via des vidéos de présentation [Fothergill *et al.*, 2012] [Hoffman *et al.*, 2010] ou bien des animations [Sodhi *et al.*, 2012]. Même si afficher la totalité du geste dégrade la lisibilité de la scène - en particulier pour des gestes 3D - elle peut toutefois permettre à l'utilisateur d'anticiper la globalité de la forme du geste à effectuer et ainsi peut-être (1) aller plus vite et (2) mieux apprendre les gestes.

D'autres guides ne représentent que la partie à venir du geste à effectuer : une portion imminente. C'est la cas de la Crib-Sheet interactive [Kurtenbach *et al.*, 1994], d'un Hierarchical Marking Menu [Kurtenbach *et al.*, 1994] ou de LightGuide [Sodhi *et al.*, 2012]. Ce dernier propose différentes modalités visuelles projetées sur la main de l'utilisateur afin d'afficher la direction à suivre (une flèche 3D par exemple). N'afficher que la partie imminente à venir peut augmenter la lisibilité de la scène 3D, mais au détriment de la capacité d'anticipation de l'utilisateur pour l'exécution d'un geste.

Bien que ces deux options - affichage de portions complètes ou bien imminentes - aient été explorées dans de précédents travaux, leur impact sur la qualité du guidage n'a pas été évalué.

Résumé

Notre outil révèle qu'aucun guide ne satisfaisait nos besoins pour guider des utilisateurs sur un ensemble complet de geste 3D afin d'interagir à distance avec des objets physiques augmentés. Néanmoins, nous pouvons adapter un guide existant, à savoir OctoPocus, à notre cas d'étude. Pour cela, nous avons identifié des caractéristiques clés pour lesquelles nous proposons d'explorer différentes alternatives. Ces caractéristiques sont :

- *pour le retour d'information rétroactif* : la fréquence de mise à jour. Le guide peut utiliser une mise à jour discrète ou bien une mise à jour continue. La mise à jour continue est supposée aider l'utilisateur à corriger son geste pendant l'exécution de gestes 2D. Nous voulons évaluer l'impact d'une mise à jour continue lors de l'exécution de gestes 3D effectués en l'air et comparer cette stratégie avec celle utilisant une mise à jour discrète en montrant le geste effectué une fois l'exécution terminée.
- *pour le retour d'information prédictif* : l'affichage des portions de gestes ainsi que la correction de cet affichage en fonction de la position courante de la main de l'utilisateur.
  Le guide peut afficher les portions complètes restantes des gestes ou

simplement les portions imminentes à venir, impactant ainsi la complexité visuelle de la scène.
Le guide peut aussi laisser l'affichage ancré à une position prédéfinie, ou bien re-localiser l'affichage en fonction de la position de la main de l'utilisateur, impactant ainsi la stabilité de l'affichage.

Ces options de conception restent à évaluer au regard de la qualité du guidage qu'elles fournissent. Nous proposons d'évaluer ces alternatives avec OctoPocus3D, un guide conçu pour guider des gestes 3D grâce à des tubes 3D affichés sur un écran distant.

Notre approche de conception repose donc sur l'extension d'OctoPocus aux gestes 3D. Nous présentons la conception d'OctoPocus3D qui servira de base à nos études expérimentales. Même si plusieurs techniques de guidage de gestes 3D existent, aucune d'entre elles ne permet le guidage d'un large ensemble de gestes, nécessaire lorsque l'on souhaite interagir avec un environnement intelligent.

9.3 Conception d'OctoPocus3D

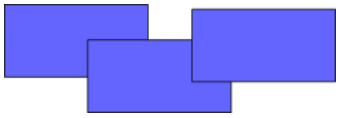
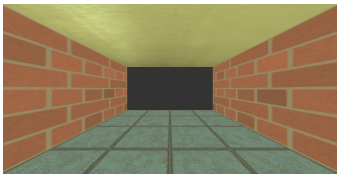
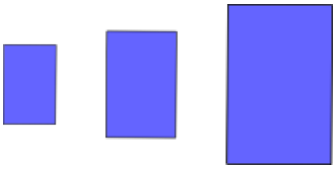
OctoPocus3D utilise une modalité de sortie visuelle afin de présenter les gestes à l'utilisateur et de le guider pendant l'exécution d'un geste. Comme annoncé en paragraphe 9.2.2.3, la modalité, définie comme une association dispositif/langage $< d, L >$, utilise un écran distant comme dispositif de sortie d. Nous nous concentrons ici sur la description du langage d'interaction L. Dans un premier temps, nous décrivons comment ce langage d'interaction présente les gestes 3D par des tubes 3D. Nous explorons ensuite les différentes options pour le retour d'information rétroactif et le retour d'information prédictif influençant ce langage d'interaction.

9.3.1 Affichage des gestes 3D

Le guide représente les gestes 3D par des tubes 3D et des étiquettes avec le nom des commandes associées placées près des tubes correspondants. Nous fournissons aussi un retour d'information rétroactif de base avec (1) une sphère blanche représentant la main de l'utilisateur dans la scène 3D et (2), un tube blanc représentant le chemin parcouru par la main jusqu'ici.
Nous avons examiné les indices visuels proposés dans la littérature qui aident à la compréhension d'une scène graphique 3D, en particulier de la perception de la profondeur. La littérature suggère environ 18 indices visuels pour la perception spatiale [Ware, 2004]. Toutefois, les effets de ces indices sont fortement couplés à la tâche de l'utilisateur [Ware, 2004]. Par exemple, la perspective linéaire rend plus difficile la réalisation d'une tâche de

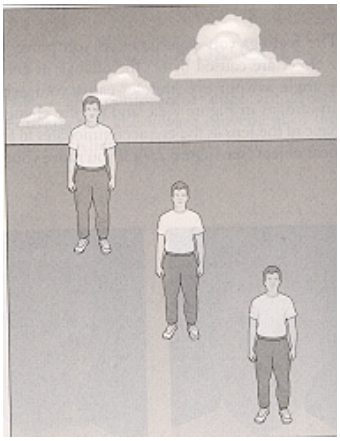
rotation, alors qu'elle correspond au meilleur indice visuel lors d'une tâche de translation [Ware, 2004]. Le nombre important de combinaisons possibles de ces 18 indices visuels rend donc impossible une exploration systématique de toutes leurs associations [Ware, 2004]. Pour la conception d'OctoPocus3D, nous avons suivi un processus itératif avec des tests utilisateur informels qui nous a conduit à choisir des éléments visuels tout en minimisant la surcharge visuelle de la scène 3D. Les Tableau 9.1, Tableau 9.2, Tableau 9.3, Tableau 9.4 et Tableau 9.5 illustrent [1] les indices visuels proposés dans la littérature concernant la perception de la profondeur. Nous décrivons ensuite les solutions de conception gardées pour la version finale d'OctoPocus3D puis celles non retenues lors de notre processus itératif de conception.

Table 9.1 – Indices graphiques [Cipiloglu *et al.*, 2010], aussi appelés monoculaires [Cutting et Vishton, 1995] ou monoculaires statiques [Ware, 2004]

Nom	Description
Occultation 	L'occultation donne de l'information de profondeur sur l'ordre des objets lorsqu'ils se chevauchent.
Perspective linéaire 	La distance entre deux lignes parallèles devient de plus en plus petite à mesure que les lignes s'éloignent du spectateur.
Gradient de taille 	La taille d'un objet est inversement proportionnelle à la distance le séparant du spectateur. Plus un objet sera grand, plus le spectateur aura l'impression qu'il est proche.

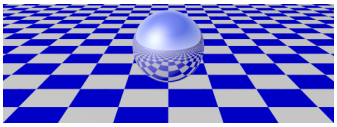
1. Certaines images proviennent des travaux de [Cipiloglu *et al.*, 2010], de [Ware, 2004] ou de [Goldstein, 2002]

hauteur relative



En présence d'une ligne d'horizon, les objets au-dessus de cette ligne semblent proches si leur base est haute (les nuages sur l'image). A l'inverse, les objets sous cette ligne semblent proches si leur base est basse (les personnages sur l'image).

Gradient de texture



Plus une surface texturée est éloignée, plus sa texture aura un aspect lisse et fin.

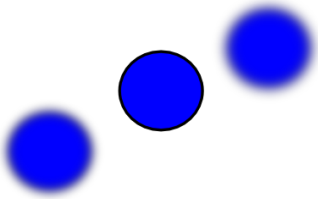
Luminosité relative

L'intensité lumineuse d'un objet varie avec sa profondeur. Un objet brillant est plus enclin à être perçu proche du spectateur.

Perspective aérienne

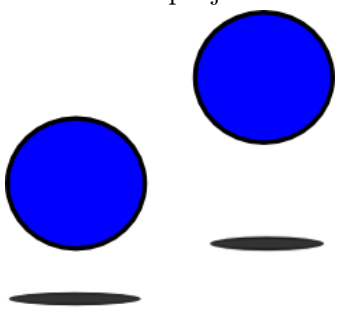
Les objets éloignés peuvent apparaitre flous et bleuâtres à cause de la diffusion de la lumière dans l'atmosphère.

Profondeur de focalisation



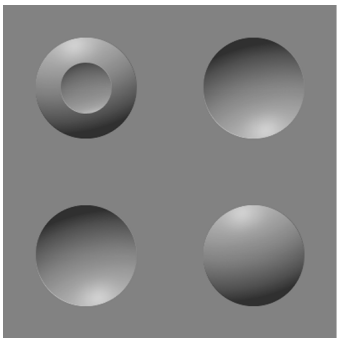
Les objets éloignés de l'endroit où le spectateur fait la mise au point visuelle semblent flous.

Ombres projetées



Les ombres projetées sur le sol facilitent la perception des positions relatives grâce à la distance entre l'objet et son ombre projetée.

Nuances d'ombres



Les nuances d'ombres permettent de percevoir la surface des objets afin de distinguer des objets convexes ou concaves. Le cerveau suppose généralement que la lumière vient d'au-dessus l'objet.

TABLE 9.2 – Indices oculomoteurs [Cipiloglu *et al.*, 2010] [Cutting et Vishton, 1995] aussi appelés binoculaires [Ware, 2004]

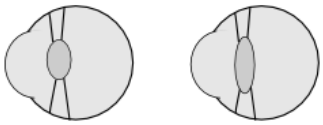
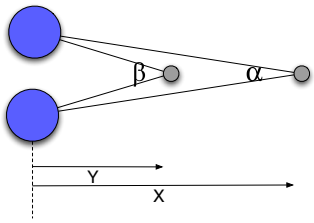
Nom	**Description**
Accommodation 	Il s'agit du processus de distorsion de la lentille de l'oeil afin de faire la mise au point sur un point donné. Cette accommodation varie avec la distance à laquelle se trouve l'objet.
Convergence 	Il s'agit de la fixation vers un point précis afin d'obtenir une unique image binoculaire. Une augmentation de l'angle de convergence signifie un rapprochement de l'objet.

TABLE 9.3 – Indices binoculaires [Cipiloglu *et al.*, 2010] [Cutting et Vishton, 1995] [Ware, 2004]

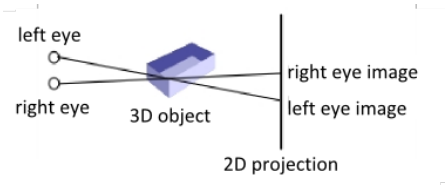
Nom	Description
Stéréoscopie 	L'écart entre les deux yeux amène le cerveau à traiter deux images rétiniennes légèrement différentes. Le cerveau arrive ensuite à interpréter la différence entre ces images comme une distance de l'objet.

TABLE 9.4 – Indices reliés à la notion de mouvement [Cipiloglu *et al.*, 2010], aussi appelés monoculaires dynamiques [Ware, 2004]

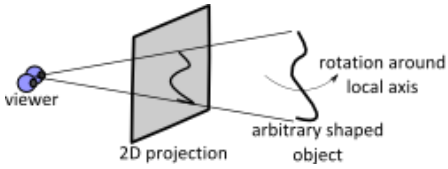
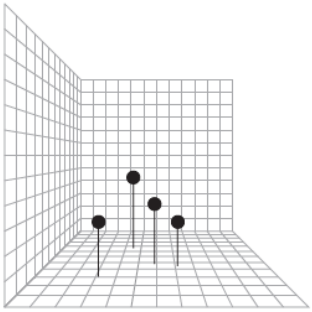
Nom	Description
Parallaxe de mouvement	C'est l'effet remarqué en regardant par la fenêtre d'un véhicule en mouvement : l'image des objets proches se déplace plus vite que celle des objets éloignés.
Profondeur cinétique 	La rotation d'un objet permet au spectateur de résoudre des ambigüités dues à la projection d'un objet 3D sur un plan 2D.

TABLE 9.5 – Indices artificiels [Ware, 2004]

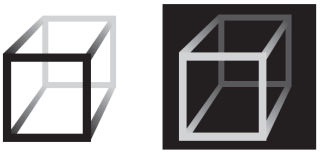
Nom	Description

Lignes d'ancrage



Comme les ombres projetées, des lignes d'ancrage perpendiculaire à un plan peuvent apporter des informations quant aux positions relatives des objets.

Luminance de proximité



Eclairage artificiel pouvant fournir une indication de profondeur.

9.3.1.1 Solutions retenues pour la version finale d'OctoPocus3D

Nous décrivons et justifions les solutions retenues pour la représentation des gestes 3D au regards des indices pour la perception de la profondeur.
Puisque OctoPocus3D affiche les gestes sur un écran 2D, certains indices tels que les indices oculomoteurs (accommodation et convergence) (Tableau 9.2) ou bien binoculaires (stéréoscopie) (Tableau 9.3) n'ont pas été considérés.

9.3.1.1.1 Occultation et transparence (Tableau 9.1) L'occultation intervient lorsque un objet en cache un autre du point de vue de l'utilisateur. C'est un indice visuel qui fournit une information binaire : un objet est devant / derrière un autre.
Nous représentons donc la main de l'utilisateur avec une sphère blanche afin que celui-ci puisse évaluer s'il se trouve derrière ou devant un geste. Il reste cependant le problème d'occultation d'un geste par un autre. Comme OctoPocus, nous avons choisi d'appliquer de la transparence sur les tubes afin que l'utilisateur puisse voir à travers les gestes tout en ayant des informations de profondeur. Les gestes sont donc composés de deux parties : un préfixe (33% de la portion totale du geste) et un suffixe (portion restante à effectuer après le préfixe) rendu plus transparent que le préfixe (Figure 9.12). Les préfixes et suffixes sont mis à jour de manière continue en fonction des mouvements de l'utilisateur. A partir des 33% finaux du geste, seul le préfixe est affiché.

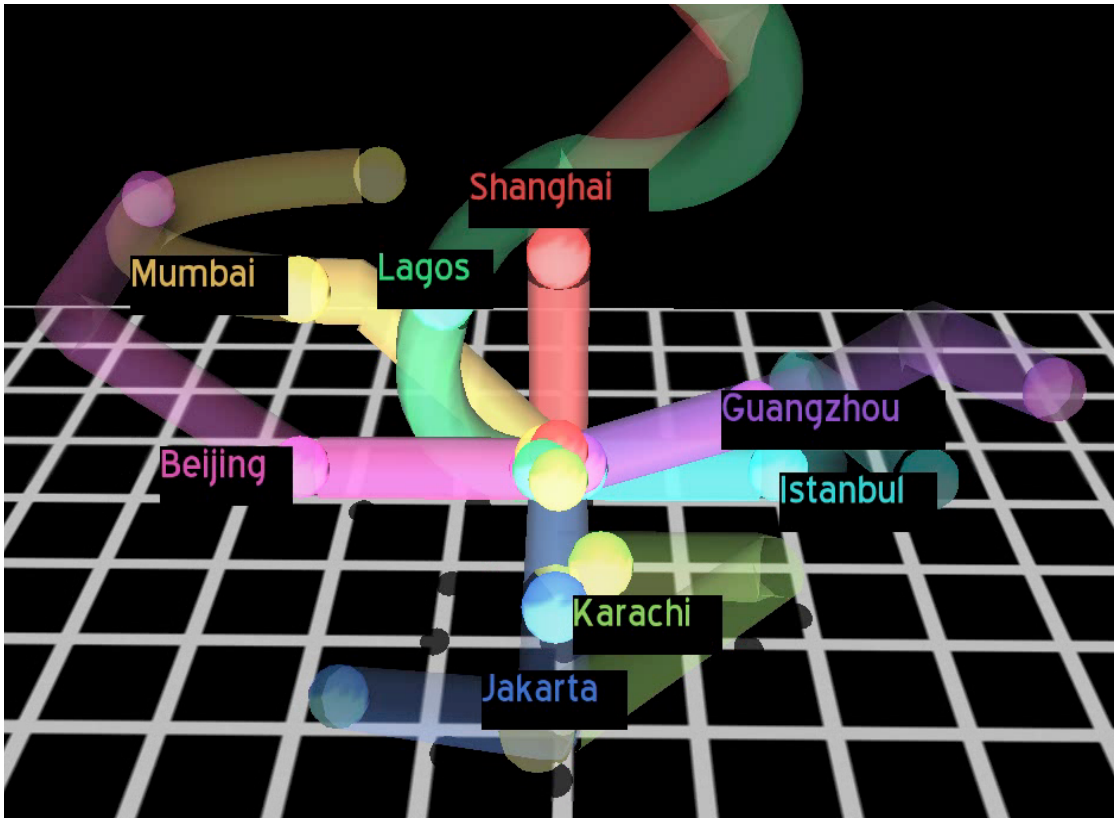


FIGURE 9.12 – Version finale d'OctoPocus3D représentant les gestes avec des tubes 3D.

9.3.1.1.2 Perspective linéaire (Tableau 9.1) Afin de souligner la perspective linéaire, nous avons ajouté un plan horizontal texturé (Figure 9.12). Nous avons choisi une texture composée de rayures blanches sur fond noir afin d'accentuer le contraste de la scène 3D. Ce plan donne à la fois des informations de profondeur (lignes parallèles convergeant vers l'horizon), mais aussi une indication sur le point de vue de la scène : vue de dessus avec un angle de 45 ° .

9.3.1.1.3 Éclairage et nuances d'ombres (Tableau 9.1) Nous avons porté une attention toute particulière à l'éclairage de la scène 3D. Bien que le plan soit trop éloigné pour que les ombres des gestes aient une utilité, les ombres créées par les gestes sur les gestes eux-mêmes fournissent de l'information quant à l'orientation des coins ainsi que des courbes.

9.3.1.1.4 Profondeur cinétique et couplage tête - affichage (Tableau 9.4) La profondeur cinétique est un mécanisme permettant à l'utilisateur d'obtenir des informations de profondeur par le mouvement. Sans système stéréoscopique, un objet 3D sera projeté sur la surface 2D d'affichage. Un couplage entre la position de la tête de

l'utilisateur et le point de vue dans la scène 3D permet aux utilisateurs de faire tourner la scène, et donc de révéler la forme 3D de l'objet. Pour cela, la caméra dans notre scène 3D est positionnée à la hauteur de la tête de l'utilisateur, faisant face au centre de la scène. Ce mécanisme permet à l'utilisateur de tourner autour des gestes 3D en penchant la partie supérieure de leur corps.

9.3.1.2 Solutions non retenues pour la version finale d'OctoPocus3D

Nous listons maintenant les solutions de conception testées durant notre processus de conception itératif. Nous décrivons les effets attendus ainsi que les raisons pour lesquelles ces solutions n'ont pas été retenues pour la version finale d'OctoPocus3D.

9.3.1.2.1 Ajouts d'éléments visuels Tout d'abord, afin d'accentuer les effets de la perspective linéaire (Tableau 9.1), nous avons ajouté quatre plans quadrillés autour des gestes : celui du bas retenu dans la version finale, ainsi que deux autres sur les côtés et un au fond de la scène, derrière les gestes. Il s'est avéré que ces plans surchargeaient la scène et créaient un effet de compression et de resserrement. Nous avons donc décidé de ne garder que le plan du bas comme expliqué précédemment.
Nous avons voulu exploiter ce plan du bas à son maximum en utilisant des lignes tombantes afin de faire ressortir les positions relatives des gestes (Tableau 9.5). Nous avons essayé trois solutions différentes pour placer l'origine de ces lignes tombantes : (1) à intervalles réguliers sur la totalité du geste, (2) aux endroits stratégiques (comme les coins par exemple) ou (3), simplement à la fin du préfixe. Il s'est avéré qu'en plus de visuellement surcharger la scène, deux facteurs ont limité l'avantage de cet indice visuel artificiel. Tout d'abord, le point de vue de la scène (vue de dessus à 45 $^\circ$) empêchait une bonne visibilité du point final des lignes sur le plan. Ensuite, le fait que l'on cherche à exécuter des mouvements continus empêche de se concentrer à la fois sur le geste à suivre et sur le plan du bas. Aussi, cet indice visuel apportait peu lors de l'exécution d'un geste.

Nous avons ensuite testé l'ajout d'éléments géométriques artificiels. Comme pour les lignes tombantes, nous avons placé des sphères, puis ensuite des cubes à différents endroits du geste (intervalles réguliers, endroits clés ou fin du préfixe). Ces formes géométriques supplémentaires devaient aider à la perception de la direction du geste grâce à des indices d'occultation locale (Tableau 9.1), de perspective (Tableau 9.1), des gradients de taille (Tableau 9.1) et des nuances d'ombres (Tableau 9.1) (Figure 9.13). Cette idée n'a pas été retenue car les formes géométriques apportaient trop de complexité à la scène, surtout en considérant l'occultation ajoutée. Si nous rendions ces formes semi-transparentes, alors les indices visuels apportés n'étaient plus présents : on ne distinguait plus les contours des cubes par exemple, rendant la perception de son orientation et donc du geste difficile. Nous avons cependant gardé l'ajout de deux sphères sur les gestes : une en fin de geste et une en fin de préfixe (Figure 9.12). Cette dernière avait l'avantage d'accentuer la

perception du mouvement du préfixe lors de l'exécution du geste (Tableau 9.4).

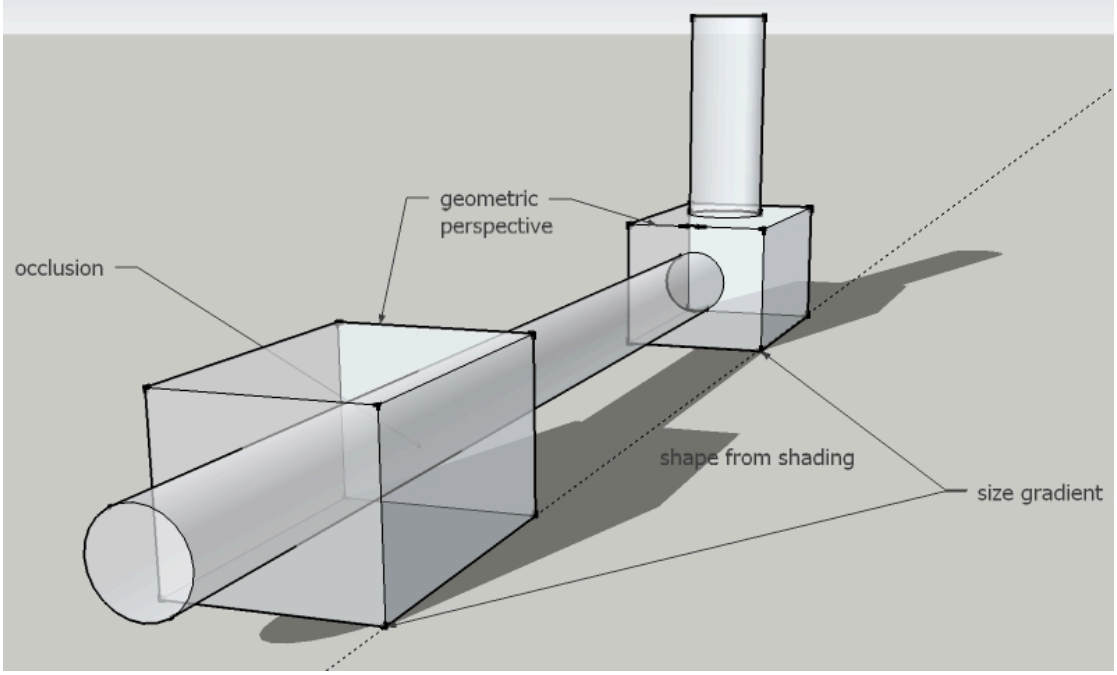


FIGURE 9.13 – Ajout de cubes répartis sur les tubes afin d'ajouter des indices de perspective, de l'occultation locale, des gradients de taille, et des nuances d'ombres.

9.3.1.2.2 Exploitation des propriétés graphiques des tubes Afin d'éviter d'encombrer la scène, nous avons décidé de profiter du fait que nos gestes soient représentés par des tubes.
Ces tubes peuvent être coupés à intervalles réguliers afin de visualiser des gradients de taille (Tableau 9.1) grâce aux cercles de ces plans de coupe, mais aussi d'ajouter un indice visuel à propos de l'orientation des gestes. Il s'est avéré qu'avec l'éclairage de la scène, ces plans de coupe n'avaient pas l'effet escompté (Figure 9.14), d'autant plus avec de la semi-transparence.
Nous avons aussi expérimenté l'ajout d'une texture aux tubes afin d'accentuer la perspective linéaire (Tableau 9.1). De plus, le gradient de texture (Tableau 9.1) donnait des indications sur l'orientation des gestes. Avec une texture, les tubes deviennent opaques et créent de l'occultation. En appliquant de la semi-transparence, la texture devient difficile à exploiter. Aussi, nous n'avons pas gardé cette solution.

Nous avons exploré des formes géométriques pouvant être utilisées pour représenter les gestes autres que des tubes droits, en particulier des tubes en spirales. Nous avons testé deux approches : combiner les deux solutions (tubes droits et tubes en spirales) (Figure 9.15a) ou n'utiliser que les spirales (Figure 9.15b). Utiliser des spirales (avec ou sans tubes à l'intérieur) apporte de l'occultation locale (Tableau 9.1), mais aussi accentue l'effet de la perspective linéaire (Tableau 9.1) et du gradient de taille (Tableau 9.1). Comme

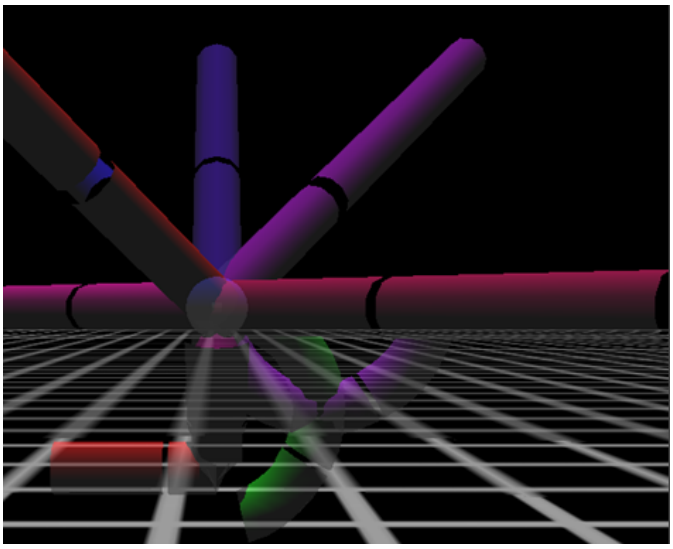


FIGURE 9.14 – Ajout de plans de coupe aux tubes afin de visualiser des gradients de taille et ajouter un indice visuel quant à l'orientation du geste.

pour la texture, les spirales peuvent donner une indication d'éloignement. Finalement, les nuances d'ombres (Tableau 9.1) indiquent aussi l'orientation du geste.
La scène visuelle n'a pas été retenue à cause de sa complexité. Il convient d'approfondir les options visuelles pour cette solution de représentation des gestes avec des spirales. Aussi nous avons choisi de mettre en suspens cette piste de recherche et de continuer à explorer les options de conception pour le comportement de la technique de guidage en utilisant des tubes 3D (extension directe des guides 2D).

9.3.2 Retour d'information rétroactif : terminal et/ou continu

Avec les techniques [Bau et Mackay, 2008] et [Freeman *et al.*, 2009], les gestes étant moins susceptibles d'être reconnus sont représentés par des traits plus minces que ceux des autres gestes. Si le score courant passe sous un seuil prédéfini, le geste correspondant n'est alors plus affiché. D'autres guides utilisent la transparence plutôt que l'épaisseur des traits [Anderson et Bischof, 2013]. Nous avons choisi de faire varier l'épaisseur des tubes. Cette solution présente l'avantage de simplifier la scène 3D pendant l'exécution d'un geste en (1) diminuant certains rayons et (2), faisant disparaitre les gestes à priori inutiles. Nous illustrons ces deux points à la Figure 9.16.

Comme souligné dans le paragraphe 9.2.3, l'impact d'un retour d'information rétroactif concurrent (en 2D ou en 3D) sur le taux de reconnaissance ainsi que sur le temps d'exécution d'un geste par des novices n'a pas été évalué. Il est évident que ce type de mécanisme simplifie le rendu de la scène 3D. Mais est-ce que les novices utilisent réellement l'information fournie par le rayon des tubes pour corriger leur geste en cours d'exécution ? Nous sommes donc intéressés par l'évaluation d'un retour d'information

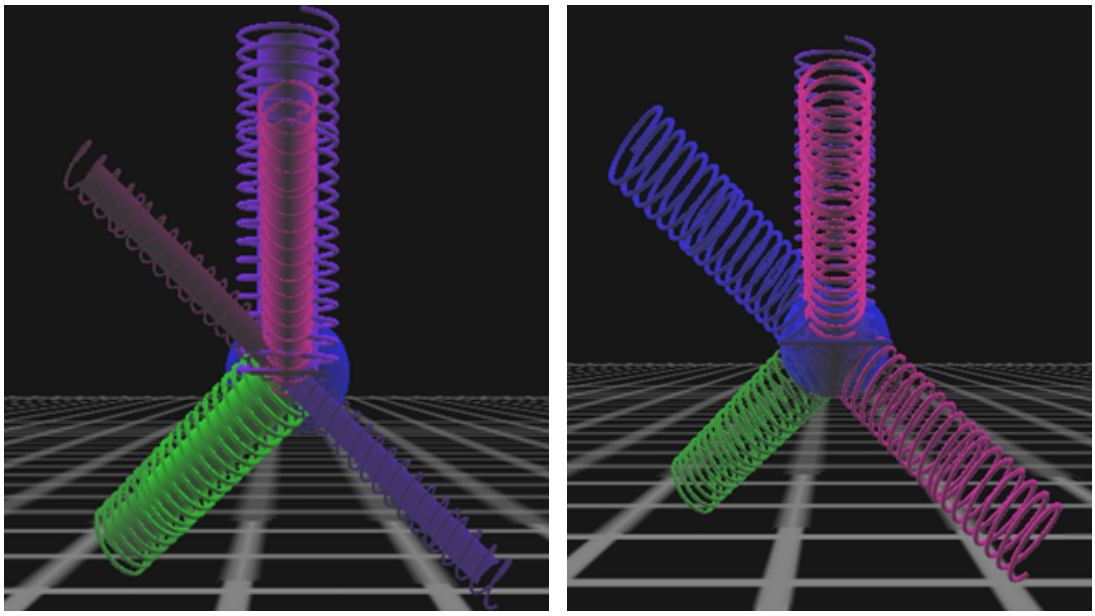
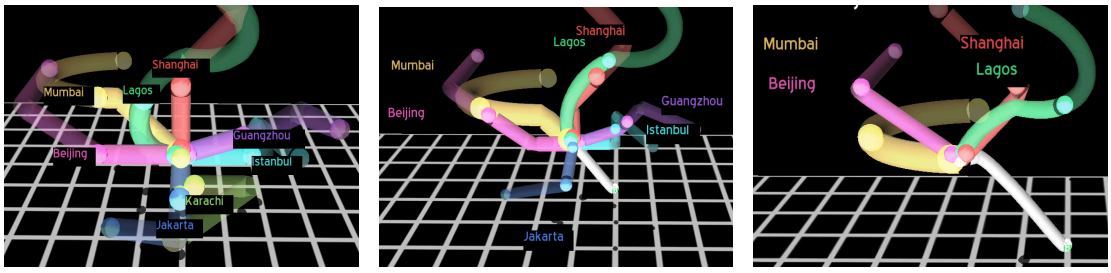


FIGURE 9.15 – Représentation du jeu de geste (a) avec des tubes droits entourés de spirales, (b) avec seulement des spirales.

rétroactif concurrent (Figure 9.16) sur le taux de reconnaissance ainsi que le temps d'exécution. Nous voulons aussi comparer ce retour d'information rétroactif concurrent avec les autres options mises en évidence par l'étude des techniques existantes (paragraphe 9.2.3) : aucun retour d'information rétroactif, un retour d'information rétroactif terminal ou bien une combinaison des deux types de retour d'information rétroactif.
Le retour d'information rétroactif terminal conçu fournit deux informations (Figure 9.17). Tout d'abord, la couleur du tracé du geste effectué donne une information binaire : il devient vert (resp. rouge) si le geste est correct (resp. incorrect). Deuxièmement, le guide affiche aussi le geste qui était à réaliser. Ainsi, les utilisateurs peuvent explorer la scène 3D grâce au couplage tête-caméra afin de mieux visualiser les différences entre le geste effectué et le geste à effectuer.

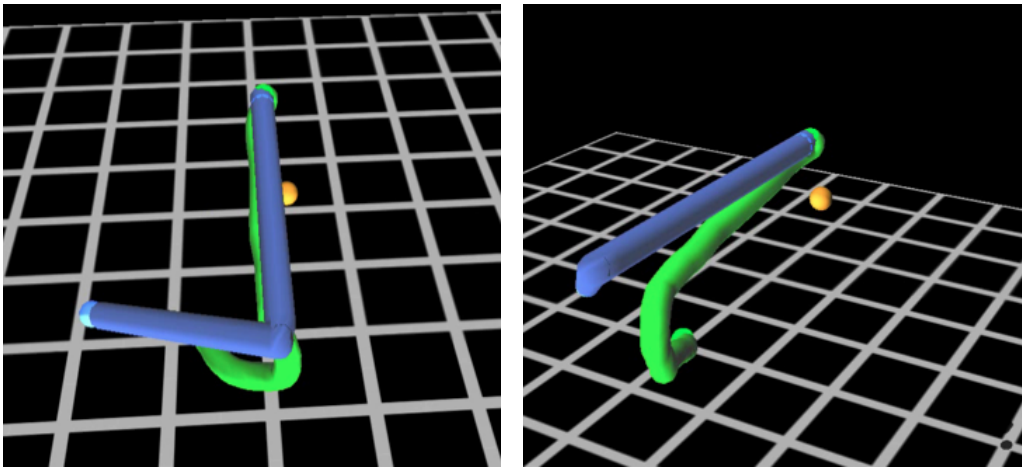
9.3.3 Retour d'information prédictif : origine du guide et portions des gestes affichées

Le mécanisme de retour d'information prédictif permet à l'utilisateur de savoir quelles sont les commandes disponibles ainsi que les gestes à effectuer pour déclencher ces commandes. Avec la modalité de sortie choisie (à savoir une présentation graphique des gestes affichés par des tubes 3D sur un écran distant), l'un des premiers défis est d'afficher plusieurs gestes 3D à la fois de manière compréhensible. En effet, la troisième dimension apporte une difficulté supplémentaire non présente lors du guidage de gestes 2D : la profondeur. Pour faciliter la compréhension de la scène visuelle nous avons d'abord examiné les indices



(a) (b) (c)

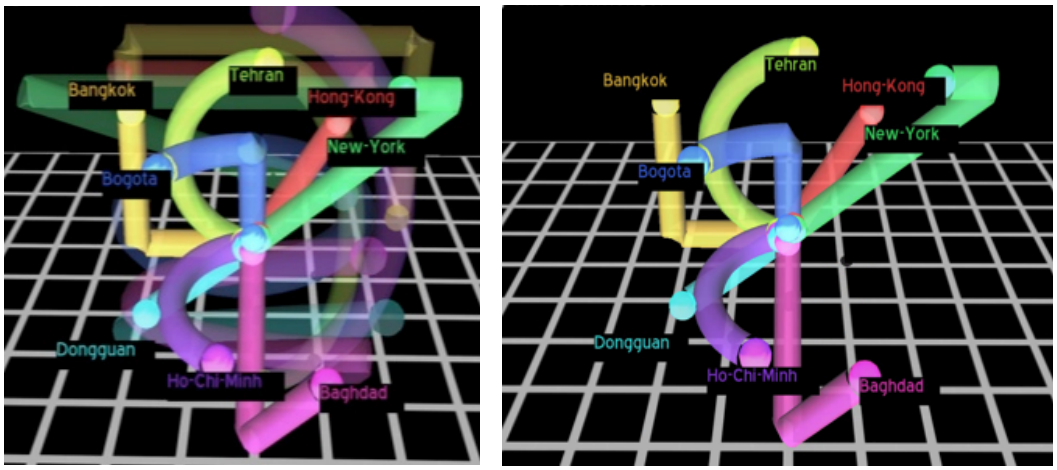
FIGURE 9.16 – Illustration du retour d'information rétroactif concurrent. Le mécanisme de retour d'information prédictif affiche la portion restante complète des gestes à effectuer à partir de la représentation graphique de la main. (a) Scène d'origine. (b) L'utilisateur commence à suivre un geste vers le haut-gauche de l'écran. Le rayon des gestes sur la droite de la scène ayant peu de probabilité d'être reconnus a diminué (retour d'information rétroactif concurrent). (c) Les gestes avec un score de reconnaissance sous un seuil limite ont disparu (retour d'information rétroactif concurrent).



(a) (b)

FIGURE 9.17 – Illustration du retour d'information rétroactif terminal et de la profondeur cinétique. Le geste à effectuer est le geste bleu. Le geste exécuté par l'utilisateur est en vert. (a) A partir du point de vue d'origine. (b) A partir d'une perspective inclinée sur la droite.
Nous rappelons que ce retour d'information rétroactif terminal n'est possible que si le système connait le geste que l'utilisateur souhaite réaliser, comme en situation d'entraînement.

visuels pour représenter les gestes 3D. Les solution explorées et choisies ont été détaillées dans le paragraphe 9.3.1.
La complexité visuelle de la scène dépend aussi de plusieurs autres facteurs, tels que les gestes eux-mêmes ou le nombre de gestes dans le jeu de gestes à afficher. Aussi, au



(a) Portions restantes complètes (b) Portions à venir

FIGURE 9.18 – Représentation du jeu de geste numéro 3. (a) Portions restantes complètes à effectuer des gestes. (b) Portions imminentes à venir des gestes.

delà de la présentation visuelle des gestes 3D par des tubes, nous avons identifié deux paramètres de conception issus de notre étude des techniques existantes (paragraphe 9.2.2 et paragraphe 9.2.3) : la quantité d'information affichée et la stabilité de cet affichage.

La première option de conception est la quantité d'information à afficher via les tubes 3D (Figure 9.18). Le guide peut afficher les portions restantes complètes des gestes (comme OctoPocus [Bau et Mackay, 2008]), une portion imminente à venir des gestes (comme YouMove [Anderson *et al.*, 2013]) ou seulement la direction à suivre (comme LightGuide [Sodhi *et al.*, 2012]). Ce facteur influence directement la complexité de la scène 3D en modifiant la quantité d'information affichée par les tubes 3D (Figure 9.18). Pour la condition *portions à venir*, nous avons choisi de ne représenter que le préfixe du geste (33% du geste).

La seconde option de conception concerne l'origine du guide. Pour une technique de guidage 2D, l'affichage du guide est souvent re-localisé sous le curseur de la souris. Ainsi, à tout instant, la représentation des gestes est centrée sous le curseur. Cependant, (1) cette solution pourrait être moins précise qu'une représentation fixe [Anderson et Bischof, 2013] et (2), cette solution est susceptible d'apporter du bruit et ainsi augmenter la difficulté pour interpréter une scène 3D déjà encombrée. De plus, les tremblements inhérents à des gestes 3D effectués en l'air sans support peuvent rendre la scène plus instable qu'en effectuant des gestes 2D sur support. Ces deux raisons nous ont motivés à considérer une représentation fixe : les tubes 3D sont centrés dans la scène et seulement la représentation de la main se déplace parmi les tubes.

Résumé

Nous avons suivi un processus itératif avec des tests utilisateur informels afin d'explorer différentes possibilités pour accentuer la perception de la profondeur dans notre scène 3D. Ce processus itératif nous a permis d'affiner la représentation visuelle des gestes faite avec des tubes en utilisant avec parcimonie les notions d'occultation (semi-transparence d'une partie des gestes), de perspective linéaire (plan texturé), d'éclairage, d'ombres et de profondeur cinétique (couplage tête-caméra). Nous avons ensuite décrit nos différents mécanismes de retour d'information rétroactif et retour d'information prédictif.

9.4 Évaluations expérimentales

Tandis que nous avons suivi un processus itératif avec des tests utilisateur informels pour définir la représentation des gestes par des tubes à l'écran, nous avons mené deux expériences pour les choix de conception du retour d'information rétroactif et celui prédictif (Figure 9.19).
Pour l'expérience 1 dédiée au mécanisme de retour d'information rétroactif, le retour d'information prédictif affiche les portions restantes complètes des gestes à partir du centre de la scène 3D. L'efficacité de cette solution a été prouvée dans le contexte de guidage de gestes 2D [Anderson et Bischof, 2013]. Pour l'expérience 2, nous serons alors en mesure d'explorer différentes options du retour d'information prédictif en considérant le meilleur retour d'information rétroactif.

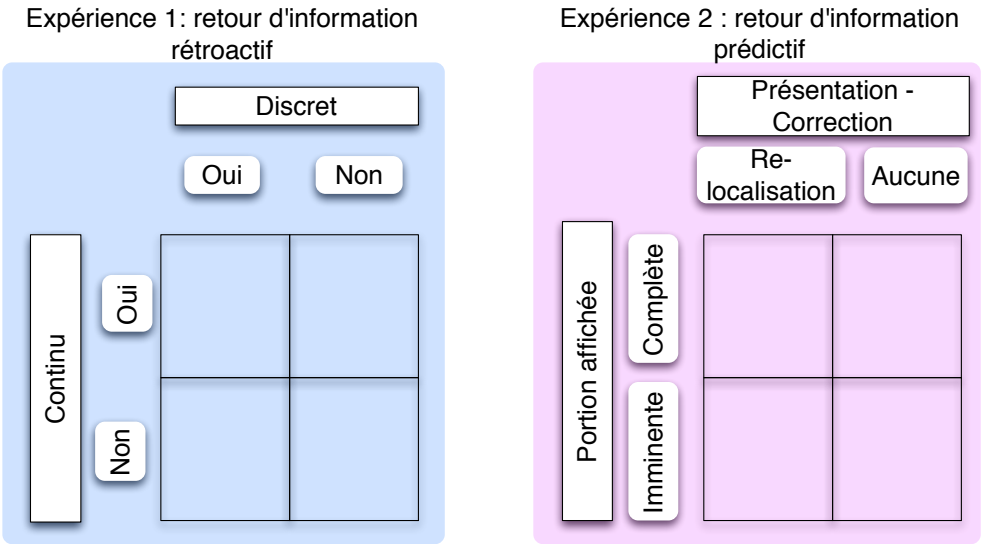


FIGURE 9.19 – Résumé des deux expériences au regard des options de conception évaluées.

9.4.1 Ressources

Nous décrivons les ensembles de gestes ainsi que le matériel et la configuration communs à nos deux expériences.

9.4.1.1 Conception des jeux de gestes pour les évaluations

Notre objectif est de comparer différentes options de conception pour notre technique de guidage OctoPocus3D. Toutefois, si les gestes sont trop faciles à reconnaitre, nous ne pourrons pas observer de différences entre ces variations. Nous avons donc créé différents jeux de gestes délibérément difficiles qui comprennent de l'auto-occultation (soit une partie d'un geste cachée par une partie du même geste), des gestes avec des formes inhabituelles, mais aussi de l'occultation entre différents gestes (soit un geste derrière un autre geste), du chevauchement de gestes (soit des gestes partageant une même portion) et de la proximité - à la fois spatiale mais aussi au niveau de la forme des gestes. Idéalement, ce type de situation devrait être évité. Mais l'un des objectifs souhaitables pour une technique de guidage est d'autoriser un riche vocabulaire de gestes et ainsi d'améliorer la puissance expressive de l'interaction gestuelle. De fait, l'augmentation du nombre de gestes dans un jeu donné est susceptible de créer les difficultés mentionnées ci-dessus. Nous avons donc créé 4 jeux de 8 gestes chacun pour un total de 32 gestes (Figure 9.20). Dans le cadre de nos expériences en laboratoire, chaque geste était relié à un nom de ville plutôt qu'à un réel nom de commande.

9.4.1.2 Équipement et configuration

Les deux expériences utilisent le même équipement et la même configuration. Le logiciel, développé en C++, utilise Ogre comme moteur de rendu 3D et est exécuté sur un ordinateur portable avec un processeur 2.40 GHz. La communication avec la Kinect est prise en charge par les bibliothèques OpenNi2 et Nite2. La Kinect permet le suivi des articulations principales du corps humain. Nous avons considéré (1) la main dominante pour l'exécution des gestes et (2) l'axe torse/tête pour le couplage entre la perspective de la scène 3D et la tête de l'utilisateur.
La reconnaissance de gestes utilise l'implémentation de l'algorithme Dynamic Time Warping de GRT[2]. Les modèles de gestes utilisés par le système de reconnaissance sont générés mathématiquement par le système. Les classes de gestes étaient donc créer artificiellement et ne comportaient qu'une seule instance d'un geste *parfait*.
Pendant l'expérience, les utilisateurs étaient debout et situés 2 mètres en face de la Kinect et de l'écran (Figure 9.21). Puisque l'initialisation (début du geste) et la terminaison (fin du geste) n'étaient pas l'objet d'étude, nous les avons capturées grâce à un clic de souris tenu dans la main gauche (non dominante).

2. http://www.nickgillian.com/software/grt (lien valide en septembre 2015)

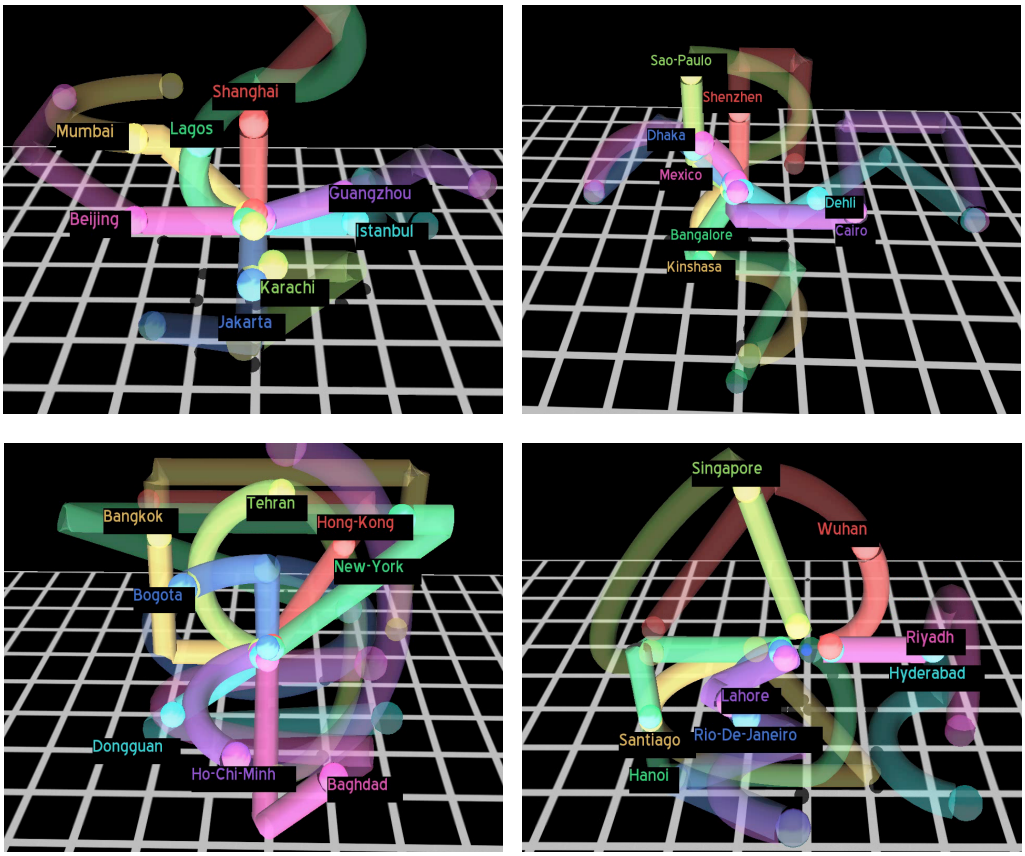


FIGURE 9.20 – Les quatre jeux de gestes utilisés dans nos expériences. Afin de présenter l'ensemble des gestes, les jeux de gestes sont ici illustrés avec un retour d'information prédictif affichant les portions complètes des gestes.

9.4.1.3 Etude préliminaire : OctoPocus3D *vs* vidéos de démonstration

Avant même d'évaluer les différentes versions d'OctoPocus3D, nous voulions savoir si le concept global permettait en effet de guider les utilisateurs. Pour cela, nous avons évalué si notre prototype pouvait mener à de meilleurs taux de réussite que des vidéos montrant les gestes à effectuer. En effet, il est courant d'utiliser une vidéo de démonstration comme technique de guidage ou bien comme point de comparaison dans la littérature [Hoffman *et al.*, 2010] [Kühnel *et al.*, 2011] [Sodhi *et al.*, 2012]. C'est aussi une solution courante dans des applications commerciales telles que le menu de préférences du pavé tactile d'Apple ou bien la télévision intelligente de Samsung [3]. La version d'OctoPocus3D évaluée dans cette étude préliminaire affiche la portion complète des gestes à réaliser centrée dans la scène 3D.

Pour ce pilote, 4 participants ont utilisé les deux techniques sur deux jeux de gestes.

3. http://www.samsung.com/ph/smarttv/common/guide_book_3p_si/main.html (lien valide en septembre 2015)

FIGURE 9.21 – Dispositif expérimental situé en contexte à l'entrée d'un restaurant.

OctoPocus3D a mené à un taux de réussite d'environ 80% (M=80,6%, 95% CI=[67,4%, 93,7%]) alors que la vidéo a conduit à un taux de réussite inférieur à 50% (M=46.9%, 95% CI=[29.3%, 64.4%]). Ces résultats ont confirmé deux hypothèses : (1) OctoPocus3D permet effectivement de guider les utilisateurs, il est donc intéressant d'étudier ce type de guidage et (2), nos jeux de gestes sont suffisamment difficile pour nécessiter du guidage.

9.4.2 Expérience 1 : retour d'information rétroactif

Nous nous intéressons aux alternatives de conception pour le retour d'information rétroactif en considérant un retour d'information rétroactif concurrent, un retour d'information rétroactif terminal, leur combinaison ou bien aucun retour d'information rétroactif.

9.4.2.1 Participants et plan d'expérience

20 sujets droitiers ont participé à l'expérience. (4 femmes, 16-47 ans, M=28.4, SD=7.82). Puisque nous voulions évaluer OctoPocus3D sur des novices, aucun participant n'était membre du laboratoire et aucun n'avait de formation en informatique. Cinq d'entre eux étaient habitués à l'interaction gestuelle avec des jeux vidéos. Les gestes de ces jeux vidéos avaient été appris grâce à un ami (4 d'entre eux) ou bien grâce à un tutoriel (1 d'entre eux). Par conséquent, aucun participant n'avait utilisé une technique de guidage de gestes 3D.

L'expérience a duré environ 90 minutes. Nous avons utilisé un protocole à mesures répétées intra-participant. Les facteurs expérimentaux étaient les différentes versions du retour

d'information rétroactif : concurrent (CONC, oui/non), et terminal (TERM, oui/non) et le nombre de répétitions REP (1, 2, 3, 4, 5, 6) afin d'analyser la progression des utilisateurs au cours du temps. La session était composée de 4 blocs (Figure 9.22), un pour chaque combinaison de retour d'information rétroactif : concurrent, terminal, les deux à la fois ou aucun. Les participants ont utilisé chaque version sur des jeux de gestes différents composés de 8 gestes chacun (les quatre jeux de gestes décrit précédemment et représentés à la Figure 9.20). Avant chaque bloc, le comportement du retour d'information rétroactif était expliqué à l'utilisateur. Il était demandé aux utilisateurs d'effectuer chaque geste 6 fois, sachant que les gestes étaient présentés au hasard à chaque répétition. L'ordre de présentation de chaque version du retour d'information rétroactif était contrebalancé selon un carré latin 4×4. Les gestes n'étant pas un facteur expérimental, l'ordre des jeux de gestes était inchangé. L'ordre de présentation des techniques permet déjà d'utiliser chaque systèmes sur tous les jeux de gestes.
Ce protocole d'expérimentation a entraîné 6 répétitions $\times$ 8 gestes $\times$ 4 techniques $\times$ 20 participants = 3840 acquisitions. A la fin de la session, nous avons demandé aux participants de classer les techniques selon leurs préférences : de 1 (la technique préférée) à 4 (la technique la moins préférée).

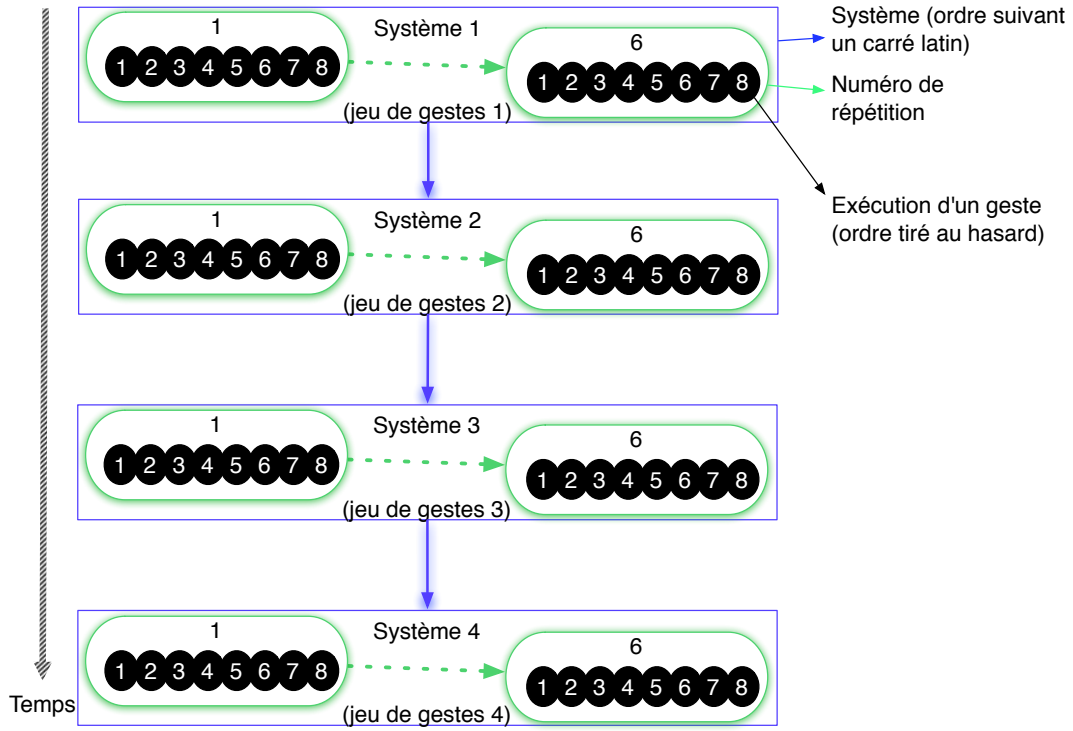


FIGURE 9.22 – Protocole d'expérimentation pour évaluer les options de retour d'information rétroactif. Quatre blocs (un pour chaque technique avec un jeu de gestes dédié), chacun composé de 6 répétitions elles-mêmes composées de 8 exécutions de gestes.

A la différence de notre évaluation des techniques de désambigüisation, nous ne formulons

pas d'hypothèse quant aux performances des techniques évaluées. En effet, nous sommes dans une phase d'exploration en considérant des options de conception proposées par les travaux existants mais non évaluées jusqu'ici.

9.4.2.2 Résultats

Les principales mesures dépendantes étaient la précision, à savoir le taux de reconnaissance obtenu avec une version de la technique de guidage donnée et la durée d'exécution, à savoir le temps écoulé entre l'étape d'initialisation (premier clic de souris) et de terminaison (dernier clic de souris). A tout moment, les participants pouvaient réinitialiser le guide en cliquant sur le bouton du milieu de la souris.

9.4.2.2.1 Taux de reconnaissance Nous avons effectué une analyse de variance à mesures répétées. Les résultats ne montrent aucun effet du retour d'information rétroactif concurrent CONC [$F(1.19) = 4.27$, $p > 0.05$], terminal TERM [$F(1.19) = 3.16$, $p > 0.05$] ou encore du nombre de répétition REP [$F(5.95) = 1.84$,$p > 0.05$] sur le taux de reconnaissance.

Premièrement, il semble que dès la troisième répétition, les participants réussissent par eux-mêmes à combler l'écart entre les différents retours d'information rétroactifs visible lors des répétitions 1 et 2 (Figure 9.23). Par conséquent, il n'y a pas de différences entre les techniques durant 4 répétitions sur 6. Deuxièmement, les différents retour d'information rétroactif montrent peu ou aucune variation du taux de reconnaissance au cours du temps. Cela démontre ainsi qu'un retour d'information rétroactif concurrent et/ou terminal n'a pas aidé les participants à progresser comme nous l'avions prévu.

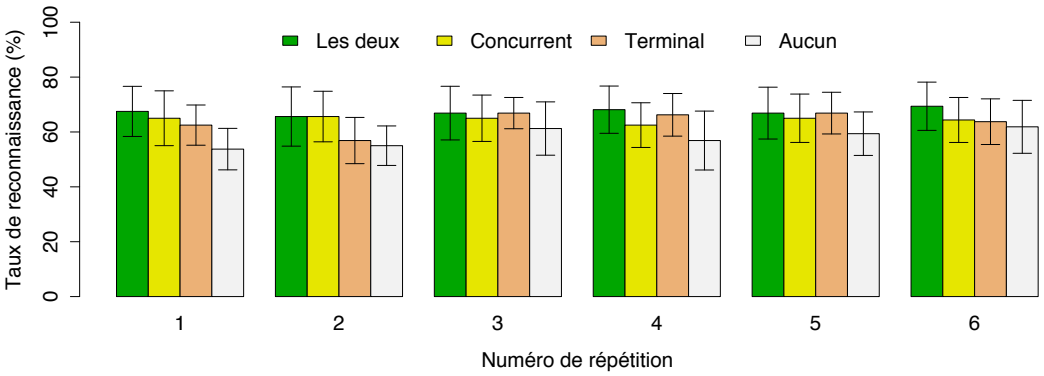


FIGURE 9.23 – Taux de reconnaissance (en pourcentage) pour chaque combinaison de retour d'information rétroactif durant les 6 répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%.

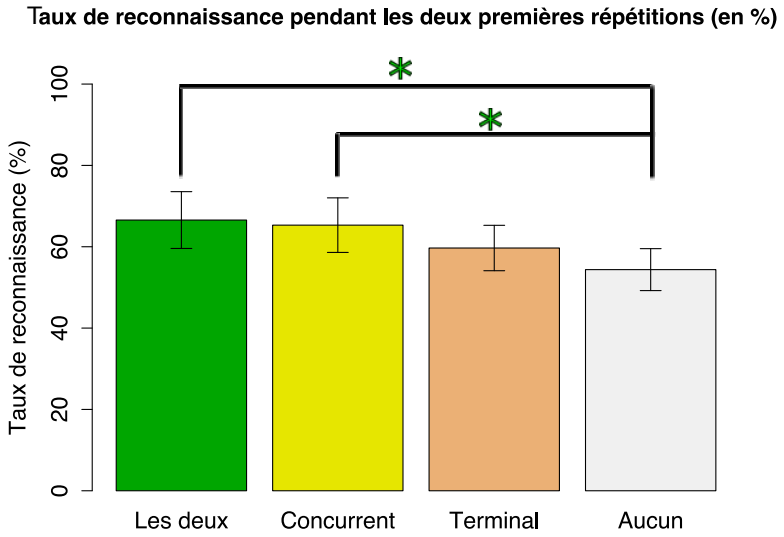


FIGURE 9.24 – Taux de reconnaissance (en pourcentage) pour chaque combinaison de retour d'information rétroactif agrégés sur les deux premières répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%.

Nous affinons notre analyse en se concentrant sur les deux premières répétitions. Il y a un effet significatif de CONC sur le taux de reconnaissance [$F(1.19) = 8.20$, $p < 0.01$] : un retour d'information rétroactif concurrent mène à un meilleur taux de reconnaissance qu'une technique ne proposant aucun retour d'information rétroactif (Figure 9.24). A l'inverse, le retour d'information rétroactif terminal TERM n'a pas d'effet significatif sur ce taux de reconnaissance pendant les deux premières répétitions [$F(1.19) = 1.17$, $p > 0.05$].

Nous pouvons tirer deux conclusions à partir de ces résultats. Premièrement, un retour d'information rétroactif concurrent permet un meilleur taux de reconnaissance par rapport à aucun retour d'information rétroactif pendant la phase de familiarisation avec la technique de guidage. Ce résultat est probablement dû à l'impact qu'a ce retour d'information rétroactif concurrent sur la représentation des gestes par des tubes 3D : durant l'intervalle de temps nécessaire à l'exécution d'un geste, le retour d'information rétroactif concurrent simplifie rapidement la scène 3D. Deuxièmement, les participants ont présenté une adaptation relativement rapide avec toutes les techniques, prouvant qu'ils peuvent être tout aussi précis quel que soit le retour d'information rétroactif fournit après deux répétitions seulement. Ce résultat est important dans un contexte où l'on cherche à introduire l'interaction gestuelle à des utilisateurs novices. De plus, ce résultat peut être mis en relation avec de précédents travaux montrant que le retour d'information rétroactif concurrent est néfaste pour l'apprentissage moteur des gestes. Désormais, nous savons aussi qu'au bout de la troisième répétition, ce retour d'information rétroactif n'est

plus utile pour les novices.

9.4.2.2.2 Temps d'exécution Une analyse de variance à mesures répétées montre un effet important du nombre de répétition REP sur le temps d'exécution [F(5.95)=35.06, p<0.001]. Des comparaisons multiples par paires effectuées avec des t-tests et une correction de Bonferroni montrent une accélération significative jusqu'à la troisième répétition [$p < 0.05$] (Figure 9.25). Bien que la sixième répétition soit significativement différente de la troisième [$p < 0.001$], les répétitions 5 et 6 ne sont pas significativement différentes, révélant ainsi un palier dans la progression des participants.

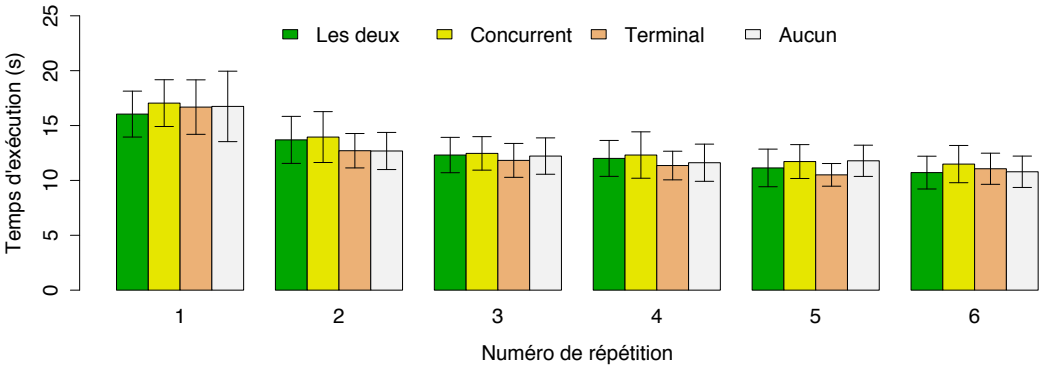


FIGURE 9.25 – Temps d'exécution (en secondes) pour chaque combinaison de retours d'information rétroactif sur les 6 répétitions. Les barres d'erreur indiquent les intervalles de confiance à 95%.

Il est intéressant de noter que nous n'avons pas trouvé d'effet significatif de CONC [$F(1.19) = 1.15$, $p > 0.05$] et TERM [$F(1.19) = 0.55$, $p > 0.05$] sur le temps d'exécution. Alors qu'on pourrait penser que les participants ralentissent avec CONC afin de corriger leurs gestes pendant l'exécution en s'appuyant sur les informations fournies par le rayon des tubes. Bien au contraire, la plupart des participants ont rapporté ne pas avoir exploité ce retour d'information rétroactif concurrent. Cette affirmation est confirmée par le classement qualitatif effectué en fin de session : les participants ont préféré les deux alternatives fournissant du retour d'information rétroactif terminal (les deux et terminal) [$\chi^2(3) = 28.48$, $p < 0.001$] (Figure 9.26).

9.4.2.2.3 Discussion Il s'avère que les participants n'ont pas utilisé le retour d'information rétroactif concurrent (d'après leurs retours et le fait qu'ils n'aient pas ralenti pour corriger leurs gestes). Mais puisque le retour d'information rétroactif concurrent a tout de même un effet significatif sur le taux de reconnaissance pendant les deux premières répétitions, nous supposons que son bénéfice provient du fait qu'il simplifie la scène au cours de l'exécution d'un geste. Par conséquent, le retour d'information rétroactif

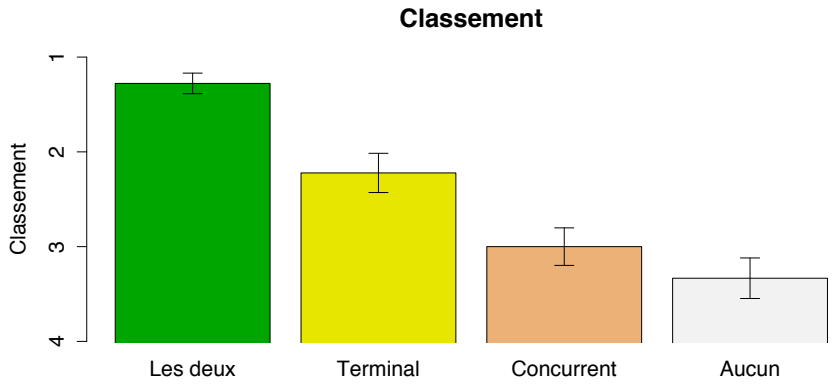


FIGURE 9.26 – Classement qualitatif de chaque combinaison de retours d'information rétroactif (la technique préférée est classée première). Les barres d'erreur indiquent les intervalles de confiance à 95%.

concurrent facilite la prise en main de la technique de guidage, mais les utilisateurs ne prennent pas le temps de corriger leurs gestes.
Si le système connait le geste que l'utilisateur souhaite exécuter (lors d'une phase d'entrainement par exemple), alors le guide devrait inclure à la fois le retour d'information rétroactif concurrent (meilleur taux de reconnaissance lors des premiers essais) et terminal (préférence des utilisateurs). Néanmoins, si aucune phase d'entrainement permettant d'inclure un retour d'information rétroactif terminal n'est prévue, n'utiliser qu'un retour d'information rétroactif concurrent est une solution viable.

9.4.3 Expérience 2 : retour d'information prédictif

Nous nous concentrons maintenant sur les deux paramètres liés au retour d'information prédictif : la quantité d'information affichée et sa stabilité. Dans le cadre de cette évaluation expérimentale, notre guidage peut fournir du retour d'information rétroactif concurrent (performances) et terminal (préférences).

9.4.3.1 Participants et plan d'expérience

24 sujets droitiers ont participé à l'expérience (10 femmes, 18-43 ans, M=25.0, SD=7.7). Aucun d'entre eux n'avaient participé à la première expérience. Comme pour l'expérience 1, aucun d'entre eux n'avait suivi une formation en informatique. 14 participants avaient déjà utilisé de l'interaction gestuelle avec des jeux vidéos. Parmi eux, 5 ont appris les gestes grâce à des vidéos de démonstration ou des images, 5 grâce à des tutoriels et 4 grâce à une démonstration d'un ami. Donc aucun participant n'avait déjà utilisé une technique de guidage de gestes 3D.

L'expérience a duré environ une heure. Nous avons utilisé un protocole à mesures répétées intra-participant. Puisque le retour d'information prédictif joue un rôle important lors de la prise en main d'une technique, nous voulions évaluer l'impact de ses différentes variations *avant* que l'utilisateur apprenne les gestes. Nous avons donc conçu l'expérience en réduisant les possibilités d'apprentissage au cours du temps. Les participants ont utilisé une technique de guidage sur un jeu de gestes donné une seule fois avant d'utiliser une autre technique sur un autre jeu de gestes (Figure 9.27). Ils ont répété l'opération trois fois. Avant chaque utilisation d'une technique, nous rappelions verbalement ses caractéristiques aux utilisateurs. Les gestes étaient présentés de manière aléatoire et l'ordre de présentation des techniques était contrebalancé selon un carré latin 4×4. Ce protocole a mené à 8 gestes $\times$ 4 techniques $\times$ 3 répétitions $\times$ 24 participants $=$ 2304 acquisitions.
Les variables indépendantes étaient la portion de geste affichée POR (imminente, complète) et la correction de l'origine du guide COR (aucune, re-localisation). La condition *imminente* affichait uniquement une portion correspondant aux 33% du geste à venir. La condition *aucune* laissait l'origine du guide au centre de la scène 3D alors que la condition *re-localisation* replaçait à chaque instant l'origine du guide au même endroit que la représentation graphique de la main de l'utilisateur. En fin de session, les utilisateurs devaient indiquer leurs préférences quant aux techniques évaluées : de 1 (technique préférée) à 4 (technique la moins préférée).

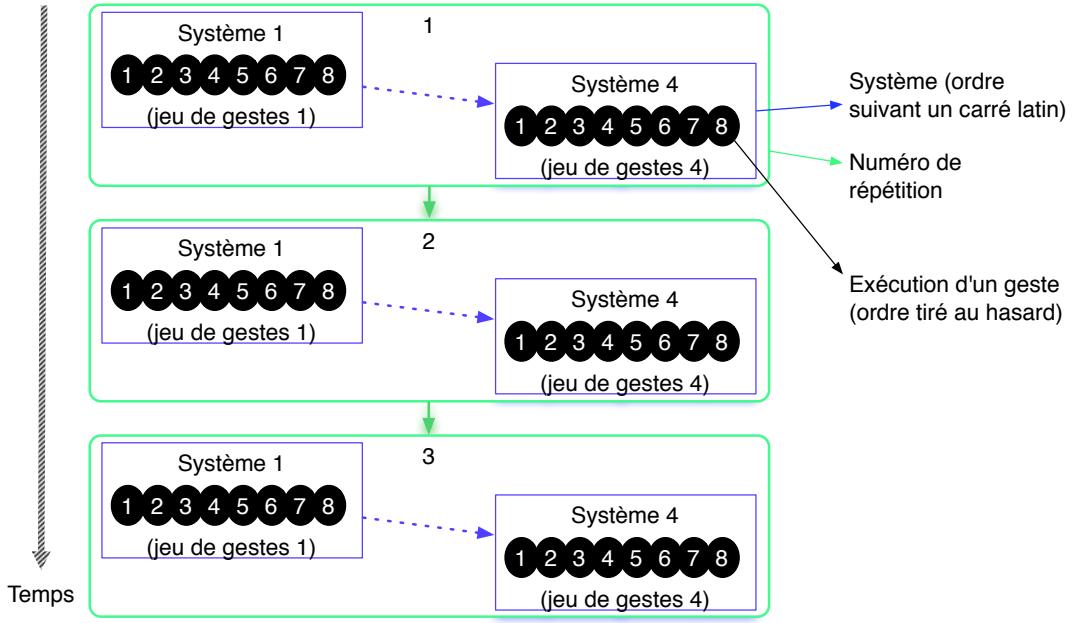


FIGURE 9.27 – Protocole d'expérimentation pour évaluer les options de retour d'information prédictif.

Comme pour l'évaluation expérimentale des options de conception du retour d'information rétroactif, nous nous plaçons dans une stratégie d'exploration. Aucune hypothèse n'est

faite sur les options de conception du retour d'information prédictif.

9.4.3.2 Résultats

Comme pour l'expérience 1, les principales mesures indépendantes étaient le taux de reconnaissance ainsi que le temps d'exécution obtenus avec chaque guide.

9.4.3.2.1 Taux de reconnaissance Une analyse de variance à mesures répétées n'a révélé aucun effet significatif de la portion affichée POR [$F(1.23) = 1.09$, $p > 0.05$] ou de la correction de l'origine du guide COR [$F(1.23) = 1.34$, $p > 0.05$] sur le taux de reconnaissance (Figure 9.28).

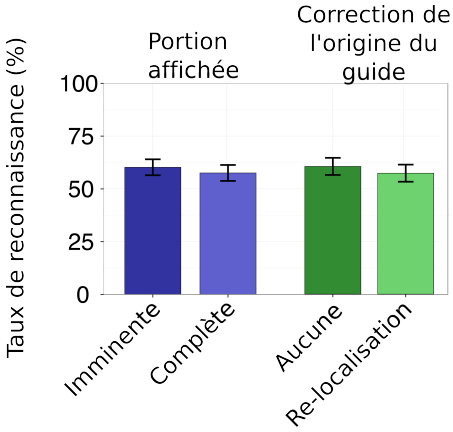


FIGURE 9.28 – Taux de reconnaissance (en pourcentage) obtenus en fonction des portions affichées et de la correction de l'origine du guide. Les barres d'erreur représentent les intervalles de confiance à 95%.

N'avoir aucun résultat significatif est étonnant. En effet, ces alternatives - que ce soit les portions affichées ou bien le déplacement du guide pendant l'exécution d'un geste - affecte grandement la complexité visuelle de la scène 3D. Cependant, nous ne trouvons aucun effet significatif de ces options de conception sur le taux de reconnaissance. Nous pouvons donc dire que si un effet existe, il est minime. En revanche, il n'est pas possible de dire si les options sont strictement équivalentes. En effet, avec 24 participants, la puissance statistique de nos tests n'est pas suffisante pour déclarer ces options comme équivalentes. En effet, la puissance de notre test est de 10% pour un petit effet, 40% pour un effet moyen et 77% pour un grand effet.

9.4.3.2.2 Temps d'exécution Une analyse de variance à mesures répétées révèle un effet significatif de la portion affichée POR sur le temps d'exécution d'un geste [$F(1.23) = 9.14$, $p < 0.01$], l'affichage d'une portion imminente menant à une exécution plus rapide que l'affichage du geste complet (Figure 9.29). Ce résultat montre que pour la rapidité

d'exécution, la simplicité de l'affichage prévaut sur la capacité d'anticipation allouée par l'affichage des gestes complets.
Nous n'avons trouvé aucun effet de la correction de l'origine du guide COR sur le temps d'exécution [F(1.23) = 2.86, p > 0.05].

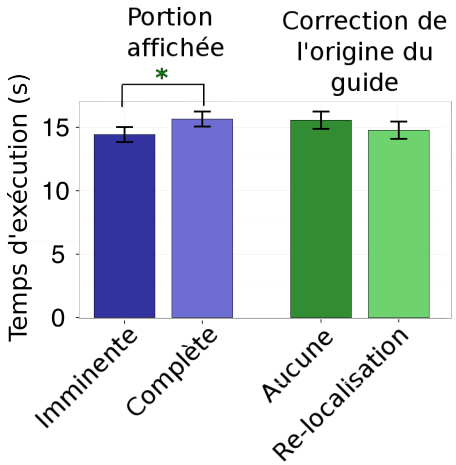


FIGURE 9.29 – Temps d'exécution (en secondes) obtenus en fonction des portions affichées et de la correction de l'origine du guide. Les barres d'erreur représentent les intervalles de confiance à 95%.

9.4.3.2.3 Discussion Puisque (1) nous n'avons trouvé aucun effet significatif sur le taux de reconnaissance, (2) afficher les parties imminentes à venir conduit à des temps d'exécution plus rapide qu'afficher la totalité du geste à compléter et (3), les préférences des utilisateurs ne font pas ressortir de technique particulière [$\chi^2(3) = 1.38$, $p > 0.05$], nous basons nos recommandations de conception sur le temps d'exécution. Il est préférable pour notre guide de ne montrer que les parties imminentes à venir des gestes.
Nous notons toutefois que ce résultat est valide pour les performances d'utilisateurs novices. Nous pouvons en effet questionner l'effet d'une représentation partielle sur l'apprentissage des gestes dans le cas où l'utilisateur n'aurait pas accès à du retour d'information rétroactif terminal. En effet, le fait de ne jamais visualiser le geste en entier pourrait négativement influencer la mémorisation de la forme du geste.

Résumé

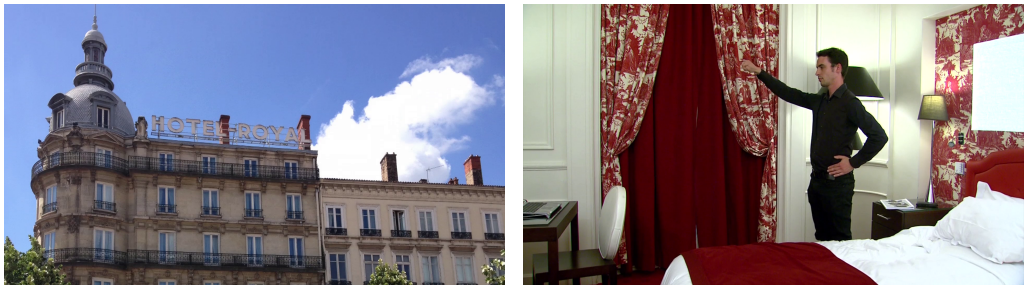
Nos expériences montrent que :

1. seul un retour d'information rétroactif concurrent a un impact bénéfique sur la reconnaissance des gestes. Cet impact est présent seulement pendant les deux premières répétitions. Nous supposons que ce retour d'information rétroactif concurrent n'aide pas les utilisateurs à se corriger, mais facilite l'exécution d'un geste en réduisant la complexité visuelle de la scène.
2. l'origine du guide déterminée par le retour d'information prédictif n'a pas

d'impact sur les performances d'exécution de geste. La stabilité globale de la scène n'est donc pas un critère critique du guidage.

3. les portions de gestes affichées par le retour d'information prédictif impactent la rapidité d'exécution des gestes. L'affichage des parties imminentes des gestes permet une exécution plus rapide que l'affichage des portions complètes, montrant ainsi que la complexité visuelle de la scène prévaut sur l'anticipation de la forme des gestes.

9.5 Déploiement *in situ*



(a) (b)

FIGURE 9.30 – Déploiement *in situ* à l'hôtel Royal de Lyon (a) vue de l'extérieur, (b) vue de l'intérieur. Un utilisateur exécute un geste pour contrôler la lumière de la chambre. La main sur la hanche indique que le geste est commencé.

Au delà des expérimentations en laboratoire, nous avons déployé OctoPocus3D en contexte réel, après les expériences contrôlées en laboratoire. Les deux démonstrateurs ont été installés dans un bureau et dans une chambre d'hôtel pour le contrôle de la lumière.
L'objectif du bureau était de tester l'ensemble du système - de son installation à son utilisation - dans un contexte hors laboratoire, permettant ainsi l'intégration des travaux de tous les partenaires du projet. Les sources de lumières devaient donc être installées et branchées sur la BOX. Le système installé devait pouvoir être contrôlé par des utilisateurs finaux. Quatre ambiances lumineuses étaient définies : travail, réunion, projection et éteint. Trois personnes ont pu tester OctoPocus3D dans ce contexte. Par manque de disponibilité des locaux, OctoPocus3D n'a pas été déployé sur le long terme. La version d'OctoPocus3D déployée utilisait un un retour d'information rétroactif concurrent et représentait les portions totales restantes des gestes avec une origine dynamique. De plus, cette version utilisait un mécanisme d'initialisation, c'est-à-dire le début d'exécution d'un geste (paragraphe 2.2.1.2), reprenant la forme de théière proposée par [Walter *et al.*, 2013] (Figure 9.30b). Cette initialisation était confondue avec le déclenchement du guidage (paragraphe 8.1.2.1).
La chambre d'hôtel était installée à l'hôtel Royal de Lyon (Figure 9.30a), pendant une

semaine. Plus d'une centaine de personnes ont visité la chambre, parmi eux des partenaires du projet, leurs collègues, des décideurs et des journalistes. Il ne s'agissait pas d'une étude utilisateur, mais de communication.

Bien que ces deux démonstrateurs n'aient pas mené à des résultats quantitatifs ou qualitatifs, nous avons pu (1) déployer et tester notre guide OctoPocus3D dans des environnements extérieurs au laboratoire et (2) le relier à une architecture finale attendue dans les spécifications du projet DELight, contrôlant ainsi des éclairages à LED via une communication avec la BOX.

9.6 Conclusion

Dans ce chapitre, nous avons décrit notre technique de guidage de gestes 3D. Cette technique, extension d'un guide 2D (OctoPocus), facilite l'utilisation de l'interaction gestuelle en proposant de montrer les commandes existantes ainsi que les gestes associés. OctoPocus3D présente visuellement ces gestes sous formes de tubes 3D affichés sur un écran distant. Au delà de la présentation visuelle des gestes par des tubes, plusieurs options de conception sont possibles pour les mécanismes de retour d'information rétroactif et de retour d'information prédictif. Toutefois, l'impact de ces différentes options n'avait pas été évalué pour le guidage d'utilisateurs novices. Il est par conséquent difficile de concevoir un guide sans savoir l'impact de ces options de conception sur l'utilisabilité. Nous avons donc effectué deux expériences pour enrichir les connaissances dans le domaine du guidage de plusieurs gestes représentés en 3D sur un écran distant.

Tout d'abord, nous avons montré qu'un retour d'information rétroactif concurrent simplifiant visuellement la scène 3D avait pour conséquence un meilleur taux de reconnaissance comparé à une technique ne fournissant aucun retour d'information rétroactif. La technique de guidage ainsi obtenue, OctoPocus3D, peut être utilisée sans phase d'entrainement au préalable qui inclurait un retour d'information rétroactif terminal, c'est-à-dire sans aucune session pendant laquelle la technique montrerait le geste effectué ainsi que le geste à réaliser pour permettre à l'utilisateur de les comparer. De plus, de précédents travaux ont montré l'influence négative d'un retour d'information rétroactif concurrent sur l'apprentissage moteur [Anderson et Bischof, 2013]. Nous ajoutons à ce fait connu qu'un retour d'information rétroactif concurrent n'a d'effet que pendant la prise en main de la technique. Il est donc tout-à-fait envisageable de supprimer ce retour d'information rétroactif de manière à s'adapter à l'expertise de l'utilisateur.

Ensuite, nous avons montré qu'un affichage ne montrant que la portion imminente à venir des gestes permettait une exécution plus rapide qu'un affichage montrant la totalité restante des gestes. En effet, même si cette option empêche les utilisateurs d'appréhender la forme globale des gestes, cette option présente l'avantage de réduire la complexité visuelle de la scène 3D, une propriété essentielle dans le cadre du guidage d'un ensemble

de gestes 3D.
Il convient de dire que la stabilité de l'affichage n'a eu aucun effet, que ce soit sur le taux de reconnaissance ou bien le temps d'exécution. Une hypothèse serait que la stabilité globale de la scène n'impacte pas les performances puisque les utilisateurs se concentrent localement sur la partie autour de la représentation de leur main. Cette hypothèse concorde avec le fait que les utilisateurs ne profitent pas de l'affichage complet des gestes. Il faudrait toutefois réaliser d'autres expériences afin d'analyser en détail cette hypothèse.

Enfin nous avons déployé la technique en contexte réel en considérant le contrôle de lumières dans un bureau et une chambre d'hôtel. Pour ceci nous avons inséré la technique OctoPocus3D au sein d'une architecture globale incluant une BOX pour le contrôle des lumières LED dans une pièce.

9.7 Travaux futurs

Cette étude peut être étendue selon deux directions complémentaires de la modalité de sortie pour présenter les gestes : la première dédiée au langage d'interaction L de la modalité de sortie, la seconde dédiée au dispositif d de sortie.

Tout d'abord, nous pourrions explorer plus en détail la manière de représenter les gestes avec OctoPocus3D. Comme mentionné précédemment (paragraphe 9.3.1.2.2), nous pourrions explorer plus en détail l'utilisation de formes géométriques autre que des tubes pour cette représentation. OctoPocus3D pourrait aussi considérer la dynamicité de la représentation graphique [Bernsen, 1994]. Par exemple, l'apparition des gestes pourrait bénéficier d'une animation lorsque le guidage est déclenché. Un autre exemple serait une rotation automatique autour de la scène 3D en fin d'exécution de mouvement. Ces exemples peuvent potentiellement ajouter de la perception spatiale grâce à la profondeur cinétique (information de profondeur obtenue grâce au mouvement) sans demander d'effort supplémentaire de la part de l'utilisateur.
Finalement, il est envisageable d'explorer différents dispositifs de sortie pour afficher les gestes. En considérant l'axe *perspective* de notre espace de conception (Figure 9.31), nous pouvons concevoir différentes installations pour afficher OctoPocus3D selon un point de vue égocentrique (comme lors d'une superposition du guide sur la main de l'utilisateur), couplé (comme lors d'une projection du guide sur un miroir semi-transparent comme le fait YouMove [Anderson *et al.*, 2013]) ou bien exocentrique (comme lors de l'affichage du guide sur un écran distant sans couplage tête-caméra).

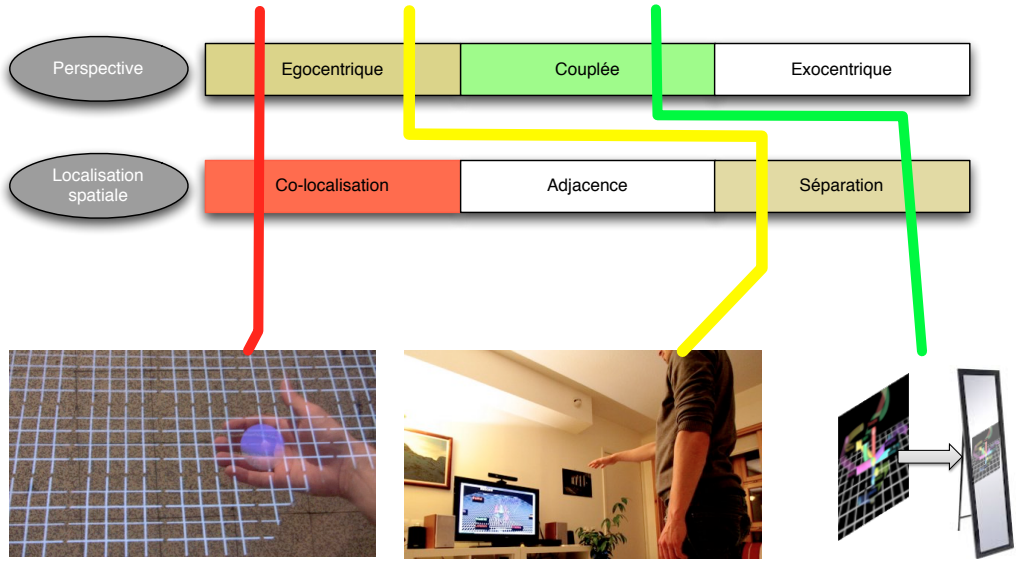


FIGURE 9.31 – Exploration de différents dispositifs de sortie afin d'explorer les axes *perspective* et *localisation spatiale* de notre espace de conception.

10 Conclusion de la partie contrôle

Sommaire

10.1 Conclusion sur le guidage **167**

10.2 Travaux futurs sur le guidage **168**

10.1 Conclusion sur le guidage

Nous avons présenté nos travaux pour contrôler des objets physiques augmentés par le geste. Plus précisément, nous avons étudié le guidage des gestes nécessaire lorsque l'utilisateur ne connait pas les commandes disponibles ainsi que les gestes associés à ces commandes.

Nous avons décrit un espace de conception (chapitre 8) pour caractériser les techniques de guidage en considérant deux sources d'information : le retour d'information rétroactif et le retour d'information prédictif (paragraphe 8.1.1.1). Ces deux catégories d'information sont elles-mêmes composées de quatre groupes d'axes suivant quatre questions : *Quand*, *Quoi*, *Comment* et *Où* (paragraphe 8.1.1.2). L'espace de conception comprend 35 axes (paragraphe 8.1) et peut décrire les 60 étapes de guidages existantes.
Afin de faciliter son utilisation, nous avons présenté un outil disponible en ligne (paragraphe 8.2.1). Nous avons montré via des scénarios comment utiliser cet outil afin de valider notre espace de conception et pouvoir décrire, comparer et concevoir des techniques de guidage (paragraphe 8.2.2).

Nous avons illustré le pouvoir génératif de notre espace de conception par la création d'une technique de guidage : OctoPocus3D (paragraphe 9.3). Deux évaluations expérimentales ont permis d'explorer des options de conception issues de notre espace (paragraphe 9.2.1) ainsi que leurs alternatives au regard de nos deux catégories d'information : le retour d'information rétroactif et le retour d'information prédictif.

Retour d'information rétroactif (paragraphe 9.4.2)
Nous avons comparé deux stratégies pour aider l'utilisateur à se corriger en fonction de la fréquence de mise à jour du retour d'information rétroactif. Cette mise à jour peut être continue, c'est-à-dire pendant l'exécution du geste (*concurrent*), ou bien discrète, c'est-à-dire à la fin de l'exécution du geste par exemple (*terminal*). Le retour d'information rétroactif concurrent permettait aux utilisateurs de savoir comment le système interprétait leurs gestes en temps réel alors que le retour d'information rétroactif terminal leur permettait de comparer le geste effectué avec le geste à réaliser.
Les résultats montrent que :

1. seul un retour d'information rétroactif concurrent permet aux utilisateurs d'obtenir de meilleurs scores de reconnaissance, et ce pendant les deux premières répétitions seulement. Ce résultat consolide les travaux existants qui montrent que ce retour d'information rétroactif concurrent a un impact négatif sur l'apprentissage moteur des gestes [Anderson et Bischof, 2013].
2. les utilisateurs n'ont pas conscience d'utiliser le retour d'information rétroactif concurrent et préfèrent un retour d'information rétroactif terminal.

Retour d'information prédictif (paragraphe 9.4.3)
Nous avons comparé deux axes chacun reliés à une difficulté de la représentation visuelle de nos gestes. Le premier axe concerne les *portions affichées* de gestes et influence la complexité visuelle de la scène 3D. Le second axe concerne la *localisation du guide* et influence la stabilité visuelle de la scène 3D.
Les résultats montrent que :

1. afficher des portions imminentes des gestes permet une exécution plus rapide qu'un affichage des portions complètes. Cela montre que la difficulté de compréhension de la scène prévaut sur la faculté d'anticiper les gestes.
2. la stabilité globale de la scène n'a aucun impact sur les performances des utilisateurs.

Au delà des expérimentations en laboratoire, nous avons aussi démontré l'intégration de notre technique OctoPocus3D en contexte réel pour le contrôle de lumières, dans un bureau et une chambre d'hôtel. Pour cela, la technique a été intégrée dans une architecture globale incluant une Box centralisant le contrôle des lumières LEDs et classiques au sein d'une pièce.

10.2 Travaux futurs sur le guidage

Nous envisageons d'effectuer d'autres itérations entre l'étude de l'espace de conception et l'étude de techniques de guidage. Comme pour la désambigüisation, la création de nouvelles techniques de guidage peut amener à l'ajout d'options de conception dans notre espace pouvant elles-mêmes mener à la création de nouvelles techniques.

Plusieurs pistes, telles que l'exploration de différents dispositifs d'affichage, ont été propo-

sées pour continuer l'exploration de notre espace grâce à OctoPocus3D (paragraphe 5.5). Mais nous pourrions nous intéresser aussi à des sujets connexes, tels que l'apprentissage des gestes ou bien le guidage de gestes autres que symboliques dynamiques (paragraphe 2.2) :

— Les évaluations expérimentales effectuées concernent les utilisateurs novices. Il serait intéressant d'étudier comment OctoPocus3D facilite l'apprentissage des gestes. Par exemple, ne représenter que des portions imminentes à venir des gestes permet de meilleurs temps d'exécution que des portions complètes. Toutefois, dans le cas où la technique ne fournit pas de retour d'information rétroactif terminal représentant tout le geste, l'utilisateur n'aurait jamais de vision globale de la forme géométrique. Cette option pourrait potentiellement influencer la mémorisation et l'apprentissage des gestes. Cette question de recherche est à explorer.

— OctoPocus3D guide des gestes symboliques dynamiques 3D. Il serait intéressant d'explorer le guidage d'autres types de gestes tels que des gestes symboliques statiques.

  De plus, OctoPocus3D ne guide que la forme géométrique de nos gestes symboliques dynamiques 3D. Il serait intéressant d'explorer des pistes pour (1) utiliser d'autres caractéristiques des gestes, telle que la vitesse ou l'échelle par exemple et (2) guider cette information supplémentaire. En effet, la commande pourrait alors bénéficier d'un paramètre : allumer avec une intensité de 50% par exemple. Le défi de recherche serait alors :

  - le choix de la caractéristique du geste que l'on pourrait utiliser (la pression, la vitesse, l'échelle, etc). Il faudrait aussi
  - chercher comment utiliser cette caractéristique de manière optimale. Nous pouvons envisager de relier la caractéristique à un paramètre de commande de façon continue ou bien discrète [Vatavu *et al.*, 2013].
  - la représentation graphique de la caractéristique ainsi que du paramètre de la commande.

Enfin, maintenant qu'une technique de guidage d'un ensemble de gestes 3D existe, nous pouvons envisager d'étudier des aspects de l'interaction gestuelle 3D jusqu'ici inexplorés. Par exemple, nous pourrions étudier ce qui rend un jeu de gestes difficile en caractérisant les différents types de difficulté de l'interaction gestuelle : la difficulté d'exécution, la difficulté de guidage, la difficulté de reconnaissance ou encore la difficulté ressentie comme pour les gestes 2D [Vatavu *et al.*, 2011]. Par exemple, une simple ligne droite est facile à exécuter, mais peut être difficile à guider si elle est orientée dans le sens de la caméra et cachée par d'autres gestes. Au contraire, un geste difficile à exécuter peut être facilement guider s'il se trouve seul et dans un plan orthogonal au point de vue de la caméra. Ces deux difficultés (exécution et guidage) sont encore différentes de la difficulté à faire reconnaître un geste par le système ou bien la difficulté ressentie par l'utilisateur. Nous envisageons d'étudier ces types de difficultés en utilisant OctoPocus3D.

11 Conclusion

Sommaire

11.1 Résumé des contributions 171
11.1.1 Sélection d'objets physiques augmentés 172
11.1.2 Contrôle d'objets physiques augmentés 173
11.2 Perspectives envisagées des travaux de thèse 174

Nous rappelons les contributions et résultats expérimentaux de nos travaux de thèse puis définissons des perspectives qui découlent de ces travaux.

Les travaux présentés concernent l'interaction à distance avec des objets physiques augmentés dans un contexte d'intelligence ambiante (chapitre 2). Nous avons exploré l'utilisation de l'interaction gestuelle (partie I et partie II) afin de répondre à ce besoin interactionnel. Pour cela, nous avons distingué deux sous-tâches de l'interaction : la sélection et le contrôle.

Pour conclure ce manuscrit, nous rappelons les contributions de nos travaux de thèse (paragraphe 11.1). En accord avec nos objectifs initiaux, ces contributions concernent les aspects conceptuels et pratiques de la sélection et du contrôle d'objets physiques augmentés. Nous présentons ensuite des perspectives de recherche pour étendre nos travaux (paragraphe 11.2).

11.1 Résumé des contributions

Nos travaux sont dédiés à la sélection et au guidage du contrôle d'objets physiques augmentés. Pour chacune de ces deux sous-tâches, nous proposons tout d'abord un espace de conception pour organiser les options de conception qui s'offrent aux concepteurs. Nous proposons et étudions ensuite des techniques d'interaction pour les utilisateurs finaux.

11.1.1 Sélection d'objets physiques augmentés

Pour la sélection des objets physiques augmentés, nous nous sommes intéressés aux gestes déictiques, c'est-à-dire aux gestes de pointage. Plus précisément, à l'étape de désambigüisation nécessaire lorsque plusieurs objets sont pointés par un volume de sélection (partie I). Une première contribution est conceptuelle avec l'établissement d'un espace de conception. Nous apportons aussi des contributions pratiques sous la forme de deux techniques de désambigüisation.

Notre espace de conception (chapitre 4) adopte deux points de vue pour décrire les techniques de désambigüisation : le point de vue de l'*Interaction* et le point de vue du *Système de Désambigüisation*. Le premier point de vue, l'*Interaction*, est composé de deux axes permettant de capturer les changements en terme d'interaction entre l'étape de pointage absolu et celle de désambigüisation. Le second point de vue, le *Système de Désambigüisation*, est composé de 9 axes pour caractériser ce que le système doit connaître à propos du monde physique. Plus particulièrement sur la nature des objets (groupe *Quoi*) et leur localisation (groupe *Où*). Nous montrons le pouvoir descriptif de cet espace en décrivant et classant 7 techniques existantes puis nous soulignons son pouvoir génératif en concevant deux nouvelles techniques de désambigüisation par exploration des options de conception du point de vue *Interaction*.

Nos deux techniques de désambigüisation, P2Roll (Physical Pointing Roll) et P2Slide (Physical Pointing Slide) maintiennent l'attention de l'utilisateur dans le monde physique, sur les objets eux-mêmes, que ce soit pour l'étape de pointage ou pour celle de désambigüisation (chapitre 5). P2Roll utilise un roulis du poignet (entrée relative au corps) pour naviguer parmi les objets pré-sélectionnés par le pointage. P2Slide utilise un mouvement horizontal du pouce sur un écran tactile de téléphone (entrée relative à un dispositif).
Nous avons évalué ces deux techniques en les comparant à une technique existante affichant les noms des objets sur l'écran d'un téléphone, forçant ainsi l'utilisateur à changer son attention des objets physiques eux-mêmes à une représentation numérique de ces objets. Les résultats expérimentaux montrent que P2Roll et P2Slide sont plus rapides que la liste. Toutefois, P2Roll et P2Slide manquent de précision pour le cas de fortes densités d'objets proches dans la pièce, c'est-à-dire avec plus de 8 objets sélectionnés par le pointage. Néanmoins, cette situation est aujourd'hui rare dans un contexte domestique tel que la maison. Les participants ont toutefois montré une nette préférence pour les techniques conservant l'attention dans le monde physique.

Selon notre démarche de travail, après avoir étudié la sélection par un pointage volumique suivi d'une désambigüisation pour l'étape de sélection, nous avons étudié l'étape suivante, à savoir l'étape de contrôle.

11.1.2 Contrôle d'objets physiques augmentés

Pour le contrôle d'objets physiques augmentés, nous avons considéré un autre type de gestes : des gestes 3D symboliques dynamiques. Plus précisément, nous nous sommes intéressés au guidage de ce type de gestes nécessaire pour révéler aux utilisateurs novices d'une part (1) les commandes disponibles et d'autre part (2) les gestes associés (partie II). Comme pour les travaux ci-dessus sur la sélection, nos contributions sont à la fois conceptuelles avec la création d'un espace de conception et pratiques avec la conception et l'évaluation expérimentale d'une nouvelle technique de guidage.

Notre espace de conception (chapitre 8) considère deux catégories d'information fournies par une technique de guidage pour caractériser le comportement d'un guide : le retour d'information rétroactif (feedback) – concernant des actions passées- et le retour d'information prédictif (feedforward)- concernant des actions futures. Chaque catégorie est composée de quatre groupes d'axes pour répondre à quatre questions : *Quand*, *Quoi*, *Comment* et *Où*. L'espace de conception est organisé selon un total de 35 axes et caractérise 60 étapes de guidage distinctes, pour tous les types de gestes définis au paragraphe 2.2. Afin d'exploiter pleinement cet espace, nous proposons un outil disponible en ligne permettant de décrire, comparer et créer des techniques de guidage (paragraphe 8.2). Nous exposons par des scénarios comment différents concepteurs peuvent utiliser notre espace ainsi que la base de connaissance qui lui est rattachée. Nous montrons le pouvoir génératif de notre espace de conception avec la création d'un guide de gestes 3D symboliques dynamiques.

Notre guide, OctoPocus3D, représente les gestes par des tubes 3D affichés sur un écran distant (chapitre 9). OctoPocus3D étend le concept originalement proposé par OctoPocus [Bau et Mackay, 2008] pour guider des gestes 2D. Nous utilisons notre espace afin d'explorer des options de conception proposées pour des gestes 2D que nous voulons évaluer dans le cas de gestes 3D en l'air.

- La fréquence de mise à jour du retour d'information rétroactif : celui-ci peut être mis à jour en continu (retour d'information rétroactif concurrent) ou bien de manière discrète, comme à la fin de l'exécution d'un geste (retour d'information rétroactif terminal). Le retour d'information rétroactif concurrent vise à aider l'utilisateur à se corriger pendant l'exécution d'un geste en montrant les scores de reconnaissance courants par l'intermédiaire du rayon des tubes 3D. Le retour d'information rétroactif terminal quant à lui permet à l'utilisateur de se corriger pendant une phase d'entrainement en montrant le geste effectué comparé au geste à exécuter.
- Les portions de gestes affichées par le retour d'information prédictif : le guide peut afficher les portions complètes restantes des gestes pour aider l'utilisateur à appréhender la totalité de la forme géométrique du geste ou bien seulement des portions imminentes à venir pour faciliter la compréhension de la totalité de la scène 3D. Cette option de conception affecte donc la complexité visuelle de la scène 3D et la capacité d'anticipation de la forme des gestes.

— L'origine du guide définie par le retour d'information prédictif : celle-ci peut être fixe ou bien localisée sur la représentation numérique de la main de l'utilisateur à tout instant. Cette option de conception affecte donc la stabilité visuelle de la scène 3D.

Pour le comportement du retour d'information rétroactif, les résultats montrent que seul le retour d'information rétroactif concurrent a un effet positif sur les taux de reconnaissance obtenus, mais seulement pendant les deux premières répétitions, après quoi les utilisateurs peuvent utiliser OctoPocus3D avec ou sans retour d'information rétroactif. Ces derniers montrent toutefois une nette préférence pour le retour d'information rétroactif terminal. Pour le comportement du retour d'information prédictif, les résultats montrent que seules les portions affichées affectent la rapidité du temps d'exécution des gestes : l'affichage des portions imminentes permet une exécution plus rapide que l'affichage des portions complètes, montrant ainsi que la complexité visuelle prévaut sur la capacité d'appréhension de la forme des gestes.

Nos différentes contributions ainsi que nos résultats expérimentaux permettent d'envisager différentes perspectives pour des futurs travaux.

11.2 Perspectives envisagées des travaux de thèse

Nous identifions plusieurs perspectives et pistes de recherche aux travaux présentés dans ce manuscrit. Nous déterminons quatre niveaux d'abstraction pour ces perspectives (Figure 11.1) :

— *Niveau 1* : niveau des contributions pratiques.
Ce niveau constitue des extensions des techniques d'interaction conçues (Figure 11.1). Ces extensions font suite aux évaluations expérimentales menées des techniques (Figure 11.2). Nous avons listé au paragraphe 5.5 des pistes de recherche pour améliorer nos techniques de désambigüisation P2Roll et P2Slide au regard de la précision dans de fortes densités. Par exemple, nous pourrions découpler le mécanisme de validation (retirer le pouce de l'écran du téléphone) du mécanisme de sélection afin d'éviter des erreurs au dernier moment. Notre technique de guidage OctoPocus3D (chapitre 9) peut être étendue en explorant d'autres dispositifs de sortie que l'écran distant. Nous avons juste commencé l'implémentation de dispositifs de sortie égocentrique et couplé en projetant OctoPocus3D sur différents supports semi-réfléchissant afin d'évaluer l'impact du point de vue sur le guidage. Ces pistes de recherche sont des extensions directes de nos contributions pratiques.

— *Niveau 2* : niveau des sous-tâches de sélection et du contrôle.
Nous avons listé des pistes de recherche pour améliorer ou bien étendre nos travaux dans de nouvelles directions au regard :
  - D'une part de la désambigüisation (paragraphe 6.2). Par exemple une piste

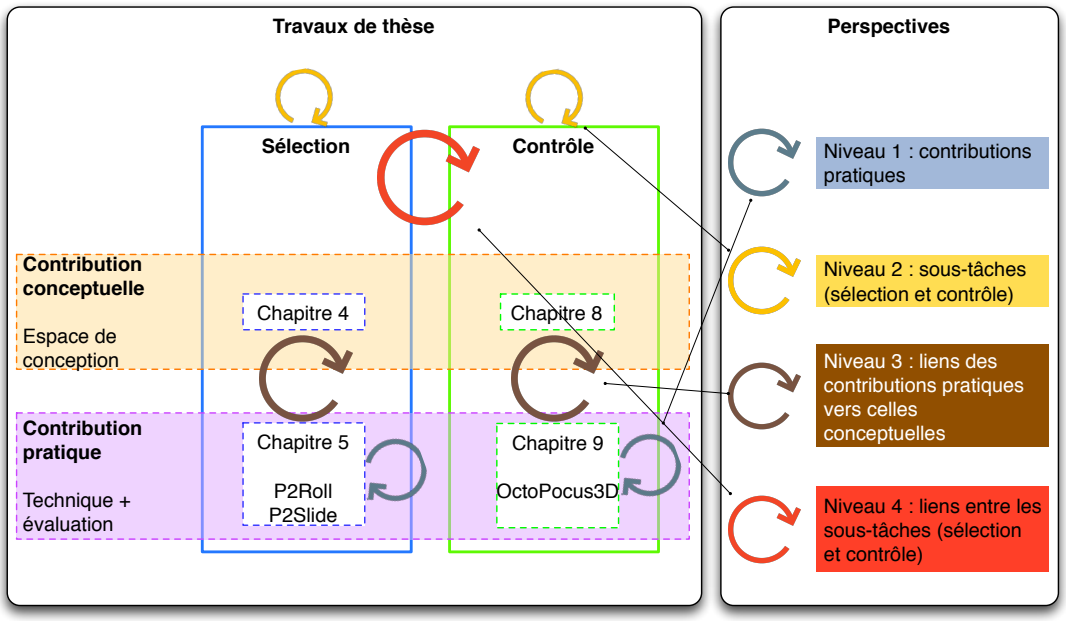


FIGURE 11.1 – Les différents niveaux d'extension à nos travaux de thèse. 1) Niveau de chaque contribution pratique. 2) Niveau de chaque sous-tâche – sélection et contrôle. 3) Niveau pour faire le lien des contributions pratiques vers celles conceptuelles. 4) Niveau global pour relier les sous-tâches de sélection et de contrôle.

pertinente de travaux est d'étudier le comportement des utilisateurs en contexte réel. Pour cet axe de recherche, il convient d'évaluer en situation réelle différentes densités et tailles d'objets (1) pour établir des limites quant au choix de la solution de pointage (rayon ou bien volume) (2) pour identifier la densité maximale à partir laquelle l'utilisateur préfère se rapprocher et ainsi réduire le nombre d'objets dans le volume de sélection plutôt qu'essayer d'être précis.

- D'autre part du guidage de gestes (paragraphe 10.2). Par exemple il est prometteur d'étudier le guidage de d'autres types de gestes qui ne soient pas dynamiques symboliques et de comprendre les différences dans les options de conception. Nous avons juste commencé à étudier cet aspect en considérant le passage d'un paramètre lors d'une commande gestuelle. En s'inspirant des Control Menus [Pook *et al.*, 2000], l'utilisateur pourrait effectuer un geste pour allumer la lumière (qui soit guidé par exemple avec OctoPocus3D) puis un geste vertical continu pour contrôler l'intensité lumineuse.

— *Niveau 3* : niveau reliant les contributions pratiques vers celles conceptuelles (Figure 11.2).
Nous avons construit nos espaces de conception de façon itérative à partir de l'existant, puis nous avons conçu et évaluer des techniques en exploitant le pouvoir génératif de nos espaces. Il convient maintenant d'établir le lien inverse de nos

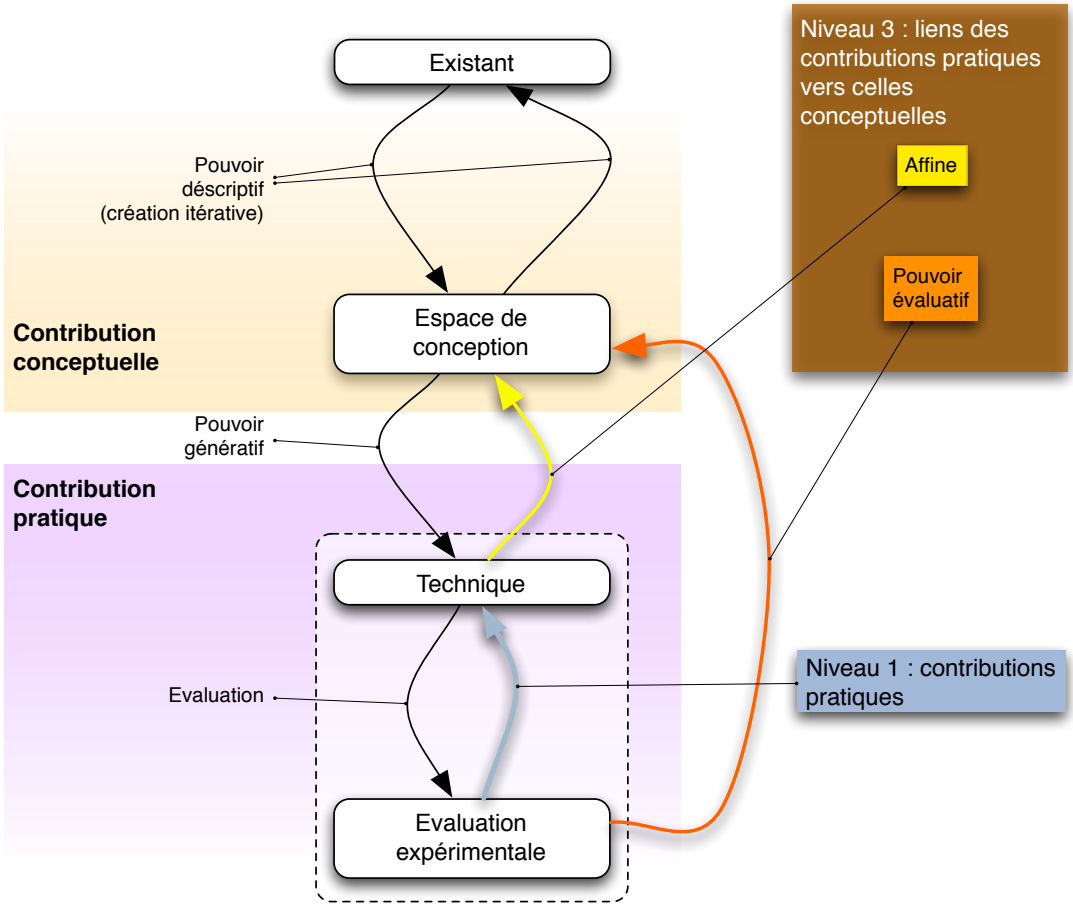


FIGURE 11.2 – Détail du niveau 3 avec le lien pouvant mener à l'enrichissement de nos espaces de conception (*affine*) et le lien permettant l'établissement du *pouvoir évaluatif* de nos espaces grâce aux résultats expérimentaux. Le schéma montre aussi le niveau 1 permettant les améliorations de nos techniques à partir des évaluations expérimentales.

techniques vers nos espaces de conception permettant éventuellement d'enrichir nos espaces. De plus établir le lien de nos évaluations expérimentales vers nos espaces permettra l'établissement du pouvoir évaluatif (guide pour choisir les options de conception) de nos espaces de conception.

— *Niveau 4* : niveau global, à savoir le lien entre la partie sélection et la partie contrôle.
Nous avons itéré sur chacune des parties (sélection et contrôle) en les considérant de façon indépendante dans une première étape afin de décomposer la question de recherche initiale sur l'interaction gestuelle 3D avec des objets physiques. Il convient maintenant de considérer leur couplage. Une future étape consiste à relier les deux sous-tâches de sélection et de contrôle et en particulier de travailler sur la transition entre les deux sous-tâches toutes deux exploitant des gestes 3D. Une piste

de recherche envisagée est l'exploration de l'utilisation du plafond comme support du guidage. Le plafond offre une large surface, souvent blanche, inutilisée et dégagée afin d'afficher des informations numériques dans le monde physique. Cette piste offre de nombreuses opportunités d'investigation. En effet, cela nous permettrait (1) d'explorer des solutions techniques pour afficher le guide au plafond en fonction de l'objet sélectionné, (2) d'explorer un guidage laissant l'attention de l'utilisateur dans le monde physique pendant l'étape de contrôle pour le cas de commandes avec un résultat ambiant – visuel ou sonore par exemple - , (3) d'explorer l'utilisabilité du plafond comme surface d'affichage pour l'interaction au quotidien dans un contexte d'intelligence ambiante ou encore (4) d'explorer l'influence d'une telle perspective et localisation (axes de notre espace de conception) durant le guidage sur les performances de l'utilisateur.

Enfin, il serait intéressant d'étendre ces travaux à une interaction sortant du schéma classique sélection puis contrôle d'un objet précis. En effet, sur le continuum existant (Figure 11.3) entre un habitat traditionnel et un habitat totalement intelligent, c'est-à-dire dans lequel l'environnement serait en mesure de répondre aux besoins de l'utilisateur sans intervention explicite de sa part, nous nous sommes intéressés au premier degré d'interaction : le contrôle d'un objet de l'environnement physique. Ce paradigme d'interaction étudié dans nos travaux est envisageable lorsque l'objet interactif est aussi l'objet modifié par le contrôle, comme lorsque l'utilisateur souhaite allumer sa télévision. Une extension proposée à ce paradigme est le contrôle de lieux composés de plusieurs objets en vue d'un but de haut niveau. Par exemple, nos travaux pourraient être étendus pour que lorsque l'utilisateur sélectionne le canapé et effectue un geste 3D de la main pour la commande "lecture", alors les ambiances lumineuses et sonores s'adaptent en contrôlant plusieurs objets tels que des sources lumineuses et des haut-parleurs. Cette extension permettrait (1) de contrôler finement les objets interactifs grâce à nos travaux et (2), de contrôler globalement plusieurs objets en fonction d'un but de haut niveau.

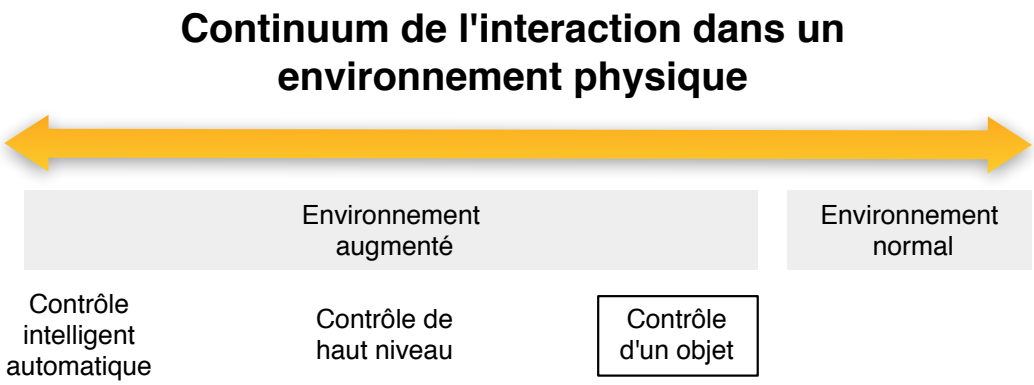


FIGURE 11.3 – Continuum de l'interaction dans un environnement physique.

Bibliographie

[Aigner *et al.*, 2012] Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindlbauer, D., Ion, A., Zhao, S. et Koh, J. T. K. V. (2012). Understanding Mid-Air Hand Gestures : A Study of Human Preferences in Usage of Gesture Types for HCI. Rapport technique.

[Ailisto *et al.*, 2006] Ailisto, H., Pohjanheimo, L., Välkkynen, P., Strömmer, E., Tuomisto, T. et Korhonen, I. (2006). Bridging the physical and virtual worlds by local connectivity-based physical selection. *Personal and Ubiquitous Computing*, 10(6):333–344.

[Alexander *et al.*, 2012] Alexander, J., Han, T., Judd, W., Irani, P. et Subramanian, S. (2012). Putting your best foot forward : investigating real-world mappings for foot-based gestures. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, pages 1229–1238.

[Aliakseyeu *et al.*, 2012] Aliakseyeu, D., Meerbeek, B., Mason, J., van Essen, H., Offermans, S., Wiethoff, A., Streitz, N. et Lucero, A. (2012). Designing interactive lighting. *In Proceedings of the Designing Interactive Systems Conference on - DIS '12*, pages 801–802, New York, New York, USA. ACM Press.

[Anderson et Bischof, 2013] Anderson, F. et Bischof, W. F. (2013). Learning and performance with gesture guides. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 1109–1118, New York, New York, USA. ACM Press.

[Anderson *et al.*, 2013] Anderson, F., Grossman, T., Matejka, J. et Fitzmaurice, G. (2013). YouMove : Enhancing Movement Training with an Augmented Reality Mirror. *In Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 311–320, New York, New York, USA. ACM Press.

[Argelaguet et Andujar, 2013] Argelaguet, F. et Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers and Graphics*, 37(3):121–136.

[Asano *et al.*, 2005] Asano, T., Sharlin, E., Kitamura, Y., Takashima, K. et Kishino, F. (2005). Predictive interaction using the delphian desktop. *In Proceedings of the 18th annual ACM symposium on User interface software and technology - UIST '05*, pages 133 – 141, New York, New York, USA. ACM Press.

[Bacim *et al.*, 2013] Bacim, F., Kopper, R. et Bowman, D. a. (2013). Design and evaluation of 3D selection techniques based on progressive refinement. *International Journal of Human-Computer Studies*, 71(7-8):785–802.

[Balakrishnan, 2004] Balakrishnan, R. (2004). "Beating" Fitts' law : virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies*, 61(6):857–874.

[Balakrishnan et MacKenzie, 1997] Balakrishnan, R. et MacKenzie, I. S. (1997). Performance differences in the fingers, wrist, and forearm in computer input control. *In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, pages 303–310, New York, New York, USA. ACM Press.

[Bau et Mackay, 2008] Bau, O. et Mackay, W. E. (2008). OctoPocus : A Dynamic Guide for Learning Gesture-Based Command Sets. *In Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*, pages 37–46, New York, New York, USA. ACM Press.

[Beaudouin-Lafon, 2004] Beaudouin-Lafon, M. (2004). Designing interaction, not interfaces. *In Proceedings of the working conference on Advanced visual interfaces - AVI '04*, pages 15–22, New York, New York, USA. ACM Press.

[Bennett *et al.*, 2011] Bennett, M., McCarthy, K., O'Modhrain, S. et Smyth, B. (2011). *Human-Computer Interaction – INTERACT 2011*, volume 6946 de *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Berard et Rochet-capellan, 2012] Berard, F. et Rochet-capellan, A. (2012). Measuring the Linear and Rotational User Precision in Touch Pointing. *In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS '12*, pages 183–192, New York, New York, USA. ACM Press.

[Bernsen, 1994] Bernsen, N. O. (1994). A revised generation of the taxonomy of output modalities.

[Bernsen *et al.*, 1998] Bernsen, N. O., Dybkjær, H. et Dybkjær, L. (1998). *Designing interactive speech systems : from first ideas to user testing*. Springer Science & Business Media.

[Blanch *et al.*, 2004] Blanch, R., Guiard, Y. et Beaudouin-Lafon, M. (2004). Semantic Pointing : Improving Target Acquisition with Control-Display Ratio Adaptation. *In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '04*, volume 6, pages 519–526, New York, New York, USA. ACM Press.

[Blanch et Ortega, 2009] Blanch, R. et Ortega, M. (2009). Rake Cursor : Improving Pointing Performance with Concurrent Input Channels. *In Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, pages 1415–1418, New York, New York, USA. ACM Press.

[Bowman *et al.*, 2001] Bowman, D., Wingrave, C., Campbell, J. et Ly, V. (2001). Using pinch gloves for both natural and abstract interaction techniques in virtual environments. Rapport technique 0106.

[Bragdon *et al.*, 2009] Bragdon, A., Zeleznik, R., Williamson, B., Miller, T. et LaViola, J. J. (2009). GestureBar : Improving the Approachability of Gesture-based Interfaces. *In Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, pages 2269–2278, New York, New York, USA. ACM Press.

[Cadoz, 1994] Cadoz, C. (1994). Le geste canal de communication homme/machine : la communication 'instrumentale'. *Technique et Science Informatiques*, 13(1):31–61.

[Card *et al.*, 1991] Card, S. K., Mackinlay, J. D. et Robertson, G. G. (1991). A morphological analysis of the design space of input devices.

[Casiez *et al.*, 2012] Casiez, G., Roussel, N. et Vogel, D. (2012). 1e Filter : A Simple Speed-based Low-pass Filter for Noisy Input in Interactive Systems. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, pages 2527–2530.

[Chavez *et al.*, 2008] Chavez, F. d. l. O., de Vega, F. F., Olague, G., Montero, J. L. et Chávez, F. d. l. O. (2008). An independent and non-intrusive laser pointer environment control device system. *In Proceedings of the 5th international conference on Pervasive services - ICPS '08*, pages 37–46, New York, New York, USA. ACM Press.

[Cipiloglu *et al.*, 2010] Cipiloglu, Z., Bulbul, A. et Capin, T. (2010). A framework for enhancing depth perception in computer graphics. *In Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization - APGV '10*, volume 1, pages 141–148, New York, New York, USA. ACM Press.

[Cockburn *et al.*, 2011] Cockburn, a., Quinn, P., Gutwin, C., Ramos, G. et Looser, J. (2011). Air pointing : Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human Computer Studies*, 69(6):401–414.

[Cohé et Hachet, 2012] Cohé, A. et Hachet, M. (2012). Beyond the mouse : Understanding user gestures for manipulating 3D objects from touchscreen inputs. *Computers & Graphics*, 36(8):1119–1131.

[Colaço *et al.*, 2013] Colaço, A., Kirmani, A., Yang, H. S., Gong, N.-W., Schmandt, C. et Goyal, V. K. (2013). Mime : Compact, Low-Power 3D Gesture Sensing for Interaction with Head-Mounted Displays. *In Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 227–236, New York, New York, USA. ACM Press.

[Coutaz et Crowley, 2008] Coutaz, J. et Crowley, J. (2008). Plan "intelligence ambiante" : défis et opportunités. *Document de réflexion conjoint du comité d'experts "Informatique Ambiante" du département ST2I du CNRS et du Groupe de Travail "Intelligence Ambiante" du Groupe de Concertation Sectoriel "Informatique Ambiante" du département ST2I du CNRS et du Groupe d*, 1.

[Crossan *et al.*, 2008] Crossan, A., Williamson, J., Brewster, S. et Murray-Smith, R. (2008). Wrist rotation for interaction in mobile contexts. *In Proceedings of the*

10th international conference on Human computer interaction with mobile devices and services - MobileHCI '08, pages 435–438, New York, New York, USA. ACM Press.

[Cutting et Vishton, 1995] Cutting, J. E. et Vishton, P. M. (1995). Perceiving Layout and Knowing Distances. *In Perception of Space and Motion*, volume 5, chapitre 3, pages 69–117. Elsevier.

[Davis et O'Connell, 2000] Davis, B. J. et O'Connell, J. (2000). Shoulder, elbow and wrist components of physiologic tremor amplitudeas measured using a laser penlight. *European neurology*, 43(3):152–154.

[Delamare *et al.*, 2013a] Delamare, W., Coutrix, C. et Nigay, L. (2013a). Designing disambiguation techniques for pointing in the physical world. *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems - EICS '13*, pages 197–206.

[Delamare *et al.*, 2013b] Delamare, W., Coutrix, C. et Nigay, L. (2013b). Mobile pointing task in the physical world : balancing focus and performance while disambiguating. *In Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13*, pages 89–98, New York, New York, USA. ACM Press.

[Delamare *et al.*, 2015a] Delamare, W., Coutrix, C. et Nigay, L. (2015a). A Tool for Optimizing the Use of a Large Design Space for Gesture Guiding Systems.

[Delamare *et al.*, 2015b] Delamare, W., Coutrix, C. et Nigay, L. (2015b). Designing guiding systems for gesture-based interaction. *In Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS '15*, pages 44–53, Duisburg, Germany. ACM Press.

[Drury *et al.*, 1987] Drury, C. G., Montazer, M. A. et Karwan, M. H. (1987). Self-Paced Path Control as an Optimization Task.

[Dubois *et al.*, 2002] Dubois, E., Nigay, L. et Troccaz, J. (2002). Assessing continuity and compatibility in augmented reality systems.

[Feiner, 2003] Feiner, A. (2003). The flexible pointer : An interaction technique for selection in augmented and virtual reality. *In Conference supplement of ACM symposium on user interface software and technology - UIST' 03*, volume 03, pages 81–82, New York, New York, USA. ACM Press.

[Fishkin, 2004] Fishkin, K. P. (2004). A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing*, 8(5):347–358.

[Fitts, 1954] Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47:381 – 391.

[Forsberg *et al.*, 1996] Forsberg, A., Herndon, K. et Zeleznik, R. (1996). Aperture based selection for immersive virtual environments. *In Proceedings of the 9th annual ACM symposium on User interface software and technology - UIST '96*, pages 95–96, New York, New York, USA. ACM Press.

[Fothergill *et al.*, 2012] Fothergill, S., Mentis, H., Kohli, P. et Nowozin, S. (2012). Instructing people for training gestural interactive systems. *In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, pages 1737–1746, New York, New York, USA. ACM Press.

[Freeman *et al.*, 2009] Freeman, D., Benko, H., Morris, M. R. et Wigdor, D. (2009). ShadowGuides : Visualizations for In-Situ Learning of Multi-Touch and Whole-Hand Gestures. *In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '09*, pages 165–172, New York, New York, USA. ACM Press.

[Ghomi *et al.*, 2013] Ghomi, E., Huot, S., Bau, O., Beaudouin-Lafon, M. et Mackay, W. E. (2013). Arpège : Learning Multitouch Chord Gestures Vocabularies. *In Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces - ITS '13*, pages 209–218, New York, New York, USA. ACM Press.

[Goldstein, 2002] Goldstein, E. B. (2002). Perceiving Depth and Size. *In Sensation and Perception*, numéro 1, pages 225–268.

[Grossman et Balakrishnan, 2006] Grossman, T. et Balakrishnan, R. (2006). The design and evaluation of selection techniques for 3D volumetric displays. *In Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*, page 3, New York, New York, USA. ACM Press.

[Hinckley *et al.*, 2006] Hinckley, K., Guimbretiere, F., Baudisch, P., Sarin, R., Agrawala, M. et Cutrell, E. (2006). The springboard : multiple modes in one spring-loaded control. *In Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, pages 181–190, New York, New York, USA. ACM Press.

[Hoffman *et al.*, 2010] Hoffman, M., Varcholik, P. et LaViola, J. J. (2010). Breaking the status quo : Improving 3D gesture recognition with spatially convenient input devices. *In 2010 IEEE Virtual Reality Conference (VR)*, pages 59–66. IEEE.

[Kamal *et al.*, 2014] Kamal, A., Li, Y. et Lank, E. (2014). Teaching motion gestures via recognizer feedback. *In Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14*, pages 73–82, New York, New York, USA. ACM Press.

[Karam et Schraefel, 2005] Karam, M. et Schraefel, m. c. (2005). A Taxonomy of Gestures in Human Computer Interactions. pages 1–45.

[Kela *et al.*, 2006] Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L. et Di Marca, S. (2006). Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing*, 10(5):285–299.

[König *et al.*, 2009] König, W. a., Gerken, J., Dierdorf, S. et Reiterer, H. (2009). Adaptive pointing-design and evaluation of a precision enhancing technique for absolute pointing devices. *In Human-Computer Interaction – INTERACT 2009*, volume 5726 LNCS, pages 658–671. Springer.

[Kopper *et al.*, 2011] Kopper, R., Bacim, F. et Bowman, D. a. (2011). Rapid and accurate 3D selection by progressive refinement. *In Proceedings of the 2011 IEEE*

Symposium on 3D User Interfaces - 3DUI '11, pages 67–74, Washington, DC, USA. IEEE Computer Society.

[Kopper *et al.*, 2010] Kopper, R., Bowman, D. a., Silva, M. G. et McMahan, R. P. (2010). A human motor behavior model for distal pointing tasks. *International Journal of Human Computer Studies*, 68(10):603–615.

[Kratz et Rohs, 2010] Kratz, S. et Rohs, M. (2010). A $3 gesture recognizer : simple gesture recognition for devices equipped with 3D acceleration sensors.

[Kühnel *et al.*, 2011] Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A. et Möller, S. (2011). Im home : Defining and evaluating a gesture set for smart-home control. *International Journal of Human Computer Studies*, 69(11):693–704.

[Kurtenbach *et al.*, 1994] Kurtenbach, G., Moran, T. P. et Buxton, W. (1994). Contextual Animation of Gestural Commands. *Computer Graphics Forum*, 13(5):305–314.

[Kurtenbach *et al.*, 1993] Kurtenbach, G., Sellen, A. et Buxton, W. (1993). An Empirical Evaluation of Some Articulatory and Cognitive Aspects of Marking Menus.

[Lee *et al.*, 2011] Lee, Post et Ishii (2011). ZeroN : Mid-air tangible interaction enabled by computer controlled magnetic levitation. *In Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, pages 327–366, New York, New York, USA. ACM Press.

[Lee *et al.*, 2009] Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H. et Choi, J. S. (2009). Freeze-Set-Go interaction method for handheld mobile augmented reality environments. *In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology - VRST '09*, pages 143–146, New York, New York, USA. ACM Press.

[McNeil, 1992] McNeil, D. (1992). Hand and mind : What gestures reveal about thought. *In Hand and Mind : What Gestures Reveal about Thought*, pages 105–133. University of Chicago Press.

[Milgram et Colquhoun, 1999] Milgram, P. et Colquhoun, H. (1999). A Taxonomy of Real and Virtual World Display Integration. *In Mixed Reality-Merging Real and Virtual Worlds*, pages 5–30.

[Milgram et Kishino, 1994] Milgram, P. et Kishino, F. (1994). A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329.

[Morency et Darrell, 2006] Morency, L.-p. et Darrell, T. (2006). Head gesture recognition in intelligent interfaces. *In Proceedings of the 11th international conference on Intelligent user interfaces - IUI '06*, pages 32–38, New York, New York, USA. ACM Press.

[Myers *et al.*, 2002] Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R. et Long, A. C. (2002). Interacting at a distance : measuring the

performance of laser pointers and other devices. *In Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, numéro 4, pages 33–40, New York, New York, USA. ACM Press.

[Nacenta *et al.*, 2013] Nacenta, M. a., Kamber, Y., Qiang, Y. et Kristensson, P. O. (2013). Memorability of pre-designed and user-defined gesture sets. *In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '13*, pages 1099–1108, New York, New York, USA. ACM Press.

[Ni *et al.*, 2011] Ni, T., Bowman, D. a., North, C. et McMahan, R. P. (2011). Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *International Journal of Human Computer Studies*, 69(9):551–562.

[Park *et al.*, 2000] Park, J. H., Shea, C. H. et Wright, D. L. (2000). Reduced-frequency concurrent and terminal feedback : a test of the guidance hypothesis. *Journal of motor behavior*, 32(3):287–296.

[Patel et Abowd, 2003] Patel, S. N. et Abowd, G. D. (2003). A 2-Way Laser-Assisted Selection Scheme for Handhelds in a Physical Environment. *In Proceedings of The International Conference on Ubiquitous Computing - UbiComp '03*, volume 2864, pages 200–207. Springer-Verlag, Berlin, Heidelberg.

[Pook *et al.*, 2000] Pook, S., Lecolinet, E., Vaysseix, G. et Barillot, E. (2000). Control Menus : Excecution and Control in a Single Interactor. *In extended abstracts on Human factors in computing systems - CHI '00*, numéro April, page 263, New York, New York, USA. ACM Press.

[Portillo-Rodriguez *et al.*, 2008] Portillo-Rodriguez, O., Sandoval-Gonzalez, O. O., Ruffaldi, E., Leonardi, R., Avizzano, C. A. et Bergamasco, M. (2008). Real-Time Gesture Recognition, Evaluation and Feed-Forward Correction of a Multimodal Tai-Chi Platform. *In Haptic and Audio Interaction Design*, volume 5270 LNCS, pages 30–39. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Poupyrev et Ichikawa, 1999] Poupyrev, I. et Ichikawa, T. (1999). Manipulating Objects in Virtual Worlds : Categorization and Empirical Evaluation of Interaction Techniques. *Journal of Visual Languages Computing*, 10(1):19–35.

[Qin *et al.*, 2010] Qin, Y., Shi, Y., Jiang, H. et Yu, C. (2010). Structured laser pointer : enabling wrist-rolling movements as a new interactive dimension. *In Proceedings of the International Conference on Advanced Visual Interfaces - AVI '10*, pages 163–166, New York, New York, USA. ACM Press.

[Rahman *et al.*, 2010] Rahman, A. S. M. M., Hossain, M. A. et Saddik, A. E. (2010). Spatial-geometric approach to physical mobile interaction based on accelerometer and IR sensory data fusion.

[Rahman *et al.*, 2009] Rahman, M., Gustafson, S., Irani, P. et Subramanian, S. (2009). Tilt techniques : investigating the dexterity of wrist-based input. *In Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, pages 1943 – 1952, New York, New York, USA. ACM Press.

[Raskar *et al.*, 2004] Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D. et Willwacher, T. (2004). RFIG Lamps : Interacting with a Self-DescribingWorld via Photosensing Wireless Tags and Projectors. *ACM Transactions on Graphics*, 23(3):406.

[Raskin, 1994] Raskin, J. (1994). Viewpoint : Intuitive equals familiar. *Commun. ACM*, 37(9):17–18.

[Ren et O'Neill, 2012] Ren, G. et O'Neill, E. (2012). 3D Marking menu selection with freehand gestures. *In IEEE Symposium on 3D User Interfaces - 3DUI '12*, pages 61–68. IEEE.

[Rohs et Oulasvirta, 2008] Rohs, M. et Oulasvirta, A. (2008). Target acquisition with camera phones when used as magic lenses. *In Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, pages 1409 – 1418, New York, New York, USA. ACM Press.

[Rohs *et al.*, 2009] Rohs, M., Schleicher, R., Schöning, J., Essl, G., Naumann, A. et Krüger, A. (2009). Impact of item density on the utility of visual context in magic lens interactions. *Personal and Ubiquitous Computing*, 13(8):633–646.

[Rovelo *et al.*,] Rovelo, G., Degraen, D., Vanacken, D., Luyten, K. et Coninx, K. Gestu-Wan -An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays.

[Rukzio *et al.*, 2007] Rukzio, E., Broll, G., Leichtenstern, K. et Schmidt, A. (2007). Mobile Interaction with the Real World : An Evaluation and Comparison of Physical Mobile Interaction Techniques. *In Ambient Intelligence*, pages 1–18. Springer.

[Sadri, 2011] Sadri, F. (2011). Ambient intelligence : a survey. *ACM Computing Surveys*, 43(4):1–66.

[Schmidt *et al.*, 2012] Schmidt, D., Molyneaux, D. et Cao, X. (2012). PICOntrol : using a handheld projector for direct control of physical devices through visible light. *In Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12*, volume 2, pages 379–388, New York, New York, USA. ACM Press.

[Schmidt et Wulf, 1997] Schmidt, R. a. et Wulf, G. (1997). Continuous Concurrent Feedback Degrades Skill Learning : Implications for Training and Simulation. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 39(4):509–525.

[Schönauer *et al.*, 2012] Schönauer, C., Fukushi, K., Olwal, A., Kaufmann, H. et Raskar, R. (2012). Multimodal Motion Guidance : Techniques for Adaptive and Dynamic Feedback. *In Proceedings of the 14th International Conference on Multimodal Interaction - ICMI '12*, pages 133–140, New York, New York, USA. ACM Press.

[Schwaller *et al.*, 2014] Schwaller, M., Kühni, J., Angelini, L. et Lalanne, D. (2014). Improving In-game Gesture Learning with Visual Feedback. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8512 LNCS, pages 643–653.

[Scoditti *et al.*, 2011] Scoditti, A., Vincent, T., Coutaz, J., Blanch, R. et Mandran, N. (2011). TouchOver : decoupling positioning from selection on touch-based handheld devices. *In 23rd French Speaking Conference on Human-Computer Interaction - IHM '11*, page 1, New York, New York, USA. ACM Press.

[Sodhi *et al.*, 2012] Sodhi, R., Benko, H. et Wilson, A. (2012). LightGuide : Projected Visualizations for Hand Movement Guidance. *In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, pages 179–188, New York, New York, USA. ACM Press.

[Sorel, 2012] Sorel, A. (2012). *Gestion de la variabilité morphologique pour la reconnaissance de gestes naturels à partir de données 3D.* Thèse de doctorat, Université Rennes 2.

[Steed, 2006] Steed, a. (2006). Towards a General Model for Selection in Virtual Environments. *In IEEE Symposium on 3D User Interfaces - 3DUI '06*, pages 103–110. IEEE.

[Swindells *et al.*, 2002] Swindells, C., Inkpen, K. M., Dill, J. C. et Tory, M. (2002). That one there ! Pointing to establish device identity. *In Proceedings of the 15th annual ACM symposium on User interface software and technology - UIST '02*, volume 4, page 151, New York, New York, USA. ACM Press.

[Talsma *et al.*, 2010] Talsma, D., Senkowski, D., Soto-Faraco, S. et Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration.

[Välkkynen *et al.*, 2006] Välkkynen, P., Niemelä, M. et Tuomisto, T. (2006). Evaluating touching and pointing with a mobile terminal for physical browsing. *In Proceedings of the 4th Nordic conference on Human-computer interaction changing roles - NordiCHI '06*, numéro October, pages 28–37, New York, New York, USA. ACM Press.

[Vatavu *et al.*, 2013] Vatavu, R., Casiez, G. et Grisoni, L. (2013). Small, medium, or large ? : estimating the user-perceived scale of stroke gestures. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 277–280, New York, New York, USA. ACM Press.

[Vatavu, 2011] Vatavu, R. D. (2011). The understanding of meaningful events in gesture-based interaction. *Studies in Computational Intelligence*, 332:1–19.

[Vatavu, 2012] Vatavu, R. D. (2012). Nomadic gestures : A technique for reusing gesture commands for frequent ambient interactions. *Journal of Ambient Intelligence and Smart Environments*, 4(2):79–93.

[Vatavu, 2013] Vatavu, R. D. (2013). A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *Journal of Ambient Intelligence and Smart Environments*, 5(2):187–211.

[Vatavu et Pentiuc, 2008] Vatavu, R. D. et Pentiuc, S. G. (2008). Multi-level representation of gesture as command for human computer interaction. *Computing and Informatics*, 27(6):837–851.

[Vatavu *et al.*, 2011] Vatavu, R.-D., Vogel, D., Casiez, G. et Grisoni, L. (2011). Estimating the Perceived Difficulty of Pen Gestures. *In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - INTERACT'11*, volume 6947 LNCS, pages 89–106. Springer-Verlag Berlin, Heidelberg.

[Vermeulen *et al.*, 2013] Vermeulen, J., Luyten, K., van den Hoven, E. et Coninx, K. (2013). Crossing the bridge over norman's gulf of execution. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 1931–1940, New York, New York, USA. ACM Press.

[Vernier et Nigay, 2001] Vernier, F. et Nigay, L. (2001). A framework for the combination and characterization of output modalities. *Interactive Systems Design, Specification, and Verification*, pages 35–50.

[Walter *et al.*, 2013] Walter, R., Bailly, G. et Müller, J. (2013). StrikeAPose : Revealing Mid-Air Gestures on Public Displays. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, pages 841–850, New York, New York, USA. ACM Press.

[Ware, 2004] Ware, C. (2004). *Information visualization : perception for design*. Elsevier.

[White *et al.*, 2007] White, S., Lister, L. et Feiner, S. (2007). Visual Hints for Tangible Gestures in Augmented Reality. *In 6th IEEE and ACM International Symposium on Mixed and Augmented Reality - ISMAR '07*, pages 1–4. IEEE.

[Wilson et Pham, 2003] Wilson, A. et Pham, H. (2003). Pointing in intelligent environments with the worldcursor. *In Proceedings of Interact*, pages 495–502. Citeseer.

[Wilson et Shafer, 2003] Wilson, A. et Shafer, S. (2003). XWand. *In Proceedings of the conference on Human factors in computing systems - CHI '03*, numéro 5, pages 545–552, New York, New York, USA. ACM Press.

[Wobbrock *et al.*, 2008] Wobbrock, J. O., Myers, B. a. et Aung, H. H. (2008). The performance of hand postures in front- and back-of-device interaction for mobile computing. *International Journal of Human Computer Studies*, 66(12):857–875.

[Yao *et al.*, 2013] Yao, L., Niiyama, R., Ou, J., Follmer, S., Della Silva, C. et Ishii, H. (2013). PneUI : Pneumatically Actuated Soft Composite Material s for Shape Changing Interfaces. *In Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 13–22.

[Zhai *et al.*, 1996] Zhai, S., Milgram, P. et Buxton, W. (1996). The influence of muscle groups on performance of multiple degree-of-freedom input. *In Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*, pages 308–315, New York, New York, USA. ACM Press.

[Zhang *et al.*, 2014] Zhang, B., Chen, Y.-h., Tuna, C., Dave, A., Li, Y. et Lee, E. (2014). HOBS : Head Orientation-Based Selection in Physical Spaces. *In Proceedings of the 2nd ACM symposium on Spatial user interaction - SUI '14*, pages 17–25.

[Zhang *et al.*, 2013] Zhang, Y., Han, T., Ren, Z., Umetani, N., Tong, X., Liu, Y., Shiratori, T. et Cao, X. (2013). BodyAvatar : Creating Freeform 3D Avatars using

First- Person Body Gestures. *In Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 387–396, New York, New York, USA. ACM Press.

Interaction à distance en environnement physique augmenté

Nous nous intéressons à l'interaction dans le contexte d'environnements physiques augmentés, plus précisément avec les objets physiques qui les composent. Bien que l'augmentation de ces objets offre de nouvelles possibilités d'interaction, notamment celle d'interagir à distance, le monde physique possède des caractéristiques propres rendant difficile l'adaptation de techniques d'interaction existantes en environnements virtuels. Il convient alors d'identifier ces caractéristiques afin de concevoir des techniques d'interaction à la fois efficaces et plaisantes dédiées à ces environnements physiques augmentés. Dans nos travaux, nous décomposons cette interaction à distance avec des objets physiques augmentés en deux étapes complémentaires : la sélection et le contrôle. Nous apportons deux contributions à chacun de ces champs de recherche. Ces contributions sont à la fois conceptuelles, avec la création d'espaces de conception, et pratiques, avec la conception, la réalisation logicielle et l'évaluation expérimentale de techniques d'interaction :

- Pour l'étape de sélection, nous explorons la désambiguïsation potentielle après un geste de pointage à distance définissant un volume de sélection comme avec une télécommande infrarouge par exemple. En effet, bien que ce type de pointage sollicite moins de précision de la part de l'utilisateur, il peut néanmoins impliquer la sélection de plusieurs objets dans le volume de sélection et donc nécessiter une phase de désambiguïsation. Nous définissons et utilisons un espace de conception afin de concevoir et évaluer expérimentalement deux techniques de désambiguïsation visant à maintenir l'attention visuelle de l'utilisateur sur les objets physiques.
- Pour l'étape de contrôle, nous explorons le guidage de gestes 3D lors d'une interaction gestuelle afin de spécifier des commandes à distance. Ce guidage est nécessaire afin d'indiquer à l'utilisateur les commandes disponibles ainsi que les gestes associés. Nous définissons un espace de conception capturant les caractéristiques comportementales d'un large ensemble de guides ainsi qu'un outil en ligne facilitant son utilisation. Nous explorons ensuite plusieurs options de conception afin d'étudier expérimentalement leurs impacts sur la qualité du guidage de gestes 3D.

Discipline : **Informatique**, spécialité : **Interaction Homme-Machine**
Mots Clés : **Interaction à Distance, Environnement Physique Augmenté, Pointage, Désambiguïsation, Guidage, Interaction Gestuelle, Espace de Conception.**

Thèse préparée au **Laboratoire d'Informatique de Grenoble. Equipe Ingénierie de l'Interaction Homme-Machine.**

Distant interaction in an augmented physical environment

We explore interaction with augmented physical objects within physical environments. Augmented physical objects allow new ways of interaction, including distant interaction. However, the physical world has specificities making difficult the adaptation of interaction techniques already existing in virtual environments. These specificities need to be identified in order to design efficient and enjoyable interaction techniques dedicated to augmented physical environments. In our work, we split up distant interaction into two complementary stages: the selection and the control of augmented physical objects. For each of these stages, our contribution is two-fold. These contributions are both theoretical, with the establishment of design spaces, and practical, with the design, the implementation and the experimental evaluation of interaction techniques:

- For the selection stage, we study the disambiguation potentially needed after a distal pointing gesture using a volume selection such as an infrared remote controller. Indeed, although the volume selection can facilitate the aiming action, several objects can fall into the selected volume. Thus, users should disambiguate this coarse pointing selection. We define and use a design space in order to design and experimentally evaluate two disambiguation techniques that maintain the user's focus on the physical objects.
- For the control stage, we study the guidance of 3D hand gestures in order to trigger commands at a distance. Such guidance is essential in order to reveal available commands and the associated gestures. We define a design space capturing specificities of a wide range of guiding systems. We also provide an online tool, easing the use of such a large design space. We then explore the impact of several design options on the quality of 3D gestures guidance.

Discipline: **Computer Science**, specialty: **Human-Computer Interaction**
Keywords: **Distant Interaction, Augmented Physical Environment, Pointing, Disambiguation, Guidance, Gesture Interaction, Design Space.**

Thesis prepared at the **Grenoble Informatics Laboratory. Engineering Human-Computer Interaction group.**