



FACULTE DES SCIENCES & TECHNIQUES

U.F.R. Sciences et Techniques Mathématiques, Informatique et Automatique
Ecole Doctorale IAEM LORRAINE
Département de Formation Doctorale en Informatique

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I
en Informatique

par **Charles H. A. TISON**

Mise en œuvre et évaluation d'interactions multimodales orientées communication appliquées au déplacement d'objets graphiques 3D

Soutenue publiquement le 15 novembre de l'an MMIV

Membres du jury :

Rapporteurs :	Professeur Christophe CHAILLOU	Université des Sciences et Technologies de Lille
	Professeur Eric BRANGIER	Université de Metz
Examineurs :	Professeur Jean-Paul HATON	Université Henri POINCARE, Nancy I
	Docteur Cédric DUMAS	Ecole des Mines de Nantes
Directeurs de thèse :	Professeur Noëlle CARBONELL	Université Henri POINCARE, Nancy I
	Docteur Dominique SCAPIN	Directeur de recherche, INRIA

Résumé: Les travaux décrits dans cette thèse concernent l'applicabilité d'une combinaison de modalités pour l'interaction avec les objets 3D d'un environnement virtuel afin de leur appliquer des repositionnements géométriques. L'approche « top-down » adoptée s'inspire des modalités observables au sein des expressions de commande d'utilisateurs au cours d'activités « *d'inspection de prototype* » du domaine de la conception vestimentaire. A partir d'un ensemble sélectionné de tâches, l'examen de combinaisons tâches vs modalités d'entrée permet d'établir des scénarios d'expressions de commande homogènes. La définition de profils de dispositifs conduit ensuite à la construction d'un système interactif sur lequel tester ces affectations de modalités avec des utilisateurs lors d'une expérience de type « Magicien d'Oz ». L'étude de la variabilité d'expression permet alors de poser les bases d'une grammaire d'interaction. Les résultats obtenus semblent confirmer la pertinence de ce style d'interaction.

Title: “Implementation and evaluation of a communicative like multimodal interaction applied to graphical 3D objects relocation.”

Abstract: The work described in this thesis concerns the applicability of a modalities combination in order to relocate 3D objects in a virtual reactive environment. The adopted “top-down” approach is initiated through the observation of command expressions modalities that take place during “prototype inspection” activities in the field of cloth design. A selection of a reduced set of user tasks, (3D objects relocation), helps investigate tasks vs input modalities combinations in order to determine an homogeneous set of low level interaction scenarios. The profile definition of relevant devices brings to build, afterwards, a minimal interactive system organised so as to test, with users, the relevancy of such a task modality allocation through an adapted version of an experiment known as “Wizard of Oz”. An evaluation of speech variability helps simultaneously to precise the grammar to implement. The results seem to confirm the relevancy of such an interaction style.

Mots clés: 3D, réalité virtuelle, IHM, multimodalité, voix, gestes, direction du regard, conception, ergonomie, expérimentation.

Keywords: 3D, virtual reality, HCI, multimodality, speech, gesture, gaze, modeling, ergonomoy, experimentation.

Projet MerLIn, INRIA, Unité de Recherche de Rocquencourt, 78153, Le Chesnay, France.

A Charles, Alexandre et Anghus.



« Merlin et Comedia », décembre 2000.

Remerciements.

Je remercie Noëlle CARBONELL et Dominique L. SCAPIN de m'avoir accueilli en thèse d'informatique à l'Université Henri POINCARÉ de Nancy.

Pour m'avoir fait l'honneur d'accepter de participer au jury de cette thèse, je remercie les rapporteurs, MM. Eric BRANGIER et Christophe CHAILLOU, ainsi que les examinateurs, MM. Cédric DUMAS et Jean-Paul HATON.

Du côté de la vie quotidienne au laboratoire, je remercie d'abord mon directeur de recherche à l'INRIA, Dominique L. SCAPIN, pour m'avoir accueilli au sein du projet MERLIn à Rocquencourt ces dernières années. Je le remercie également pour sa confiance et son soutien scientifique.

Je remercie les collègues de travail. Cédric BACH, dont l'approche ergonomique du même sujet a souvent orienté nos discussions sans que nous ayons jamais été en désaccord. Vincent LUQUIAUD, malgré des échanges parfois houleux autour de Java et de certains fondements de l'informatique. Noëly GRONDIN pour sa bonne humeur et sa façon de relativiser les problèmes quand le besoin s'en fait sentir. Stéphanie AUBIN pour sa gentillesse et sa patience concernant les choses administratives.

Il me faut également remercier, sans les nommer ici (loi du 6 janvier 1978), tous les utilisateurs qui ont participé à l'expérimentation, sans le sympathique concours desquels aucun résultat ne serait disponible.

Je remercie tous ceux de mes proches qui ont réussi à me supporter au long de cette longue période d'étude, et ce, malgré ma présence rendue très relative auprès d'eux faute de temps, malgré les difficultés qui n'ont pas manqué de survenir et l'inconfort que cela pouvait leur procurer, transformant pour eux cette situation en épreuve.

Je remercie en particulier Catherine, Michel, Nicole, Patience et Frédérique pour leur soutien et leur aide constante ou face aux difficultés même quand la situation, devenue trop pénible pour certains, les incita à délaissé provisoirement celles-ci aux autres pour se remettre eux-mêmes en condition.

Je remercie par ailleurs, quand bien même ils ne feraient plus partie des précédents, tous ceux de mes proches sans l'aide initiale ou passée desquels cette thèse n'aurait pu avoir lieu. En particulier, à ce titre, Denise C. M. E. *Roberte* SINCLAIRE, qui m'a encouragé, au long de nombreuses années, à progresser le long de mes centres d'intérêt malgré une attitude parfois dubitative mais cependant compréhensible quant au bien fondé de la démarche globale adoptée.

Je remercie, finalement, Pierre KLOSSOWSKI, dont la disparition au cours de cette période n'aura pas, avec tous mes regrets, permis à son petit-fils de lui présenter une progression encore partielle de ce point de vue mais néanmoins en grande partie inspirée de son propos sur la virtualité apparente du monde et ses *simulacres*.

Table des matières

TABLE DES MATIERES	1
CHAPITRE 1. INTRODUCTION	3
1.1 NATURE ET POSITION DU PROBLEME	3
1.2 PROBLEMATIQUE, APPROCHE CENTREE UTILISATEUR.....	5
1.3 PLAN DE LA THESE	7
PARTIE I	9
CHAPITRE 2. DE LA REALITE PHYSIQUE AUX ENVIRONNEMENTS VIRTUELS 3D : DEFINITIONS ET APPLICATIONS	11
2.1 DEFINITIONS ET TAXONOMIES ACTUELLES	11
2.1.1. Réalités virtuelles, augmentées, mixtes ?.....	11
2.1.2. Notion de « Présence » ?.....	14
2.2 LES DOMAINES D'APPLICATION POUR LA CONCEPTION.....	17
2.2.1. Environnements Culturels ou de loisir (Entertainment).....	18
2.2.2. Enseignement.....	20
2.2.3. Jeux.....	20
2.2.4. Environnements permettant la modification d'objets graphiques.....	22
CHAPITRE 3. MODALITES ET DISPOSITIFS D'INTERACTION AVEC LES ENVIRONNEMENTS 3D	27
3.1 MODALITES DE SORTIE.....	30
3.1.1. Modalités de sortie visuelle (affichages)	30
3.1.2. Autres Modalités de Sortie (non visuelles).....	49
3.2 MODALITES D'ENTREE	55
3.2.1. les "gestuelles".....	55
3.2.2. La parole	79
CHAPITRE 4. METAPHORES D'INTERACTION 3D ET IMPLEMENTATION LOGICIELLE	83
4.1.1. La "Navigation"	83
4.1.2. L'Adressage des Objets 3D ; Pointage 3D et Sélection.....	86
4.1.3. La "Manipulation" d'objets 3D	92
4.1.4. La "Création" d'Objets 3D	94
4.1.5. La "Modification" d'Objets 3D.....	94
4.2 CONCLUSION	98
PARTIE II	101
CHAPITRE 5. MANIPULATION DANS UN ENVIRONNEMENT VIRTUEL 3D : TACHES ET COMMANDES ELEMENTAIRES	103
5.1 CLASSIFICATION DE TACHES	104
5.1.1. Pointage, Visée et Présélection au sein d'un Environnement Virtuel.....	105
5.1.2. Sélection au sein de l'Environnement Virtuel.....	106
5.1.3. Mouvements d'objets 3D au sein de l'environnement.....	108
5.2 TYPES DE MODALITES EXISTANTES	114
5.2.1. Mouvements généraux.....	115
5.2.2. Mouvements des mains et des doigts.....	115
5.2.3. Mouvement des pieds (et des orteils).....	120
5.2.4. Mouvements de la tête.....	120
5.2.5. Mouvements faciaux : Mouvements Oculaires.....	121
5.2.6. Enonciation vocale.....	122
5.2.7. Combinaisons de modalités.....	124
5.3 TYPES DE DISPOSITIFS CORRESPONDANTS.....	131
5.3.1. Mouvements de la tête	133
5.3.2. Mouvements des yeux.....	134
5.3.3. Mouvements des mains.....	136
5.3.4. Enonciations.....	138

5.4 ROLE DES MODALITES D'ENTREE AU SEIN DE L'INTERACTION.....	140
5.4.1. <i>L'interaction au niveau le plus bas.</i>	141
5.4.2. <i>Expressions de commande</i>	150
5.4.3. <i>Les positionnements</i>	152
5.4.4. <i>Les déplacements</i>	156
5.4.5. <i>Les mouvements circulaires</i>	158
CHAPITRE 6. EXPERIMENTATION.....	163
6.1 OBJECTIF.....	163
6.2 PROBLEMATIQUE ET HYPOTHESES.....	164
6.3 METHODE.....	167
6.3.1. <i>Pré-expérience</i>	170
6.3.2. <i>Le « Magicien d'Oz »</i>	170
6.3.3. <i>Les utilisateurs sujets</i>	170
6.3.4. <i>Construction du montage expérimental</i>	172
6.3.5. <i>Montage expérimental</i>	191
6.3.6. <i>Fonctionnement de l'ensemble</i>	192
6.4 MESURES.....	194
6.4.1. <i>Mesures effectuées par le système</i>	195
6.4.2. <i>Retranscription d'informations enregistrées</i>	197
6.4.3. <i>Restitutions multimedia</i>	200
CHAPITRE 7. RESULTATS.....	203
7.1 PERFORMANCE GLOBALE DES DEUX GROUPES D'UTILISATEURS AUX TACHES PROPOSEES.....	203
7.1.1. <i>Durées des expressions de commande</i>	206
7.1.2. <i>Effectifs des apparition d'erreurs d'énoncé.</i>	207
7.1.3. <i>Usage des modalités</i>	212
7.1.4. <i>Comparaisons individuelles d'erreurs ou de particularités.</i>	219
7.1.5. <i>En résumé</i>	225
7.1.6. <i>Hypothèse</i>	226
7.2 CONCLUSION AU VU DES RESULTATS.....	227
7.2.1. <i>Trois interrogations et leurs éléments de réponse</i>	227
7.2.2. <i>Amélioration de l'expérience (magicien d'OZ) suite aux problèmes constatés</i>	230
7.2.3. <i>Amélioration du système</i>	233
CHAPITRE 8. CONCLUSION.....	237
8.1 TRAVAUX COMPLEMENTAIRES A ENVISAGER.....	238
8.2 PERSPECTIVES.....	244
CHAPITRE 9. ACRONYMES.....	246
CHAPITRE 10. RÉFÉRENCES.....	247

Chapitre 1. Introduction

1.1 Nature et position du problème

Face aux développements de l'informatique en matière de nouvelles techniques graphiques 3D, d'environnements de réalité virtuelle, de réalité augmentée, etc., face à l'engouement et aux espoirs que ces développements suscitent, de nombreuses questions se posent, notamment d'un point de vue ergonomique, en particulier en ce qui concerne une IHM (interaction Homme-Machine) permettant d'interagir avec de tels environnements. Le postulat de base étant que pour la manipulation d'objets 3D, les techniques d'interaction classiques (clavier-souris) ne sont pas nécessairement les plus adaptées.

Le but de cette thèse, dans le cadre d'une informatique d'IHM, est de permettre de définir et d'évaluer des interactions facilitant l'action de l'utilisateur sur des objets graphiques 3D pour la réalisation de leur activité.

Malgré le caractère encore récent du domaine de l'interaction avec des environnements 3D qualifiés de « virtuels », la littérature relative à ce domaine est rapidement devenue considérable. Aussi, en vue de distinguer ce qui caractérise ce nouveau domaine ouvert à l'informatique, une exploration préalable des classifications produites par de précédentes contributions permet de déterminer les avantages et inconvénients des diverses approches, métaphores mises en oeuvre, techniques, dispositifs, etc.

Pour obtenir un aperçu d'ensemble du domaine, une revue de littérature fut d'abord envisagée dans un sens général. La constatation rapide de l'étendue du domaine ainsi que l'identification d'une immaturité à long terme de certaines techniques dont en particulier celles renseignant le sens tactilo-kinesthésique chez l'Homme (ou « sens du toucher ») permit une restriction de l'étendue du problème à un sous-domaine d'interaction privé d'un tel retour d'information. En conséquence, il apparut également une seconde restriction de l'espace du problème ne tenant plus compte des techniques d'interaction dites de « manipulation directe ».

Une revue de la littérature fut donc réalisée en explorant plus particulièrement ce qui pouvait alors contribuer à une catégorie d'interactions ne requérant ni le contact physique, ni l'obtention d'une modification simultanée de l'état du système, affectant immédiatement l'objet d'une tâche utilisateur, conformément à ses actions.

Un second écueil rapidement constaté au sein de nombre de techniques d'interaction consistait, pour l'utilisateur confronté à ces types d'environnement, en la difficulté à manipuler efficacement l'équivalent 3D du « pointeur » en partie de par une difficulté chez l'homme à évaluer correctement les distances le séparant des objets, et en majeure partie de par l'aggravation de cette carence constituée par l'incapacité des techniques actuelles d'affichage à représenter la profondeur conformément à la vision humaine.

Parallèlement, des enquêtes de terrains dans le domaine de la conception textile vestimentaire furent effectuées en rapport avec l'objectif initial de cette thèse.

La compilation des écueils présentés par nombre de techniques d'interaction désignèrent ensuite les techniques orientées « communication Homme-Machine » comme candidats envisageables en entrée du système.

Une telle orientation imposa une évaluation de ce domaine dans un sens large et une prise en compte, chez l'être humain, des phénomènes observables contribuant à la « communication » en situation de commande. L'ensemble de ces phénomènes observables adressa ensuite plus particulièrement les définitions des gestuelles auxquelles furent étendus les actes de langage, ainsi que les notions de multimodalité d'entrée d'un système.

La recherche ultérieure d'un nombre minimal de modalités d'expression dont les combinaisons laissent prévoir la manifestation par l'utilisateur d'un spectre maximal d'expressions de commande en entrée mena finalement à retenir les mouvements du regard ou de la tête comme média d'expression de modalités passives de visée. Certains mouvements de la main désignèrent également cet appendice comme source d'expression complémentaire, cependant que les manifestations d'énoncés vocaux appurent sans équivalent pour enrichir le spectre d'expression.

Une étude de la mise en œuvre de combinaisons de ces trois modalités d'entrée (voire groupes de modalités), a permis d'envisager, à partir d'une définition de tâches de commande de mouvements d'objets 3D disposés sur une scène virtuelle, un langage de commande partiel qui reste à préciser pour compléter une grammaire exhaustive.

Nous verrons au chapitre 5 qu'à ce stade de l'étude, l'objectif de ce travail de recherche s'est vu modifié par le retrait du partenaire industriel fournisseur d'accès à la fois aux terrains d'étude du domaine de la conception vestimentaire ainsi qu'à une application informatisée relative au domaine.

C'est alors une étude avec des utilisateurs qui a été envisagée afin de tester d'une part la pertinence d'une interaction orientée communication incluant les actes non verbaux pour effectuer des tâches de mouvements d'objets 3D au sein d'un environnement virtuel. Elle a été envisagée également pour tester d'autre part, à cet effet, la syntaxe que nous avons définie pour l'usage des modalités gestuelles aux seuls niveaux primitifs de l'interaction (présélections, sélections), ainsi que l'orientation « langages de commande » pour la mise en œuvre combinée des trois modalités retenues afin de permettre à l'utilisateur d'effectuer des tâches de mouvements d'objets 3D. Enfin, cette étude a nécessité d'obtenir davantage d'informations sur l'usage de la modalité vocale et son impact d'ensemble sur la formulation d'expressions de commande de ces mouvements d'objets 3D.

1.2 Problématique, approche centrée Utilisateur.

L'IHM (Interaction Homme-Machine) est un domaine pluridisciplinaire qui se situe à un carrefour des sciences humaines et des sciences informatiques. Elle relève de l'étude des facteurs humains, de la psychologie, de l'ergonomie voire de la sociologie et de l'anthropologie quand des facteurs culturels sont à considérer. Elle relève également du domaine de l'informatique dont elle met les techniques en œuvre pour ajuster l'outil informatique à l'être humain en vue d'une meilleure « utilisabilité ».

Le contexte initial de cette thèse se situe au sein du projet européen Eureka "Comedia" qui adresse les activités de conception dans le domaine du stylisme-modélisme.

L'approche envisagée est centrée sur l'utilisateur dans un domaine d'application de conception. Ce type d'approche tient notamment compte des caractéristiques et besoins de ces "*Utilisateurs Concepteurs*" ainsi que des techniques disponibles pour la réalisation de tâches déterminées par l'observation et l'analyse de l'activité. Une exploration des différentes possibilités doit donc être effectuée pour obtenir un panorama de techniques disponibles dont certaines seront mises en œuvre en respectant les critères utilisateurs. Cette exploration concerne trois catégories qui adressent les tâches de l'activité, les modalités d'interaction, en particulier en entrée du système, ainsi que les dispositifs existants.

Le but reste d'établir, justifier et mettre en œuvre une construction théorique permettant finalement à l'utilisateur d'interagir avec des objets graphiques 3D de manière "efficace".

D'un point de vue IHM pour les environnements 3D, l'expérience décrite dans cette thèse adresse l'applicabilité d'une multimodalité de communication utilisateur en entrée d'un système d'interaction avec les objets 3D d'un environnement virtuel. Ceci en vue de leur appliquer des repositionnements au sein d'une scène réactive affichée.

La détermination de modalités d'entrée à retenir repose sur une approche dite « top-down » à partir d'interactions utilisateurs observées dans le domaine de la conception textile vestimentaire au cours d'une activité « *d'inspection de prototype* ».

Dans une telle approche, un ensemble de tâches utilisateur est alors défini en respect des contraintes de l'activité. L'examen de ces tâches utilisateurs en vue d'en obtenir une définition du plus général au plus particulier, a soulevé un problème de tâches « atomiques » à considérer au bas niveau d'interaction. Ceci nous a conduit à nous démarquer de la notion de « *pointeur 3D* » au profit d'une notion plus atomique, à notre sens, de « *directions* » indiquées par l'utilisateur.

De telles indications de « *directions* » constituant nos tâches « universelles » de base permettent alors de construire des tâches composites de niveau plus élevé participant de l'interaction.

L'approche « top-down » prévoit ensuite d'examiner les modalités observables qui composent les expressions de commande chez l'utilisateur au cours de son activité, et d'en retenir les plus pertinentes. Plusieurs « *gestuelles* » furent alors définies pour distinguer en particulier les mouvements de la tête et les mouvements des mains comme média d'expressions complémentaires à la parole.

Ces trois modalités d'expression (parole, gestuelle de la tête, gestuelle des mains), apparemment en effet constituer le meilleur rapport entre effectif de modalités à mettre en œuvre et spectre d'expression utilisateur qu'elles permettent.

L'examen de chacune des trois modalités retenues permet ensuite de définir les critères d'un profil théorique de dispositifs propres à mesurer et convoier les informations de modalités. Un examen de la littérature vise alors à situer, quant il existe, un dispositif répondant à ces critères pour chaque modalité retenue afin de permettre, de la part d'un utilisateur, leur mise en œuvre au sein de tâches d'interaction définies.

Les trois modalités se voient alors affecter des rôles expressifs complémentaires en fonction de leurs compétences particulières. Ces rôles sont ensuite testés séparément et de manière combinée au cours de scénarios d'interaction afin d'en homogénéiser le style de mise en œuvre. Ceci nous a conduit à distinguer et préciser d'une part la notion de « *présélection* » en tant qu'acte utilisateur éventuellement continu, involontaire, mais ajustable révélant éventuellement un objet « d'intérêt », de celle, d'autre part, de « *sélection* » réalisée à l'aide d'un événement ponctuel, du moins de courte durée produisant une interprétation immédiate, mis en œuvre de façon volontaire, éventuellement motrice, pour effectuer la « *validation* » d'une « *présélection* » en cours.

Le style d'interaction résultant nécessite alors un test visant à en confirmer la pertinence, ainsi qu'une étude de la façon dont elle est mise en œuvre par des utilisateurs afin d'en préciser la grammaire à constituer à ce niveau d'interaction.

L'expérience proposée pour répondre à ces questions consiste en des tests utilisateur réalisés à l'aide de la technique du « *Magicien d'Oz* ». Par ailleurs, des techniques apparentées à celles de la post-production cinématographique ont été mises en œuvre pour évaluer les performances à des tâches de manipulation d'objets 3D.

D'un point de vue informatique, l'expérience s'est appuyée sur la réalisation d'une application 3D interactive répartie, permettant l'enregistrement des comportements des utilisateurs au cours d'interactions multimodales avec des objets 3D. Elle s'est également appuyée sur la réalisation d'une seconde application de restitution 3D de ces enregistrements à des fins d'observation, destinée à l'expérimentateur.

D'un point de vue IHM, les résultats obtenus semblent, pour des tâches de manipulation d'objet dans l'espace, confirmer la pertinence d'une interaction orientée

« Communication » Homme-Machine sur la base d'un nombre réduit de modalités d'entrée sélectionnées.

1.3 Plan de la thèse.

Dans une première partie, nous allons décrire une revue de questions sur les Modalités et dispositifs en sortie et en entrée. Cette revue de question décrit les classifications d'environnements virtuels. Elle aborde ensuite les modalités de sortie et d'entrée d'un système informatique. Et rapporte les métaphores et techniques d'interaction pertinentes de la littérature.

Nous étudierons ensuite une mise en correspondance des tâches versus Modalités-Dispositifs en effectuant une classification des tâches utilisateur restreinte aux mouvements d'objet 3D au sein d'un environnement virtuel, puis une description des modalités existantes sur lesquelles appuyer une interaction orientée « communication » en entrée.

La description des modalités nécessaires à renseigner permettra alors de décrire des profils de dispositifs à mettre en œuvre pour convoier les informations de modalités en entrée d'un système informatique.

L'étude de combinaisons de modalités apportant chacune des compétences complémentaires permettra de jeter les bases encore mal définies d'une syntaxe relevant des langages de commande concernant les tâches retenues pour cette étude, celles de mouvements ou relocalisation d'objets 3D.

Nous décrirons ensuite le montage expérimental réalisé pour mesurer la pertinence d'une telle multimodalité d'entrée pour les utilisateurs potentiels, mesurer également la conformité des affectations de tâches aux modalités sélectionnées, et obtenir des informations d'une part sur l'impact de la parole sur les expressions de commande ainsi, d'autre part, que la variabilité des énoncés à prendre en compte pour l'établissement d'une grammaire tenant compte des usages des utilisateurs.

Nous décrirons ensuite comment sont effectuées les mesures, le déroulement de l'expérience, et les résultats obtenus avant de conclure.

Partie I

Revue de questions

Chapitre 2. De la réalité physique aux environnements virtuels 3D : Définitions et applications

L'objectif de ce chapitre est d'une part de rassembler, au travers de la littérature scientifique, les connaissances actuelles en matière ergonomique et IHM¹ sur les techniques et modalités d'interaction homme-machine avec des objets et environnements 3D, de les organiser et d'en faire le bilan critique, mais aussi d'identifier les configurations IHM candidates à l'expérimentation pour nos recherches futures, dans des contextes d'activité de conception. En effet, ce travail a débuté dans le cadre du projet européen EUREKA-COMEDIA qui vise, à long terme, à "réduire le coût et la durée de la mise au point des collections, afin de donner aux stylistes et modélistes la possibilité de donner toute la mesure de leur talent en les libérant des contraintes inhérentes à la réalisation matérielle des prototypes". Il s'agit notamment d'étudier l'intérêt et la faisabilité d'outils de conception de CAO² en 3D, adaptés à leur culture et permettant de visualiser et d'évaluer de nouvelles idées grâce à la simulation dynamique de vêtements portés par un mannequin virtuel en mouvement. Cette étude a été menée en parallèle avec d'autres études à caractère évaluatif des outils informatiques existants (Hamard, 2000, Bach, 2000) et à caractère descriptif de l'activité (Taillefer & Scapin, 2000 ; Taillefer, 2001a, 2001b).

Ce chapitre évoque tout d'abord les définitions et typologies existantes. Puis, il évoque les domaines d'application auxquels nous nous intéressons, autour de la conception d'objets 3D. La revue de questions est ensuite organisée selon divers modes de sortie, puis selon les divers modes d'entrée. La conclusion évoque les combinaisons de modalités envisagées dans le contexte de nos travaux sur la conception.

2.1 Définitions et taxonomies actuelles

2.1.1. Réalités virtuelles, augmentées, mixtes ?

Les termes utilisés en matière de réalité virtuelle recouvrent diverses notions qui ne sont pas toujours clairement définies. Plusieurs auteurs ont tenté de classer ces notions. Milgram & Kishino (1994), Milgram et al. (1994), proposent de considérer la virtualité comme un continuum sur la dimension réel / virtuel qui s'étend des systèmes de représentation de la réalité physique objective aux systèmes représentant uniquement des objets 3D purement synthétiques. Au sein de ce continuum (voir Fig. 1), ils distinguent notamment la réalité augmentée de la virtualité augmentée.

¹ Interaction Homme-Machine.

² Conception Assistée par Ordinateur.



Figure 1 Continuum Réalité-Virtualité.

A gauche de cette échelle se trouvent les systèmes de vidéo surveillance transmettant une vue partielle de la réalité objective ainsi que l'observation directe de cette réalité objective par l'utilisateur. A l'extrémité droite se trouvent les environnements purement synthétiques.

Entre ces deux extrêmes se rangeraient tous les autres environnements d'affichage-présentation d'objets graphiques 3D se superposant à l'observation possible de l'environnement objectif.

Ces recherches étant issues de travaux concernant plus particulièrement la télémanipulation, cette échelle peut par ailleurs être perçue, en termes d'action utilisateur, comme celle séparant (à gauche) une action sur le monde réel au travers de médias (télémanipulation, navigation ou action d'un robot), d'une action utilisateur dans un environnement purement synthétique (relevant donc, à droite de l'échelle, de la simulation).

Au sein de ces "environnements virtuels" ("*Virtual Environments*", VE), les auteurs distinguent une "Réalité Augmentée" ("*Augmented Reality*", AR) d'une "Virtualité Augmentée" ("*Augmented Virtuality*", AV) à partir de l'identification de six classes d'affichage pouvant concerner les "Réalités Mixtes" ou hybrides ("*Mixed Reality*", MR). La "Réalité Augmentée" (AR) se référerait ainsi à la situation dans laquelle la perception visuelle de l'environnement réel par l'utilisateur est "augmentée" au moyen d'objets graphiques virtuels. La "Virtualité Augmentée" (AV), se concevrait par l'existence de l'une des classes d'affichage permettant la présentation d'environnements complètement graphiques (ou synthétiques, qu'ils soient ou non complètement immersifs), auxquels des séquences vidéo de la "Réalité" sont ajoutées, auquel cas ce qui est augmenté n'est pas une représentation de la scène réelle mais plutôt un monde virtuel essentiellement généré par une machine, situation pour laquelle les auteurs préfèrent le terme de "*Augmented Virtuality*" (AV).

Le principe directeur adopté par les auteurs (Milgram & Kishino, 1994) est de faciliter la distinction entre "*images réelles*" et "*images virtuelles*", notamment à partir de notions de l'optique qui définissent en l'occurrence toute image qui présente une certaine luminosité à l'endroit où elle apparaît. Cette définition incluerait non seulement la vision directe ("*direct viewing*"), mais aussi les images d'un objet "*non-direct viewing*" affichées sur un écran de moniteur. Une "*image virtuelle*" peut alors être définie comme une image qui ne présente pas de luminosité à l'endroit où elle apparaît, comme par exemple les hologrammes et les images réfléchies par miroir.

Cela inclut également le cas de l'affichage stéréoscopique pour lequel chacune des images droites et gauches est une image réelle, mais dont le percept 3D spatial résultant de la fusion est virtuel.

Parallèlement à ce continuum, Milgram & Kishino (1994) proposent une autre dimension qui distingue objets réels et objets virtuels (voir Fig. 2), mais qui va dans le même sens que la précédente :

Les "objets réels", ("*Real objects*") sont les objets qui ont une existence physique réelle.

Les "objets virtuels", ("*Virtual objects*") sont les objets qui existent en essence ou par leurs effets, mais pas de façon physique ou réelle.

Pour qu'un *objet réel* soit visuellement représenté, il lui faut soit être observé directement, soit être digitalisé puis resynthétisé et représenté via un dispositif d'affichage. Pour qu'un *objet virtuel* soit visualisé, il lui faut être *simulé* car il n'existe pas de manière réelle, physique, ce qui suppose l'existence d'un *modèle* de l'objet virtuel. Les auteurs ajoutent en note qu'un objet virtuel peut être construit aussi bien à partir du modèle d'un objet non-existant que du modèle d'un objet réel existant. Le modèle d'un objet virtuel pouvant aussi être un objet réel, ce qui est le cas pour des sculptures, peintures, etc.



Figure 2. Echelle "EWK".

Une autre interprétation de cette échelle, qui rejoint la notion "d'étendue de la connaissance de l'environnement" (ou "Extent of the World Knowledge EWK" développée elle aussi par Milgram *et al.*, (1994), *cf.* Fig.2), constitue une dimension supplémentaire de leur taxonomie des "mixed reality visual displays", et établit un continuum entre les représentations des objets du monde réel au sujet desquels la machine ne dispose d'aucune information (exemple une vidéo simultanée de la réalité objective), et entre les représentations des objets virtuels d'un monde synthétique (par exemple immersif) pour lesquels la machine dispose au moins de toutes les informations de modélisation et d'affichage.

Les exemples proposés concernent ainsi, pour l'extrémité gauche de l'échelle, des représentations à des fins de télémanipulation pour lesquelles la machine ne dispose pas de modèle, il s'agira soit de vision directe, soit d'objet "blindly scanned" pour représentation, mais dans ce cas, *la machine n'a pas d'information au sujet de ce qui est représenté sur l'affichage*. L'autre extrémité de l'échelle constitue le cas dans lequel la machine dispose seule de toutes les informations concernant chaque objet présent dans la scène, y compris la position et le point de vue de l'utilisateur.

Ce souci de distinction réel vs virtuel est rejoint par les propositions terminologiques de J.P. Papin, (2001), qui considère que l'utilisateur immergé au sein d'un environnement 3D de synthèse se voit présenter un environnement à partir d'une image immersive bien réelle (d'un point de vue physique, constituée de photons de différentes longueurs d'onde), aussi réelle que l'image du monde réel qui nous entoure. Il définit donc ainsi une autre "réalité" qu'expérimente l'utilisateur et qui semble bien faire partie des "Représentations mentales". J.P. Papin, (2001), estime ainsi que l'appellation "réalité virtuelle" est une traduction mot-à-mot impropre du terme anglo-saxon "Virtual Reality", car le terme "virtuel", contrairement à la "réalité", se définit comme un synonyme de "potentiel", c'est-à-dire qui n'a, *a priori*, pas encore d'existence. Sa proposition "*d'espace vicariant*" serait inspirée de l'usage du terme "vicariant" en neurophysiologie pour désigner une partie de l'encéphale qui permet ultérieurement la récupération fonctionnelle d'une aptitude perdue, auparavant liée à une autre partie de l'encéphale qui a subi des dommages irréversibles.

Pour notre part, nous parlerons d'environnements (*virtuels-réels-mixtes*) de représentation d'objets 3D (*bien que virtuel et réel soient antonymes*), ces représentations pouvant être issues plus ou moins directement de la réalité et/ou construites plus ou moins de manière synthétique.

2.1.2. Notion de « Présence » ?

Jusqu'à présent, il s'agissait de notions et de dimensions des caractéristiques intrinsèques de ces environnements. Une autre vue des choses a trait à la manière dont ces environnements sont « perçus » et à la notion de « présence ».

L' "*Extent of Presence Metaphor*" (EPM) troisième dimension de la "taxonomy of mixed reality visual displays" proposée par Milgram en 1994, fournit une mesure du degré auquel l'utilisateur se sent "présent" au sein de la scène représentée (feeling "present" in the scene).

En incluant cette dimension, on reconnaît le fait que les "*Mixed Realities*" incluent non seulement les environnements hautement immersifs, avec une importante métaphore de présence, mais aussi l'importante classe d'affichages basée sur l'usage de moniteurs (non-immersifs i.e. "window-on-the-world" -WoW- displays) de Réalité Augmentée (AR) exocentrique.

L'axe de l'EPM s'étend depuis la métaphore selon laquelle l'observateur scrute le monde de l'extérieur au travers d'une vue monoscopique depuis un point de vue fixe, jusqu'à celle du "realtime imaging", avec laquelle les sensations utilisateur ne diffèrent pas de celles produites par "l'unmediated reality" .

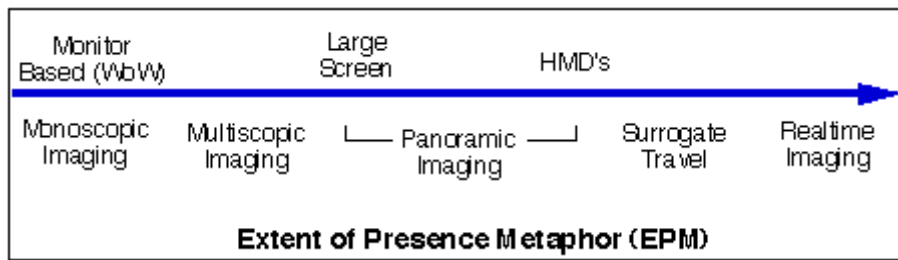


Figure 3. Echelle "EPM".

Adjacents sur l'échelle EPM à la classe de représentations WoW ("*Windows-on-the-World*") notamment produite par les moniteurs, se trouvent les représentations "multiscopiques", qui s'étendent au-delà des affichages stéréoscopiques, incluant les points de vue multiples. L'imagerie panoramique est une extension de l'imagerie "multiscopique", permettant à l'utilisateur de scruter autour de lui au sein de la scène affichée, davantage basée sur la métaphore lui donnant l'impression d'observer le monde de l'intérieur, et non plus de le scruter de l'extérieur. Le "Panoramic Imaging" peut être partiellement réalisé au moyen de grands écrans de projection, mais la métaphore résultante n'est alors valide que pour l'exécution de tâches qui seront contenues dans un volume de travail confiné. Une instanciation plus inclusive de cette classe d'affichage peut être réalisée à l'aide d'HMD.

Le "*Surrogate Travel*" se rapporte à la possibilité de déambuler à l'intérieur du monde représenté, alors que le « realtime imaging » se rapporterait lui, à la mise à jour simultanée des changements survenant dans le monde représenté.

Il n'existe pas de définition très claire de la notion de "Présence" au sein d'un environnement, ainsi, selon Shilling & Shinn-Cunningham, (2001), qui rapportent la définition très approximative de Sheridan, (1996) :

"it is generally believed that the sense of presence is dependent upon auditory, visual, and tactile fidelity". (Sheridan, 1996).

La notion d'une sensation de "Présence" au sein d'un environnement virtuel immersif est également évoquée par Slater & Usoh, (1994), qui décrivent la proprioception comme ce qui nous permet d'élaborer un modèle mental de notre propre corps, informé en permanence par un flot de données proprioceptif qui nous permet de le situer dans l'espace ainsi que ses parties. Ce flot d'informations proprioceptives entrerait en contradiction avec les autres données sensorielles si le corps est absent en environnement virtuel immersif. Les auteurs instaurent donc le "Virtual Body" comme un palliatif (visuel) représentant le corps en environnement virtuel, asservi aux dispositifs de "tracking" disponibles.

Ainsi, parmi les rares mesures de "Présence", Slater & Usoh (*op. cit.*) mènent une expérience avec deux groupes d'utilisateurs munis soit d'un "Virtual Body" (VB) ressemblant à un corps, soit d'un VB réduit à un curseur 3D-flèche poursuivant les positions successives de la main de l'utilisateur. Une notion qui se dégage de la mise en œuvre de cette expérience est le "pointeur 3D" considéré comme une réduction de

l'idée de "Virtual Body". La méthode d'évaluation employée se réfère à un modèle de l'expérience subjective, proche de la définition ci-dessus de Sheridan, qui établit qu'une telle expérience est encodée en termes de trois systèmes principaux de représentation, Visuel, Auditif et Kinesthésique. Le modèle utilisé permettrait aussi de distinguer la position perceptuelle egocentrée ou exocentrée qu'un utilisateur particulier utilisera de préférence. La sensation de "présence", rapportée à celle de "se trouver" dans l'environnement, serait réfutée en l'absence de Corps, ce que confirmerait l'expérience menée avec des utilisateurs. D'autres détails, implicitement liés à l'idée de devoir disposer d'un corps pour satisfaire à la sensation de "présence" au sein d'un environnement virtuel immersif, concerneraient d'autres indices visuels tels que l'ombre projetée du VB, la réflexion de son image sur d'autres objets, en particulier sur les miroirs qui pourraient y être présents et utilisés. Les auteurs dégagent donc l'idée de "*Body Centered Interaction*" dont la mise en œuvre nécessiterait des techniques d'interaction capables de satisfaire le sensorium (visuel, ce que perçoivent les yeux) et proprioceptif (ce que perçoit le corps). Bien que cette idée semble s'attacher à reproduire ce que l'on expérimente dans la réalité en suggérant de contraindre les techniques d'interaction en environnement virtuel immersif à celles dont on dispose dans la réalité, on peut en retirer que le problème d'absence de corps ne se pose pas de la même manière en Réalité Augmentée "See-Through" ou projetée sur grands écrans, du fait que le corps réel y est visible. Cela suggère néanmoins que disposer d'un Virtual Body dont la position ne correspond pas à ce que nous dit notre sens proprioceptif (en particulier dans le cas où l'utilisateur "contrôle" une paire de "Virtual Hands" distante sur quelque affichage), n'est pas, selon ce principe, compatible avec la sensation de présence au sein de l'environnement. On note toutefois qu'il sera préférable de disposer d'un moyen de tracking des mouvements de l'utilisateur, non seulement pour satisfaire à ce critère de "présence" si l'on a affaire à un environnement immersif au sein duquel la sensation de "présence" est souhaitée (sur le même principe que l'affichage de la présence d'un membre, d'une main, par exemple, doit correspondre à ce que fournit le sens proprioceptif), mais aussi en Réalité Augmentée collaborative afin de respecter la cohérence des scènes concernant l'obstruction visuelle des autres utilisateurs de l'espace partagé.

Autre information en faveur d'une perception, par l'utilisateur au sein d'un environnement virtuel (VE) immersif, soit de ses propres membres préhensiles, soit de leur représentation virtuelle reconstituée; il semble que plongé au sein d'un VE, face à une *tâche de préhension d'objets virtuels* ("*reaching and grasping task*") disposés devant lui sur une table (elle aussi virtuelle), l'utilisateur soit plus lent lorsqu'il ne voit pas son membre préhensile que lorsqu'il le perçoit au cours de la tâche, (Mason *et al.*, 2001).

Il existe bien sûr d'autres travaux sur des aspects plus dynamiques de la scène (ex. simulation dynamique d'un personnage de synthèse). Peu de mesures sont cependant

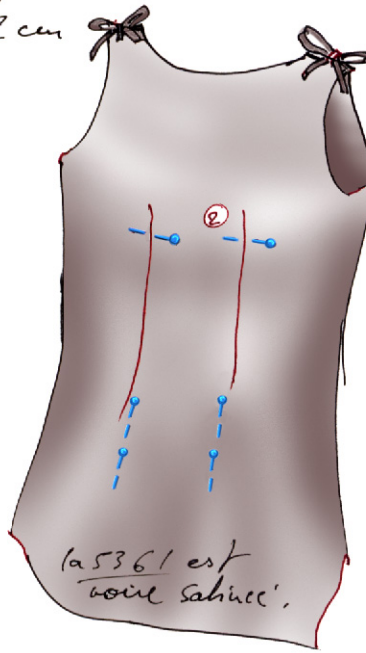
disponibles voire réalisables sur un tel sujet difficile à cerner, on y trouve toutefois nombre de travaux qui constituent autant de directions de recherche.

2.2 Les domaines d'application pour la conception.

Le domaine d'application auquel on s'intéresse plus particulièrement concerne les métiers de la conception pour lesquels une évaluation visuelle du prototype en cours de développement permet de situer les modifications ou retouches à effectuer.

Ainsi, l'inspection de prototypes de vêtements en cours de développement révèle l'usage, de la part du styliste, d'une verbalisation faisant appel à un vocabulaire propre au métier, complétée, par le modéliste, d'une sémiotique se manifestant sous la forme d'un "marquage" du prototype (à l'aide d'aiguilles) illustrant ou précisant le sens des verbalisations du styliste (cf. Fig.4 ci-après). Ces verbalisations se trouvent alors transcrites en ajoutant également des références au marquage réalisé.

L'exemple ci-dessous explique en commentaire les opérations qui seront à effectuer après retour en atelier. On y mentionne qu'il faudra "abaissier des pinces" qui seront par ailleurs à rallonger d'autant. Sachant que le modèle dispose aussi de pinces sur la face arrière, des aiguilles sont disposées en sorte de préciser l'interprétation du message verbal afin de lever une éventuelle ambiguïté pouvant survenir chez le destinataire du message (le modéliste) quant à cette interprétation.

<p>Robe 5361 idem 5357 et ② pour abaisser l'échec de 995 2cm</p> <p>V. Avt.</p>  <p>la 5361 est voire Sabine'.</p> <p>Figure 4. Exemple de modification sur un prototype.</p>	<p><u>Exemple de consignes du styliste destinées au modéliste:</u></p> <p>"Abaisser les pinces de 1,5cm et allonger de 3cm, élargir la taille. Creuser autour de la manche, fermer échancrure derrière. Modèle de base 5360 - comme épinglé. Vérifier les dimensions.-"</p> <p>("Abbassare le pinces di cm. 1,5 e allungare di cm. 3, allargare punto vita. Scavare giro manica, chiudere scollatura dietro. Capo da rif. 5360 (come spillato. Verificare le dimensioni.)").</p>
--	--

Certains exemples de VE actuellement existants et présentant un rapport avec de tels environnements fonctionnels sont évoqués ci-après.

Nous classons, pour notre part, les VE en fonction d'une capacité d'action croissante dont est muni l'utilisateur au sein de ces environnements. On trouvera ainsi d'abord les environnements à caractère culturel permettant par exemple les "visites virtuelles". On trouvera ensuite les environnements destinés à l'enseignement, puis certains environnements dédiés aux jeux, et enfin les environnements spécifiquement conçus pour leur propre modification de la part de l'utilisateur, comme les environnements de conception.

On gardera à l'esprit que deux types d'environnements diffèrent selon les objectifs pour lesquels ils sont conçus au départ, bien qu'il soit malaisé d'arrêter des différences de façon générale (par exemple, au cas par cas, et de par le fait que l'on pourrait évoquer la conception 3D sous une forme ludique, ou encore la conception 3D sous forme d'enseignement).

2.2.1. Environnements Culturels ou de loisir (Entertainment)

Il s'agit essentiellement d'environnements à caractère culturel dédiés à une forme d'enseignement ludique par exploration de la part de l'utilisateur, qui ne peut rien y modifier d'autre que sa position d'observation ("*Viewpoint*") au sein de cet environnement, ex. visite de musée virtuel ou d'environnements reconstitués tels que "Notre-Dame de Paris" avec le projet "*Virtual Reality Notre-Dame*" (cf. Fig. 5 et 6, extraites du site <http://www.vrndproject.com>), environnements interactif 3D tel que le plug-in "Atmosphère" proposé par la société Adobe en 2000, la course à la voile transmise en direct par la société Duran-Dubois la même année, ou encore le survol de terrain de skylinesoft.com, généralement destinés à une classique plate-forme écran-clavier-souris.

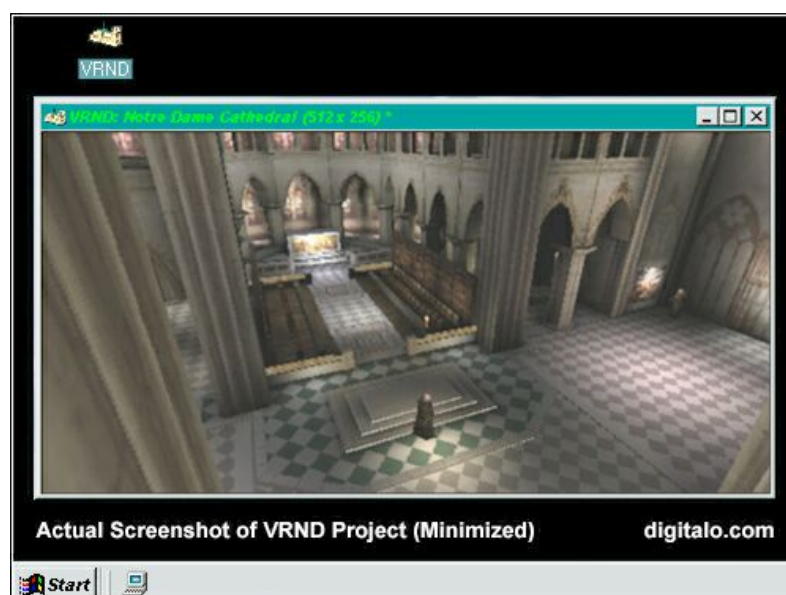


Figure 5. Exemple typique d'un VE à caractère culturel à visiter virtuellement.



Figure 6. Autre exemple d'un VE de visite virtuelle.

Une exception récente est constituée par le projet "ArcheoGuide" présenté au salon Imagina ainsi qu'à la conférence "Laval Virtual" en 2002 (Vlahakis *et al.* 2002). Le résultat de ce projet Européen consiste en un équipement de réalité augmentée portable par l'utilisateur (et visiteur pour l'occasion) sur un site extérieur d'intérêt archéologique (site situé en Grèce, à Olympie en l'occurrence, *cf.* Fig.7). Cet équipement permet au visiteur d'observer sur le site une reconstitution 3D d'un temple s'incrutant comme à l'origine en lieu et place du tas d'éboulis épars que le même visiteur peut observer à l'œil nu.



Figure 7. Incrustation visuelle de la reconstitution d'un temple dans le paysage réel.

Ces environnements, destinés à la seule exploration visuelle, se caractérisent par une "navigation utilisateur" obtenue par un déplacement de caméra(s) relatif(s) par rapport au point de vue ("*Viewpoint*") de l'opérateur. Au sein d'un tel environnement, l'utilisateur n'existe que sous la forme d'un point orienté dans l'espace, et n'y dispose d'aucun moyen d'action lui permettant de modifier l'état dans lequel se trouve l'environnement simulé. L'utilisateur y est donc réduit à un spectateur, qu'il s'agisse d'un environnement simulé statique ou dynamique. Cette catégorie se décline éventuellement par l'existence de quelques comportements préprogrammés ("Behaviors") et activables par l'utilisateur à l'aide des outils d'entrée classiques (souris). De tels comportements activables concernent généralement des changements locaux d'états prédéfinis par scénario. Un objet "porte" s'ouvrira par exemple sur sollicitation par un clic souris de la part de l'utilisateur, (ce que l'on considérera, avec Jacques Tisseau et sa "Métaphore d'Ali-Baba" (2001), comme un objet informatique disposant d'une méthode modifiant un état de l'objet parmi deux valeurs "Ouvverte" et "Fermée"). Considéré de la sorte, ce type d'environnement simulé est inclus dans l'ensemble des hypothétiques environnements de conception 3D qu'un utilisateur devra pouvoir explorer visuellement.

2.2.2. Enseignement

Certains environnements sont dédiés à l'enseignement de tâches à réaliser par la suite en environnement réel (ex. le projet SOFI de la SNCF, la simulation de conduite, la proposition de D. Mellet-d'Huart (2001, 2002), pour l'apprentissage virtuel de la soudure présenté en projet lors de la conférence VRIC 2001. La Fig.8 ci-après illustre un tel exemple d'environnement d'apprentissage de la soudure présenté en démonstration sur le salon de la conférence VRIC 2002, réalisé pour un projet AFPA au sein du consortium "PerfRV". (On y remarquera le "chalumeau" accroché sur la partie droite du portique).



Figure 8. Poste d'apprentissage de soudure virtuelle.

Ces environnements impliquent une ressemblance (avec la réalité), conduisant à simuler le plus fidèlement possible le contexte de l'environnement pour lequel on cherche à produire un enseignement. Ce besoin de fidélité est l'une des raisons qui conduisent à tenter d'enrichir le feed-Back sensoriel destiné à l'utilisateur, non plus seulement du point de vue visuel, mais aussi sonore, et de plus en plus, tactilo-kinesthésique (domaine de "l'haptique").

2.2.3. Jeux

Les applications dans le domaine des jeux présentent davantage de points communs avec un environnement virtuel de conception 3D, en particulier celles

concernant le jeu en environnement virtuel impliquant des modifications de l'environnement virtuel de la part de l'utilisateur, (cf. Starner *et al.*, 2000, ou encore Tamura, 2000), et ce, non plus seulement comme des changements de quelques états prédéfinis.

C'est précisément dans cette capacité fournie à l'utilisateur de pouvoir modifier "l'état de l'environnement" de façon dynamique (*i.e.* avec feed-Back immédiat), que se situe la principale ressemblance. Encore qu'une grande majorité de jeux existants de ce type, (ex. "*Myst*", "*Riven*", "*Amerzone*", etc.), reposent sur des scénarii d'états prédéfinis inspirés du domaine du multimédia dont les techniques reprennent celles du cinéma classique (synopsis, scénario, story-Boarding).

Les Fig. 9, 10 et 11 illustrent un exemple d'environnement de jeu présenté par H. Tamura en 2000 lors du salon Imagina 2000 (Tamura, 2000).



Figure 9. Exemple de scène de jeu non augmentée.



Figure 10. Exemple augmenté de la même Scène de jeu.



Figure 11. Vue de la scène augmentée perçue par l'un des joueurs.

2.2.4. Environnements permettant la modification d'objets graphiques.

On trouvera essentiellement, dans cette catégorie, d'une part les environnements de "*Free Form Modelling*" (ex. l'application "*Geoclay*" mise en œuvre par D. Tonnesen sur WorkBench à l'INRIA-Rocquencourt, PBVE, 2000, cf. Fig.12, qui permet la sculpture virtuelle d'un objet, réalisée à l'aide d'un second objet virtuel tenant le rôle d'outil de déformation du premier), et d'autre part les environnements plus classiques du "*Solid-Modeling*", en particulier ceux issus du récent intérêt du domaine des CAD 3D vers leur adaptation à ces nouvelles techniques (ex. Gao et al.,



Figure 12. "Geoclay", exemple de sculpture virtuelle.

2000). Citons comme exemple encore rare l'une des activités du département IGD A2 du Fraunhofer Institut (<http://www.igd.fhg.de/igd-a2/>) abritant le projet "ARCADE" ("*Advanced Realism CAD Environment*",). Ce projet a pour objectif le développement d'un système CAD 3D coopératif à manipulation directe qui supporte des dispositifs 3D en entrée comme en sortie, (Amicis et al., 2001).

On trouvera d'autre part dans cette dernière catégorie les environnements présentant un Agent conversationnel éventuellement anthropomorphe auquel l'utilisateur peut "demander" de modifier l'agencement d'un mobilier virtuel présenté visuellement, comme le «*MR Living Room*» de H. Tamura (2000), ou encore

³ Computer Assisted Design.

"Gandalf" (Thórisson, 1997), un agent de dialogue inspiré des "Actes de langages" pour communiquer avec l'utilisateur sous forme de dialogue Multimodal et piloter une application de visite virtuelle.

Une autre approche existe sous l'appellation de "*Declarative Modelling*", (Coyne & Sproat, 2001), dont un exemple illustrera le fonctionnement. L'utilisateur produit des énoncés (au clavier dans une fenêtre de saisie), descriptifs de la scène à composer, cf. exemple ci-après.

"The lawn mower is 5 feet tall. John pushes the lawn mower. The cat is 5 feet behind John. The cat is 10 feet tall."

Le résultat fourni à l'affichage consiste en la composition d'une scène peuplée de composants graphiques 3D "traduite" pour ainsi dire par le système, et illustrée, pour l'exemple présenté par Coyne et Sproat en 2001, sur la figure 13 ci-dessous. Ce principe ne s'arrête pas à la description de scènes statiques, il semble aussi expérimenté pour la description de mouvements que peuvent effectuer les objets graphiques.



Figure 13. Exemple illustrant l'approche du "Declarative Modelling".

Présentée dans cet ordre, cette succession de catégories d'environnements visuels 3D permet d'établir un classement en terme croissant d'actions permises à l'utilisateur, convergeant pour la plupart vers la manipulation directe d'Objets Graphiques 3D, avec quelques exceptions en ce qui concerne le "*Declarative Modelling*" ainsi que les études d'interaction utilisateur avec des objets gouvernés par des agents voire des représentations anthropomorphiques d'agents informatiques.

Les collecticiels.

Le sujet des collecticiels ("*Computer Supported Cooperative Work*", CSCW) n'est pas central pour nos travaux, il constitue cependant un domaine d'étude à part entière. Ce thème adressant à lui seul un vaste domaine d'étude, Le "CSCW" ne sera cité ici que pour assurer l'éventualité d'une compatibilité ultérieure au sein d'un plus vaste domaine d'interaction.

On citera par exemple des applications qui permettent l'affichage individualisé d'un environnement 3D de façon cohérente avec une scène réelle pour chacun des participants. L'exemple illustré Figure 14 et Figure 15 ci-dessous (cf. Billingham et

al., 2000), montre l'usage d'un marqueur graphique (réel) permettant au système associé à chaque utilisateur (équipé d'une caméra en entrée et de lunettes d'affichage en sortie) de repérer les position et orientation de l'objet graphique de synthèse à présenter aux yeux de chaque utilisateur.

Billinghurst *et al.*, (1997, 2000), s'intéressent à l'aspect visualisation d'applications 3D collaboratives basées sur HMD (lunettes d'affichage), comparées à des applications 3D collaboratives affichées sur écran-moniteur ("*Monitor-Based*"). Ceci, en mettant plutôt en œuvre une augmentation de la réalité (au sens de Milgram). L'intérêt de ces considérations réside en ce que, par extension, cela apporte quelques arguments en faveur d'une structure collaborative Styliste-Modéliste envisageable lors de l'évaluation du prototype de vêtements à l'aide d'affichages individualisés tels que "*See-Through HMD*".

Les auteurs cherchent ainsi à établir des arguments en faveur de l'idée selon laquelle les équipements individualisés portables ("*Wearable computers*") constituent une plate-forme idéale pour effectuer du collectif en environnement 3D. Ils évoquent les tentatives antérieures d'extension du principe des systèmes 2D au domaine du CSCW-3D (en particulier, projection sur grands écrans, tels que CAVE, WorkBench), en leur reprochant l'affichage d'une stéréoscopie centrée sur un seul utilisateur (un autre spectateur devra se tenir au plus près d'icelui pour garder une vue à peu près correcte de la scène), et en considèrent les systèmes immersifs comme une meilleure approche isolant toutefois encore trop l'utilisateur du monde réel qui l'entoure, ne permettant plus la communication interpersonnelle non verbale. Ce qui leur permet d'introduire leur nouvelle approche utilisant des "*See-Through HMD*" qu'ils baptisent "*Shared Space technique*" parce qu'elle permet la cohabitation et l'activité simultanées de plusieurs utilisateurs effectuant des tâches collaboratives à la fois dans le monde réel et dans le monde virtuel au sein d'un même lieu partagé, tout en autorisant le "*face-to-face communication*" que l'on utilise habituellement dans le cas particulier de nos activités collaboratives "*non mediated*" au sein du monde réel.

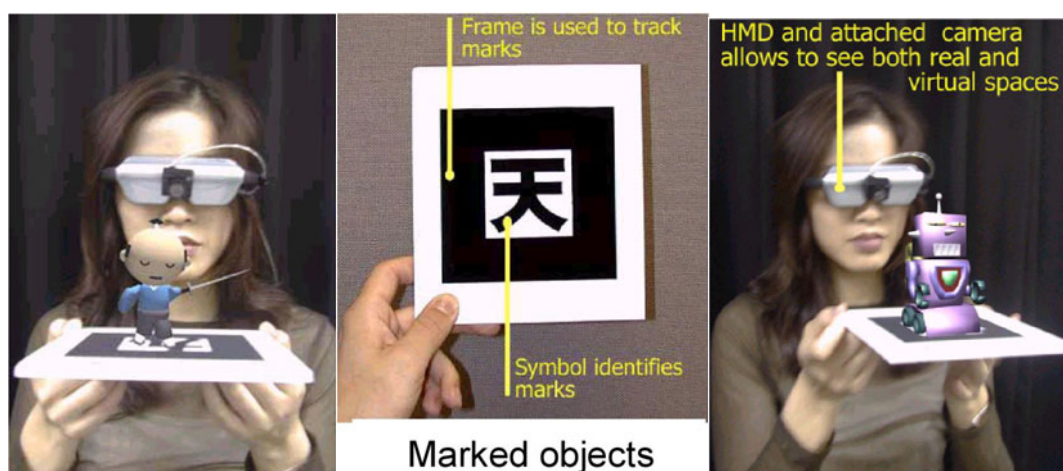


Figure 14. Marqueur graphique permettant de situer l'affichage de l'objet 3D.

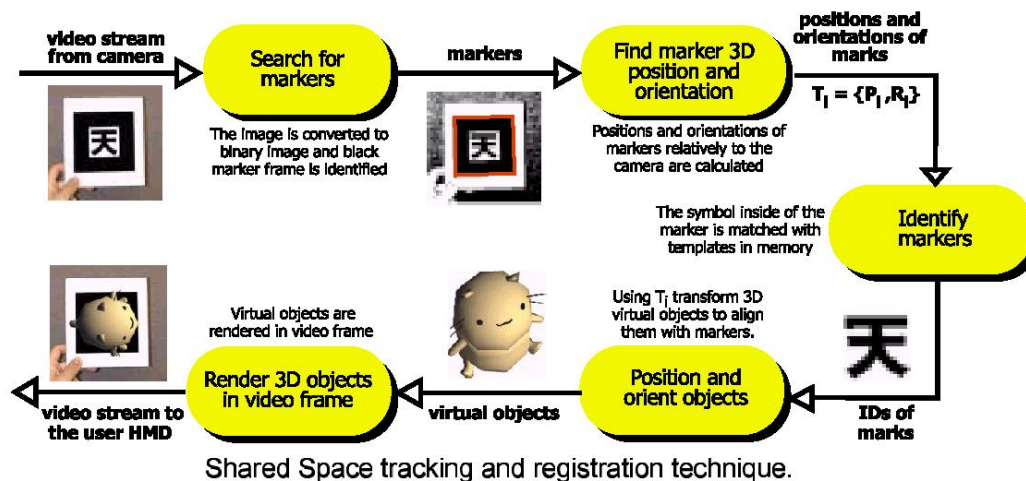


Figure 15. Illustration d'un affichage individualisé de l'environnement réel augmenté.

Spin3D constitue un autre exemple d'interaction virtuelle de plusieurs utilisateurs avec un objet graphique 3D (Dumas, 1999). Cet environnement est centré autour de l'objet de la tâche qui occupe le centre de la scène présentée et permet à chaque utilisateur participant d'intervenir sur le « document » que constitue l'objet 3D de la tâche. L'interaction est basée sur la mise en œuvre d'un pointeur 3D dont dispose chaque utilisateur éventuellement distant. Les problèmes de représentation de la profondeur ont amenés l'auteur à distinguer des indices visuels alternatifs à la stéréoscopie en trois catégories d'indices « permanents » vs « temporaires », « statiques » vs dynamiques » et « passifs » vs « actifs ».

Les textures de surface des objets constituent par exemple un indice permanent et statique, cependant qu'une ombre portée par un objet sur un plancher de référence est un indice permanent, mais dynamique. Les indices temporaires relèvent de l'interaction utilisateur, une « boîte englobante » apparaissant pour encadrer un objet graphique de la scène à l'aide d'une représentation cubique dite en « fil de fer » constituera ainsi un indice temporaire statique. Le terme d'indice « actif » définit un comportement interactif du système en réponse à une action de l'utilisateur, par exemple, l'apparition de boîtes englobantes supplémentaires dites « progressives » se manifestant à l'approche du pointeur.

Chapitre 3. Modalités et dispositifs d'interaction avec les environnements 3D

Les fondements théoriques ainsi que le vocabulaire utilisé dans le domaine de la multimodalité ne sont pas forcément très clairs ni explicites. Il semble que la communauté ne se soit mise d'accord ni sur quelque consensus ou autre standard de fait, ni sur la définition d'un vocabulaire arrêté concernant des notions pourtant déjà anciennes.

Nous allons maintenant établir un bref résumé historique non exhaustif et orienté sur les notions qui nous semblent les plus pertinentes afin de définir un ensemble de notions et un vocabulaire que nous retiendrons pour la suite.

Larson *et al.* (1992) définissent les médias comme des types de donnée qu'ils baptisent "*novel data types*" excluant les types traditionnels tels qu'entier, virgule flottante, caractère, etc. Ces "*novel data types*" distinguent les types texte, image bitmap, illustration (graphique), son, animation et full motion video. Ils définissent d'une part un système multimédia comme celui utilisant plusieurs "*novel data types*", et d'autre part une "*modalité*" comme un langage dont se sert un utilisateur pour entrer des informations dans une machine ou pour en extraire des informations, en ajoutant que les modalités sont habituellement associées à des dispositifs d'entrée-sortie.

Plusieurs modalités peuvent être utilisées pour exprimer une même information. Ainsi, une "interaction multimodale" implique l'usage de plusieurs modalités.

Coutaz, (1992), et Nigay, (1994), décrivent pour leur part comment des interactions multimodales peuvent être catégorisées en *exclusives*, *alternées*, *synergiques*, et *concurrentes* (cf. Figure 16 et Figure 17).

		USAGE DES MODALITES	
		Séquentiel	Parallèle
FUSION	Combiné	ALTERNE	SYNERGIQUE
	Indépendant	EXCLUSIF	CONCURRENT
		Sens / Pas de Sens	Sens / Pas de Sens
NIVEAUX D'ABSTRACTION			

Figure 16. Catégories d'interactions multimodales.

Toujours en 1992, Gourdol *et al.*, établissent qu'une modalité est associée à un dispositif d'entrée-sortie, (clavier, souris, microphone, affichage graphique), permettant à l'utilisateur de communiquer avec un système en utilisant les modalités correspondantes telles que taper du texte, entrer du graphisme, parler, afficher du graphisme. Un système serait multimodal quand il permet l'usage de plusieurs modalités en entrée et en sortie, l'exemple fourni évoque un système muni d'un clavier ("typing modality") et d'une souris ("graphic input modality") comme multimodal.

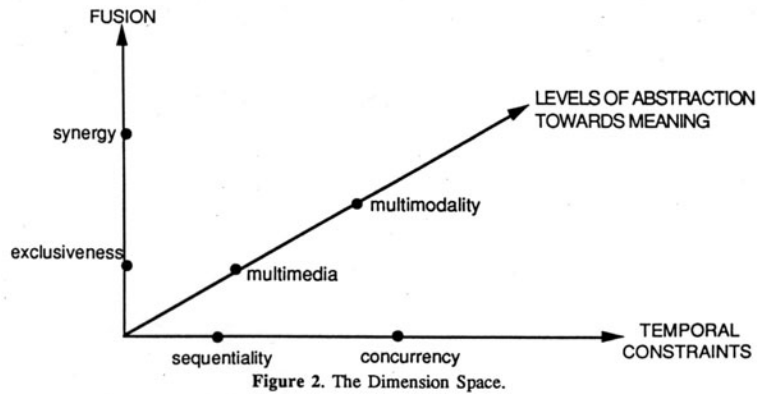


Figure 17. Dimensions impliquées par la multimodalité.

En 1993, Nigay et Coutaz, clarifient le terme de "*système multimodal*" comme supportant la communication avec l'utilisateur au travers de différentes modalités telles que la parole ("*voice*"), la gestuelle ("*gesture*") ou encore l'entrée de texte au clavier ("*typing*"). Les auteurs précisent le terme en le découpant en "multi" (plus d'un), et "modal", recouvrant aussi bien la notion de "modalité" que celle de "mode".

Une *Modalité* se référerait à un type de canal de communication utilisé pour convoyer ou acquérir de l'information, le terme recouvrirait aussi la façon dont une idée est exprimée ou perçue, ou encore la manière dont une action est effectuée.

Le *mode*, lui, se rapporterait à un état qui détermine la façon dont l'information est interprétée pour extraire ou convoyer une signification. La modalité définirait par ailleurs le type de données échangées alors que le mode déterminerait le contexte dans lequel les données sont interprétées.

Les auteurs baptisent leur point de vue de "*system-centered view*". La multimodalité définirait ainsi la capacité du système à communiquer avec l'utilisateur au moyen de différents types de canaux de communication *et* d'extraire ou convoyer automatiquement de la signification ("*meaning*"). Pour mieux préciser les choses, il est rapporté que les systèmes multimédias et multimodaux utilisent plusieurs types de canaux de communication, mais que ce qui les différencie, est la capacité d'un système multimodal à modéliser automatiquement le contenu de l'information à un haut niveau d'abstraction. Les auteurs expliquent la raison pour laquelle elles précisent que leur point de vue est "*system-centered*", en ajoutant que la perspective "*user-centered*" pourrait conduire à une définition différente. L'exemple cité rapporte qu'un certain logiciel de gestion de mails ("*the NeXT voice electronic mail*") peut, du point de vue "*user-centered*", laisser apparaître le système comme multimodal du fait qu'il permet d'inclure texte, graphisme et messages vocaux dans un mail, mais que d'un point de vue "*system-centered*", le système reste uniquement multimédia du fait qu'il n'extrait aucune "interprétation" des informations qu'il transmet, qu'en particulier, les messages vocaux sont enregistrés, mais non interprétés.

Bernsen et les modalités de sortie.

N. O. Bernsen, qui a publié en 1994 une taxonomie des *modalités de sortie*, émet un raisonnement systématique en distribuant toutes les possibilités binaires de quatre propriétés principales relevant de la communication pour obtenir un tableau de modalités "unimodales" dans lequel il élimine ensuite les combinaisons aberrantes pour n'en retenir finalement que vingt-huit. Cherchant ainsi à faire ressortir un niveau atomique permettant de caractériser une modalité sans se baser sur d'éventuels dispositifs d'entrée-sortie existants, il partage d'abord la distribution en catégories de modalités "Linguistiques" et "Non Linguistiques", puis "Analogiques" vs "Non Analogiques", "Arbitraires" vs "Non Arbitraire" et "Statique" vs "Dynamique". Il ajoute à cela trois propriétés supplémentaires sans toutefois leur opposer leur complément: "Graphiques", "Sonores" et "Tactiles". (La distribution comprend donc ainsi $24 \times 3 = 16 \times 3 = 48$ combinaisons). Il insiste, plutôt que sur des définitions relatives aux modalités dont les contours n'apparaissent qu'en fin de raisonnement, sur des concepts tels que celui de "représentations externes" auxquelles il limite volontairement l'étude, qui implique une différence entre ce qui est représenté et sa représentation. Il redéfinit de même le terme de "*modalité représentationnelle*" de départ pour son raisonnement, (trop vague dans la littérature à son goût), en énonçant d'abord que des tableaux, des bips sonores, le langage écrit ou encore la langue naturelle parlée peuvent être qualifiées de 'modalités', que les "modalités représentationnelles" ne devraient pas être confondues avec les "modalités sensorielles" de la psychologie, que les modalités elles-mêmes peuvent être unimodales ou multimodales, et enfin qu'une taxonomie de modalités est une façon de trouver un découpage, au sein de la multitude des représentations possibles d'une information. Ce point de vue se fonde sur l'observation du fait que différentes modalités de représentation présentent différentes propriétés qui les rendent aptes à représenter différents types d'information. Pour limiter l'étendue de la taxonomie envisagée, l'auteur distingue d'autre part des *médias d'expression* de base tels que le graphisme, le son et le toucher qui se caractérisent par des ensembles de qualités perceptuelles (plus précisément "utilisateur", cette fois) très différentes (respectivement visuelles, auditives et tactiles). Après examen des classes ou sous-ensembles de propriétés selon lesquels se regroupent les modalités unimodales de représentation externe, Bernsen les définit finalement d'abord de façon générique en établissant qu'une modalité se caractérise par un médium spécifique d'expression (qui ne lui est pas exclusif) et par un profil de propriétés telles que précédemment définies. Une définition plus générale du terme de 'modalité' incluant les représentations tant unimodales que multimodales consiste ensuite en ce qu'une modalité dispose d'au moins un médium d'expression et d'un profil de propriétés qui comprend au moins une propriété au sein de la liste de compléments binaires suivante : "*analogue/non-analogue, arbitrary/non-arbitrary, static/dynamic, linguistic/non-linguistic*".

Les notions relatives à la "*MultiModalité*" semblent donc s'être dégagées par un besoin de distinction avec les notions relatives au "*MultiMédia*". Nous retiendrons donc, par rapport au système, l'existence de modalités aussi bien en entrée qu'en sortie du système associées à un ou plusieurs canaux de communication. On retiendra aussi la notion selon laquelle l'usage d'une modalité implique une forme "d'interprétation" des données transmises, (ce qui ne sera pas le cas en ce qui concerne le "Multimédia" impliquant plutôt une restitution). D'autre part, le fait que plusieurs modalités transitent éventuellement via un même dispositif suggère, notamment en entrée, des traitements qui doivent permettre de les distinguer par le système.

Au cours de la section suivante, nous allons aborder diverses catégories de modalités d'entrée et de sortie du système qui sont apparues les plus importantes vis à vis de notre domaine d'intérêt. Nous aborderons d'abord les modalités de sortie, parmi lesquelles les affichages visuels semblent les plus importants, nous évoquerons ensuite rapidement les modalités relatives au rendu sonore, puis celles qui concernent le sens du toucher en général ("tactilo-kinesthésique"). Nous aborderons ensuite les catégories de modalités d'entrée les plus importantes, en particulier les diverses "gestuelles", avant de mentionner les modalités dites "vocales".

Pour chacune de ces catégories, nous mentionnerons d'abord un aspect théorique incluant les facteurs humains les plus importants à retenir, avant d'exposer divers dispositifs techniques aptes à tenir le rôle de medium d'information.

3.1 Modalités de sortie

Evoquer l'affichage comme modalité principale de sortie tient d'une part au fait que la vision prend une place prépondérante concernant des tâches de conception d'artefacts. Ce choix tient d'autre part à certaines considérations issues de l'étape d'analyse lors de la phase d'évaluation de prototypes vestimentaires, révélant en l'occurrence, pour de tels objets présentant un caractère esthétique, une inspection de prototype essentiellement basée sur l'aspect visuel qu'ils présentent.

3.1.1. Modalités de sortie visuelle (affichages)

Eléments Psychophysiques des systèmes d'affichage 3D.

La perception visuelle d'un objet 3D se complique chez l'homme de contraintes dont il serait souhaitable de tenir compte pour un usage quotidien sur de longues périodes non seulement en ce qui concerne la cohérence visuelle de ce qui est présenté à l'utilisateur via divers dispositifs d'affichage, mais aussi en ce qui concerne la notion de "confort", qui adresse également les risques encore peu connus qu'encourt un utilisateur exposé de façon prolongée à des dispositifs impliquant une adaptation visuelle anormale. Nous allons rapidement exposer ici les notions qui semblent les plus importantes.

Le "Motion Parallax"

M. McKenna évoque en 1992 l'importance du phénomène décrit sous le nom de "*Motion Parallax*" pour les affichages 3D de type moniteur, qui s'applique aussi pour les systèmes de projection (McKenna, 1992). L'article constitue une bonne introduction aux avantages et inconvénients du "*Motion Parallax*", en décrivant la nécessité d'une distorsion contrôlée de l'image affichée sur écran en fonction du déplacement des yeux (donc, de la tête) de l'utilisateur, lorsqu'il n'est plus dans l'axe normal de visualisation (qui est, pour un moniteur classique, la normale au centre de la surface de l'écran). La correction de perspective en fonction de la position des yeux de l'utilisateur par rapport à un écran fixe permet ainsi une visualisation plus fidèle à ce qui se passe dans la réalité lors de l'exploration d'un objet 3D *autour* duquel, notamment, une telle exploration est pratiquée, (ce qui aurait peut-être moins d'importance dans le cas d'une scène immersive, encore que le fait qu'on se serve d'objets 3D comme autant d'indices de la scène -les "*LandMarks*"-, puisse jouer un rôle). A noter que l'auteur rappelle que la perspective et le "*motion parallax*" constituent des *indices monoculaires intéressant la profondeur*, et qu'en conséquence, la sensation de profondeur qu'on en extrait ne requiert guère d'une part qu'un seul œil, et d'autre part qu'un affichage 2D.

Le terme "*Motion Parallax*" est ici défini comme le mouvement apparent d'un objet 3D relativement aux déplacements du point de vue du spectateur. La Figure 18 ci-après en illustre le principe de façon exagérée.

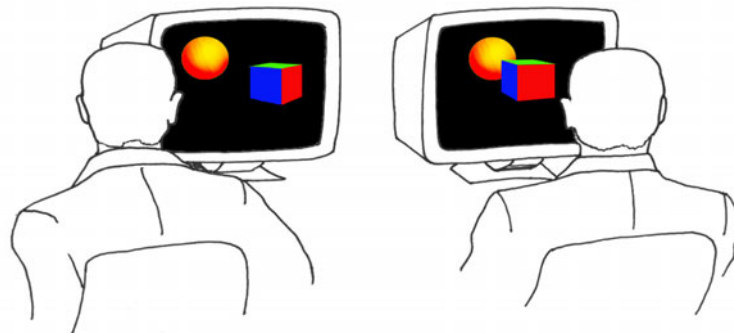


Figure 18. Correction de perspective en fonction de la position de l'observateur.

Les indices binoculaires

Trois indices binoculaires semblent particulièrement importants pour l'examen d'une scène en trois dimensions. Le "*stéréopsis*", la "*convergence*" et "*l'accommodation*".

Le "*stéréopsis*" mis en œuvre lors de l'examen d'une scène statique est le plus important des trois indices *binoculaires* intéressant la perception de la profondeur, et repose sur la disparité entre les deux images perçues respectivement par chacun des deux yeux. La Figure 19 ci-dessous illustre ce que percevrait chacun des deux yeux d'un spectateur observant la scène.

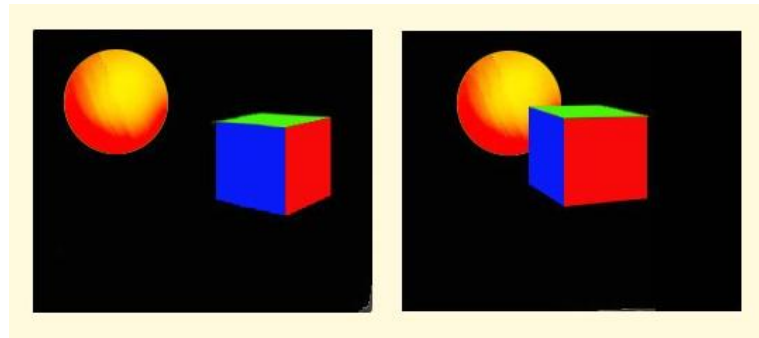


Figure 19. Disparité visuelle illustrant le stéréopsis.

L'usage de la stéréoscopie n'est réellement utile qu'en deçà d'une distance au-delà de laquelle chacun des deux yeux de l'observateur n'interprète plus qu'une même image. Avec une résolution moyenne chez l'homme au niveau de la fovéa avoisinant la minute d'angle, une approximation de cette distance pour un écart interoculaire moyen de 7cm fournit une distance d'environ 250m (écart interoculaire / [sinus (résolution)] = 0,07m / [sin(1°/60)] = 240m).

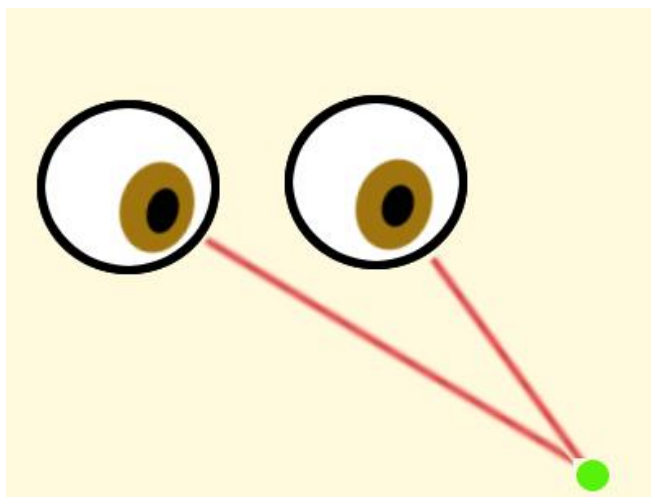


Figure 20. Convergence oculaire.

La "*convergence*" réfère au fait (Eysenck & Keane, 1998), que les yeux effectuent, pour focaliser sur un objet, une rotation vers l'intérieur plus importante pour un objet proche que pour un objet lointain, (cf. Figure 20).

Quant à "*l'accommodation*", elle consiste en la variation de vergence de l'œil par épaissement du cristallin (sur sa face extérieure), lors de l'examen d'un objet rapproché, (l'œil est au repos pour les objets situés à l'infini).

Ces trois indices binoculaires fonctionnent ensemble au sein du système nerveux humain, si l'on présente un objet à 30 cm des yeux d'un spectateur dont l'un des deux est masqué, ses deux yeux convergeront sur cet objet et accommoderont à la distance de 30 cm.

Un affichage dit "*stéréoscopique*" implémente habituellement le "*stéréopsis*" et la "*convergence*" en calculant deux images distinctes destinées respectivement à chacun des deux yeux, mais n'implémente pas "*l'accommodation*" du fait que la surface d'affichage reste à une distance constante de l'Observateur. Lorsqu'un spectateur observe un tel affichage, la perception d'une image nette implique une accommodation sur la surface de l'écran, à une distance inhabituelle sauf dans le seul cas où l'objet 3D affiché est virtuellement présent au voisinage de la surface de l'écran.

Tant que l'objet affiché est présenté virtuellement à la même distance que l'écran, "l'*accommodation*" du spectateur correspond à peu près à ce qui se passerait dans la réalité. Si l'on modifie la distance virtuelle de l'objet, (plus près ou plus loin que la surface réelle d'affichage), le spectateur devra continuer à accommoder à la surface de l'écran pour conserver la netteté de l'image, cependant que les deux autres indices binoculaires recevront une information de distance différente, introduisant ainsi une incohérence de la scène responsable de l'inconfort alors perçu.

Afin de donner une idée de ce qu'éprouve un spectateur soumis à une telle incohérence d'accommodation, il existe un exemple graphique 2D issu du domaine de l'illusion d'optique reproduit ci-dessous (Figure 21, Illustration extraite de Science & vie, Hors-Série 216, sept 2001, "L'œil et la Vision"), qui, en différents endroits observés, possède un indice de texture correspondant à différentes distances auxquelles l'œil va accommoder.

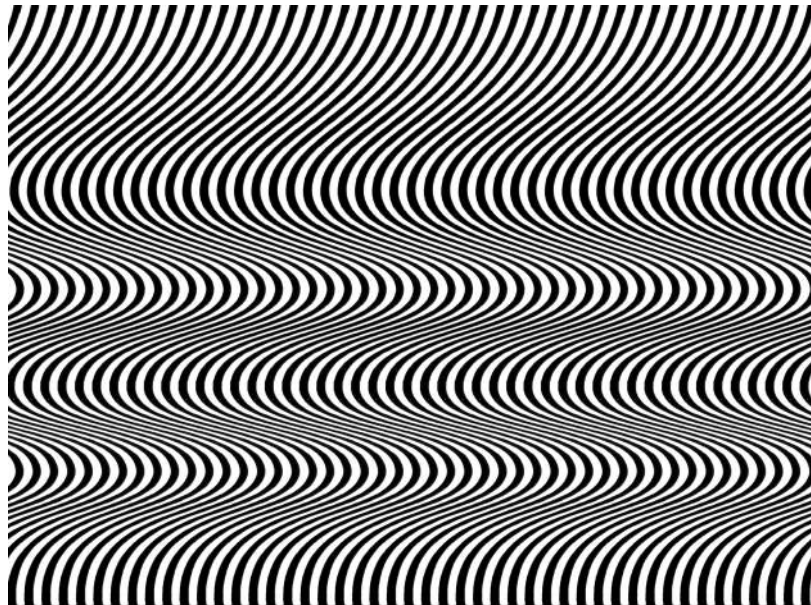


Figure 21. Exemple perturbant l'accommodation oculaire.

Des mesures effectuées quant au confort "acceptable" par l'utilisateur au sujet de cette incohérence de l'indice "Accommodation" relativement aux informations fournies par les deux autres indices binoculaires (disparité oculaire et convergence), ont fourni un intervalle de distances virtuelles acceptables comprises entre -25% (vers l'avant) de la distance utilisateur-Ecran et +60% au-delà de la surface d'affichage (Williams & Parrish, 1990), (Ware *et al.*, 1998). De part et d'autre de cet intervalle, l'affichage reste encore possible mais devient de plus en plus inconfortable et sujet à erreur d'évaluation des distances, et peu envisageable sur de longues périodes de travail. (*cf.* Figure 22 illustrant l'intervalle de profondeur "acceptable" d'affichage stéréoscopique d'objets sur écran en l'absence de traitement de l'accommodation).

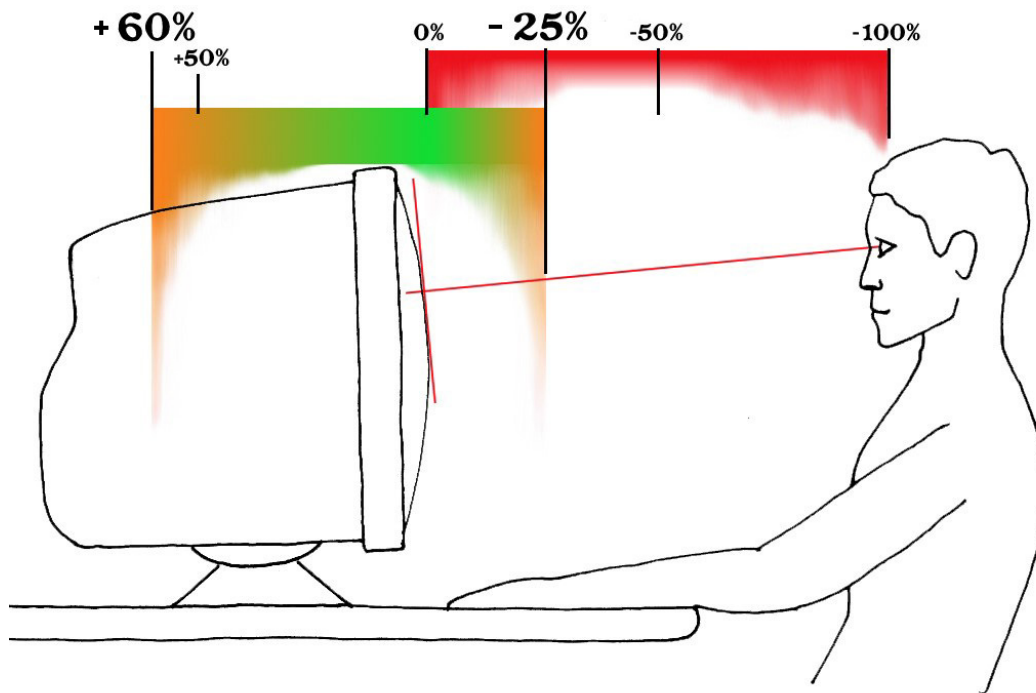


Figure 22. Intervalle de profondeur "acceptable" d'affichage stéréoscopique d'objets entre -25% et +60% de la distance à la surface d'affichage.

En 1993, Ware *et al.* expérimentent un dispositif baptisé "*FishTank*". Textuellement, "bac à poisson", traduisant la notion d'un aquarium dont un affichage sur moniteur en figure une des faces. Les auteurs caractérisent leur "*FishTank*" par l'affichage sur moniteur de l'image stéréoscopique d'une scène 3D en utilisant une projection de perspective pilotée par la position de la tête de l'observateur, et effectuent une investigation expérimentale concernant deux facteurs:

1. selon qu'une déformation de la perspective affichée est ou non couplée à la position de la tête de l'observateur.
2. selon que le stéréopsis est mis en œuvre.

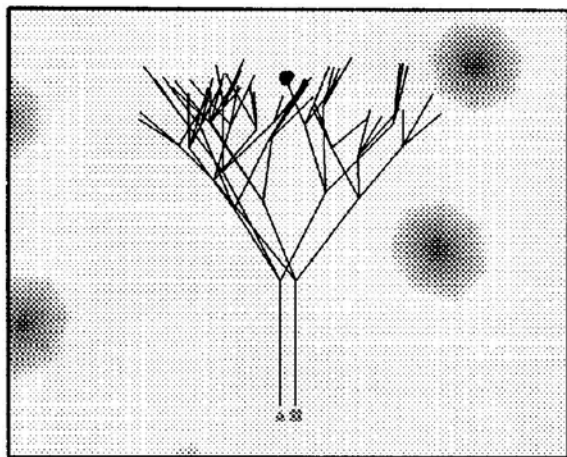


Figure 23. Le test utilisateur consiste à retrouver la racine portant la "feuille" représentée.

Les résultats évoqués concernent un test utilisateur impliquant le tracé correct, parmi deux arborescences entrelacées du chemin séparant une feuille ("leaf") de sa racine correspondante (*cf.* Figure 23).

Situation Mise en Oeuvre	Taux d'erreur Observé
affichage monoscopique	22%
affichage stéréoscopique	14,7%
uniscopie asservie à la User head position	3,2%
stéréoscopie asservie à la User head position	1,3%

Figure 24. Les résultats du test utilisateur.

Les résultats de ce test suggèrent (*cf.* Figure 24), qu'il est plus efficace, d'un point de vue perceptuel, de disposer seulement d'une image monoscopique couplée à une implémentation du "*Motion Parallax*" que de disposer seulement d'une image stéréoscopique.

Ils évoquent aussi le problème de la profondeur de champ qui nous permet dans le monde réel de conserver la netteté sur un objet par rapport à l'arrière-plan ou à d'autres objets disposés à d'autres distances. En l'absence de mesure de la focale des cristallins de l'observateur, il n'est pas possible en VR de modifier la netteté de l'arrière-plan, avec le FishTank, toutefois, (ou autre système de projection), la scène de travail étant nécessairement contrainte géographiquement entre quelques cm en avant et quelques cm en arrière de la surface d'affichage de l'écran, il est alors concevable de simuler un effet de profondeur de champ en affichant un arrière-plan (flou) pour simuler le phénomène de "*Vection background*".

Le terme "*vection*" est associé avec la sensation de déplacement que l'on éprouve lorsque c'est en fait l'environnement qui se déplace (ex. quand on est dans un train à l'arrêt et qu'on voit par la fenêtre un train voisin qui se déplace), il a souvent été rapporté, selon les auteurs, que ce sont les stimuli périphériques et distants d'un vaste champ de vision qui sont responsables de cet effet. La transposition au FishTank consiste à générer une image de fond composée d'objets circulaires situés à l'infini, donc flous pour donner l'illusion "d'out-of-focus", baptisée "*Vection Background*".

Autre problème, la stabilité lors de mouvement des yeux. En VR immersive, les yeux sont très proches des écrans pour obtenir un champ de vision suffisamment étendu, ce qui introduit une erreur lorsque l'œil bouge de l'axe normal à l'écran. Selon une information rapportée par Deering (1992), l'œil effectue des rotations autour d'un centre géométrique, lequel se trouve 6mm en arrière du centre optique (le foyer). Ce qui signifie qu'un mouvement oculaire de 40° impliquera une translation du centre de la pupille de 3,8mm. Couplé avec un écran à 8cm de l'œil, il en résultera un changement de position virtuel de quelques 3,8cm pour un objet qui serait situé à 80cm de distance. Cet effet ne peut être corrigé (pour modifier l'affichage en conséquence), qu'en mesurant les mouvements des yeux de l'observateur (on notera pour la suite qu'il y a là un argument en faveur de "*eye-tracking*" qui trouverait alors

un emploi supplémentaire en se voyant intégré à un affichage HMD). Le même effet existe en "FishTank VR", mais se trouve réduit à un déplacement de 3,8mm pour un objet qui se situerait à 80cm s'il est proche de la surface de l'écran.

Effet secondaire conséquent à l'usage du stéréopsis sur écrans plats tels que moniteurs.

Un autre phénomène à retenir en stéréoscopie projetée de surface réduite consiste en ce que de tels dispositifs d'affichage jouent le rôle de "fenêtre" ouverte sur un espace 3D, dont les bords externes cachent les objets représentés en arrière et hors de la surface d'affichage, de la même façon qu'une fenêtre normale donnant vue sur un extérieur. Des objets qui sont représentés en avant de la surface d'affichage sont

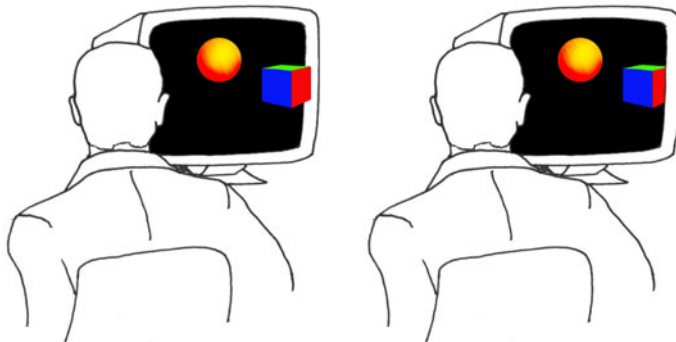


Figure 25. Illustration du phénomène de "Cancellation Effect" (ou "Window Violation").

cependant éventuellement eux aussi partiellement cachés par le bord de l'écran qui serait pourtant censé se trouver derrière ces objets, (Le "Cancellation effect" de Valyrus, 1966, cf. Ware et al., 1998). Selon McKenna qui rapporte ce phénomène auquel il donne le nom de «window violation», (McKenna, 1992), la sensation de perception de la profondeur peut, lorsque se

manifeste une telle incohérence, disparaître aussi bien sur un affichage monoscopique que stéréoscopique.

La Figure 25 illustre le problème qui peut survenir sur un dispositif d'affichage délimité par un cadre en avant duquel est virtuellement affiché un objet graphique 3D. Tant que les deux vues de l'objet graphique le circonscrivent entièrement à l'intérieur du cadre, le problème ne se pose pas (exemple, la sphère sur la même figure). Lorsque ce ne sera plus le cas, on voit à gauche ce que l'observateur devrait observer (le cube en avant de l'écran). Sur la figure de droite, le même cube dont la stéréoscopie informe l'observateur qu'il se situe en avant de la surface d'affichage subit une obstruction par le bord de l'écran supposé se trouver en arrière du cube, d'où une incohérence produisant un conflit visuel.

Effet du type de rendu utilisé pour l'affichage.

En ce qui concerne la présentation à l'affichage, le "rendu photoréaliste" semble souvent admis comme le plus adapté à toutes les situations. Si un tel rendu de qualité maximale est effectivement un objectif à atteindre pour la composition de "prise de vue" cinématographiques ou autre application simulant au mieux la réalité, ce n'est

pas forcément le cas pour des activités de conception, par exemple, d'objets manufacturés.

S. Feiner (2002) ayant fait remarquer, lors de sa présentation à la conférence de Laval, l'importance d'un rendu graphique de scène 3D permettant aux utilisateurs de percevoir plus nettement les angles ou protubérances sur un volume au sein duquel ils devront par exemple se déplacer virtuellement, c'est aussi, d'un autre point de vue, partiellement l'objet de l'une des recommandations de K. Kaur (Kaur, 1998).

Dans un domaine similaire, certains chercheurs se consacrant, dans une branche particulière du "Computer Graphics", à la représentation graphique de formes de façon non-photoréaliste, justifient en grande partie l'intérêt de leurs travaux en avançant que la représentation graphique sous forme d'illustration artistique ou technique permet de mettre en valeur certaines caractéristiques des objets représentés, que l'on ne perçoit pas forcément sur une vue photoréaliste. Pour citer l'un de ces chercheurs qui introduit le sujet mieux que nous ne saurions le faire nous-même :

(Gooch, 1998, introduction, chapter one.).

" Imagery generated by artists provides information about objects that may not be readily apparent in photographs or real life. The same goal should apply to computer-generated images. This is the driving force behind non-photorealistic rendering. The term non-photorealistic rendering (NPR) is applied to imagery that looks as though it was made by artists, such as pen-and-ink or watercolor. Many computer graphics researchers are exploring NPR techniques as an alternative to realistic rendering. More importantly, non-photorealistic rendering is now being acknowledged for its ability to communicate the shape and structure of complex models. Techniques that have long been used by artists can emphasize specific features, expose subtle shape attributes, omit extraneous information, and convey material properties. These artistic techniques are the result of an evolutionary process, conceived and refined over several centuries. Therefore, imitating some of these techniques and exploring the perceptual psychology behind technical illustration are good first steps in going beyond realistic rendering".

Plusieurs chercheurs se consacrent à l'étude de ce type de rendu graphique technique plus adapté au domaine de la conception d'Objets manufacturés, type de rendu qui est d'ailleurs l'objet, depuis l'an 2000, d'une conférence ACM bisannuelle ("*NPAR conference*").

D'un point de vue perceptuel utilisateur pour le domaine qui nous occupe, les environnements de conception et les interactions qui peuvent y prendre place, Schumann *et al.* (1996) ont étudié les effets du rendu non-photoréaliste "*sketch-like*" de représentations produites par un CAD auprès d'un échantillon d'architectes, dont les résultats suggèrent l'usage d'une telle représentation comme médium de choix pour un document modifiable car elle semblerait susciter des modifications collaboratives qui ne semblent pas se produire sur un document tel qu'issu d'une table traçante.

Freudenberg *et al.* proposent en 2001, un procédé de rendu visuel exploratoire d'un environnement virtuel basé sur un moteur de jeu modifié pour fournir un rendu d'aspect crayonné ("*pen-based*") ou aquarelle ("*pen-and-ink-based*"). Les illustrations ci-après donnent un aperçu de différents traitements sur un même objet dont l'aspect détourné rappelle les techniques utilisées par les auteurs de "Bandes Dessinées". Ces techniques ont évolué au cours du 20ème siècle pour aboutir à des résultats d'imagerie de scène dont seuls les éléments importants pour l'objet et la cohérence de scène sont clairement visibles, distingués les uns des autres et mis en valeur (*cf.* Figure 26-Figure 27).

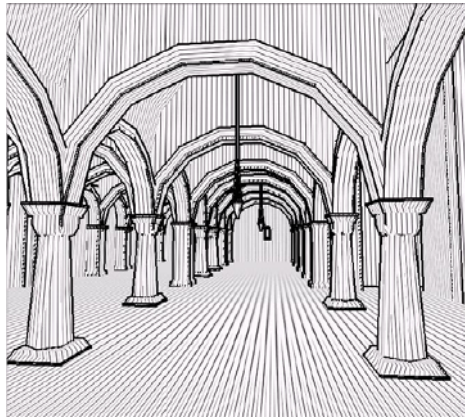


Figure 26. Un intérieur médiéval avec rendu de style "non-photoréaliste".

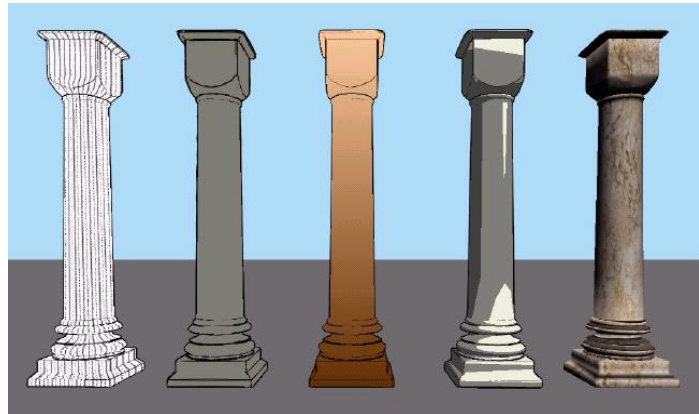


Figure 27. Différents styles de rendus graphiques non-photoréalistes d'un même objet graphique.

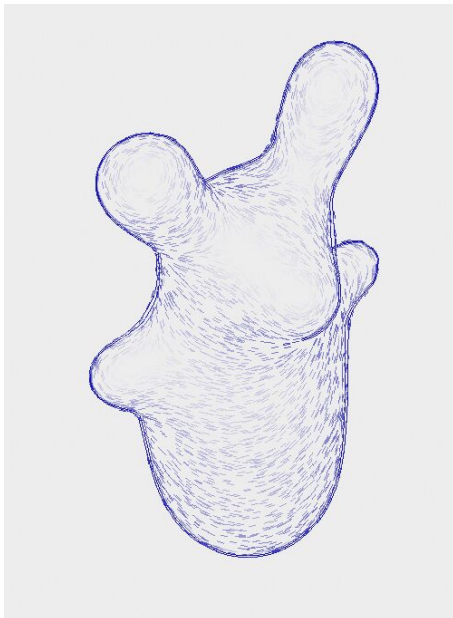


Figure 28. Exemple de rendu non-photoréaliste.

On peut aussi constater, en plus du dernier exemple évoqué, que certaines évolutions récentes utilisent des implémentations de modifications de moteurs de rendu existants pour obtenir un rendu NPR ("*Non Photorealistic Rendering*"). Bremer & Hugues proposent ainsi (1998), un ensemble de class-files reposant sur l'API Java3D de SUN afin de permettre un rendu au contour texturé de formes 3D non régulières. L'illustration ci-après, (Figure 28), en fournit un exemple qui montre ce que l'on peut ainsi obtenir avec un objet comportant des parties convexes.

Enfin, Gooch & Willemsen, (2002), semblent être les premiers à évaluer la perception spatiale de l'utilisateur basée sur un rendu non-photoréaliste en réalité augmentée immersive pour une tâche d'exploration.

Principales catégories de dispositifs d'affichage aptes à représenter des Objets Graphiques 3D

Tubes Cathodiques Classiques.

Ecrans Monoscopiques "Moniteur"

Mis à part son actuel usage pour une 3D aplatie sur la surface d'écran, il existe une tendance à estimer l'inaptitude du "moniteur" pour la 3D, il reste cependant d'autant plus utilisable dans certains contextes (ex. les WOW, "Windows on the World"), si l'on considère les arguments de Ware *et al.* (1993) évoqués au chapitre précédent, selon lesquels il sera plus efficace de disposer seulement d'une image monoscopique couplée à une implémentation de la déformation de perspective en fonction des mouvements de la tête de l'utilisateur ("Motion Parallax"), que de disposer seulement d'une image stéréoscopique.

Ecrans Cathodiques Classiques en Stéréoscopie

Ce type de configuration ne permet, utilisé avec la seule stéréoscopie, un rendu réaliste que si l'on reste en face de l'écran à une distance déterminée. Elle permet néanmoins le rendu de la profondeur si l'on respecte ces conditions, abstraction faite du problème que pose l'accommodation.

Son principe repose habituellement sur l'usage de lunettes dites "shutters" qui ne laissent voir qu'une image sur deux à chaque œil du spectateur et impliquent en conséquence des fréquences suffisantes de rafraîchissement de l'affichage pour éviter l'effet gênant connu sous le nom de "scintillement".

Les capacités de tels dispositifs d'affichage en limiteront sans doute l'usage pour des applications ne concernant que des petits objets (inférieurs aux dimensions de l'affichage) si l'on tient à la stéréoscopie.

Systèmes de projection

Les systèmes de projection se distinguent des moniteurs par le fait qu'ils utilisent un projecteur physiquement dissocié de l'écran sur lequel sera projetée l'image. Deux types de projection sont possibles, la projection directe, interdisant à l'observateur de se mouvoir entre le projecteur et la surface écran dont il occulterait alors une partie de l'image, et la rétroprojection qui lui permet de se rapprocher de l'écran translucide mais imposant un espace de projection supplémentaire de l'autre côté de la surface d'affichage.

Nous allons ci-après évoquer plusieurs exemples qui se distinguent par le type de projection utilisé, le nombre d'écrans mis en œuvre et la forme des surfaces de projection.

Projections sur plan(s).

Bien que rarement évoquée, la projection directe sur une ou plusieurs surfaces planes reste envisageable, en particulier lorsque le ou les observateurs sont consignés à une position d'observation telle qu'illustrée sur la Figure 29 ci-dessous (Raskar *et al.* 1998). Cette solution permet d'économiser le volume qui serait requis derrière la surface d'affichage pour obtenir une solution équivalente par rétroprojection. En contrepartie, aucun mouvement de l'utilisateur ne doit venir obstruer le faisceau du projecteur

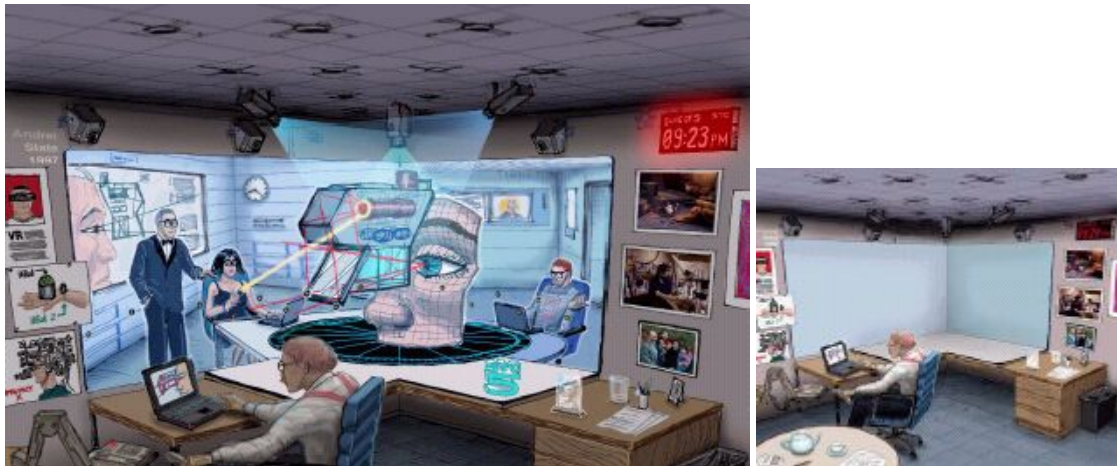


Figure 29. Illustration d'un "Bureau futuriste" impliquant la projection directe.

Dans la catégorie des techniques de rétroprojection, le "Mur de Projection" sur unique écran plat de grande taille constitue un exemple exploitable avec mise en oeuvre de la déformation de perspective et de la Stéréoscopie.

La principale amélioration par rapport aux moniteurs classiques consiste en une plus grande surface de travail et une réduction du problème d'accommodation à plusieurs mètres de distance tant que les Objets Graphiques 3D sont représentés au-delà de l'écran. Cependant, l'image reste calculée pour un unique "ViewPoint", donc, pour un petit nombre d'utilisateurs rassemblés autour du point de vue pour lequel est calculée l'image ("viewpoint"). Par ailleurs, le problème de l'accommodation subsiste à courte distance. Enfin, l'écran étant délimité, le "Window Violation" reste possible (cf. § 5.1.3).

Cas particulier de la table de projection, exemple collaboratif.

Kitamura *et al.*, (2001), ont récemment présenté un affichage stéréoscopique collaboratif rétroprojeté (pour quelques utilisateurs seulement, cf. Figure 30).

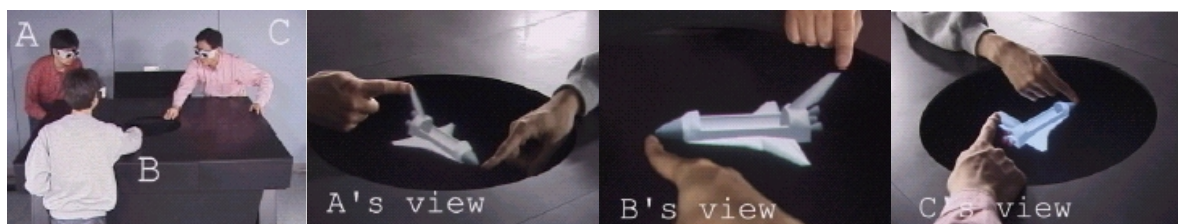


Figure 30. Exemple d'affichage collaboratif rétroprojeté sur écran unique.

Cette "table virtuelle" est un cas particulier que l'on classe également ici dans cette catégorie "Unique écran d'affichage". L'intérêt de l'exemple cité ci-dessus réside dans son usage autorisant la présence de plusieurs utilisateurs disposant chacun de son propre point de vue ("*view-point*", en réduisant toutefois notablement la surface d'affichage utilisable), alors qu'un tel dispositif ne permet jusqu'à présent de n'afficher des images calculées que pour un unique "*viewpoint*" destiné soit à un unique observateur, soit à un groupe d'observateurs rassemblés autour de l'unique "*viewpoint*".

Le principe de la rétroprojection peut également s'effectuer sur deux écrans orthogonaux synchronisés. Le cas le plus connu est sans doute celui du "*WorkBench*", qui peut se décrire comme une extension de la "table virtuelle" qui vient d'être



Figure 31. Exemple d'affichage rétroprojeté sur deux écrans.

évoquée à laquelle on ajoute un second écran qui fait face à l'observateur (cf. Figure 31, modifiée avec ajout de contour de couleur rouge délimitant les surfaces d'écran).

Un tel dispositif se compose de deux écrans plats rétroprojetés de taille moyenne et synchronisés pour la stéréoscopie, avec implémentation de la déformation de perspective. Les écrans sont disposés perpendiculairement l'un par rapport à l'autre en sorte que l'utilisateur fait à la fois face à une table virtuelle qui s'étend devant lui ainsi qu'au second écran disposé de façon plus classique face à lui. Cette disposition permet d'obtenir une plus grande surface d'affichage avec simulation d'un volume "en avant" des écrans, mais ne permet encore une fois l'affichage que pour un unique "*viewPoint*", donc, pour un seul observateur, éventuellement un groupe de quelques personnes rassemblées autour du "*viewPoint*". Par ailleurs, le problème de l'accommodation visuelle subsiste car l'observateur doit accommoder en permanence à moins de deux mètres avec des représentations d'Objets Graphiques 3D situés soit plus près, soit plus loin. Enfin, L'écran étant délimité, le "*Window Violation*", encore appelé "*Cancellation Effect*" reste possible.

Une orientation différente des deux écrans rétroprojetés commence à se démocratiser pour occuper deux murs contigus comme précédemment illustré en projection directe pour le "bureau du futur" décrit par Raskar *et al.* (1998). Le "*Wedge*" (cf. Figure 32), est ainsi présenté par "l'Australian National University" (Canberra, <http://escience.anu.edu.au/research/index.en.html>), comme une solution économique comportant deux écrans stéréoscopiques rétroprojetés apportant au regard une surface davantage panoramique.

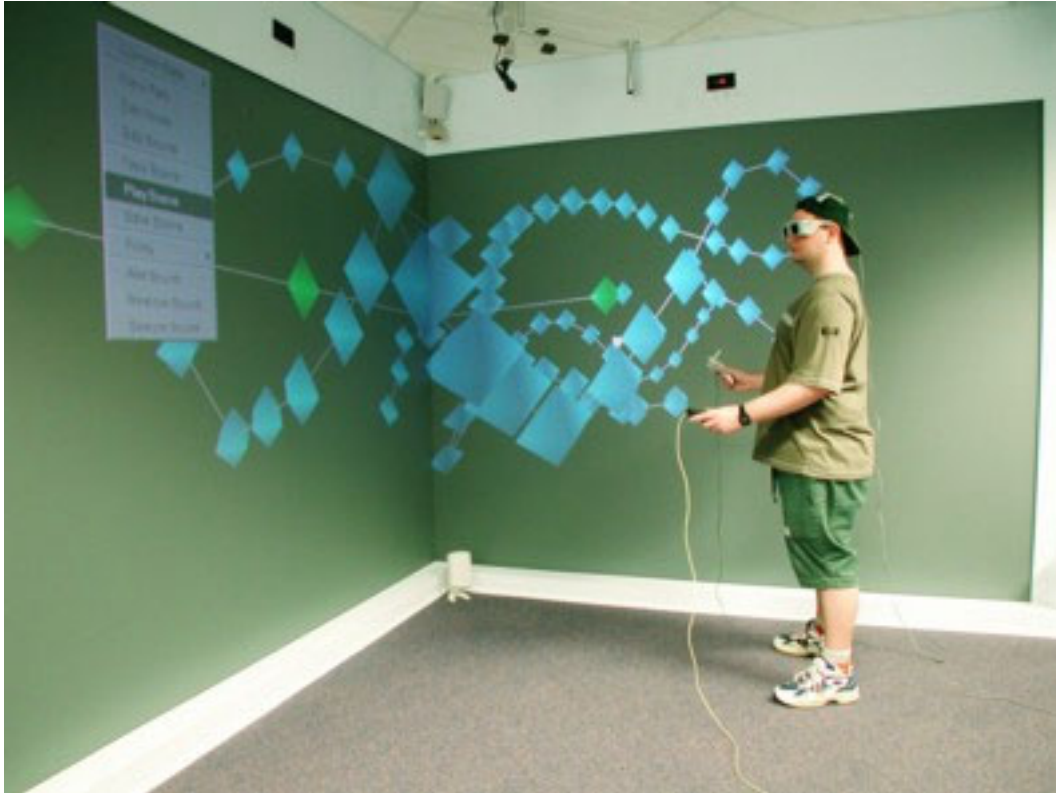


Figure 32. Le système de projection "Wedge".

Voyons maintenant le principe de la rétroprojection appliqué à plus de deux écrans orthogonaux. L'assemblage immersif d'écrans rétroprojetés le plus médiatisé est connu sous le nom de "Cave™", ou encore de "Cave-like", "Cube-like", etc.

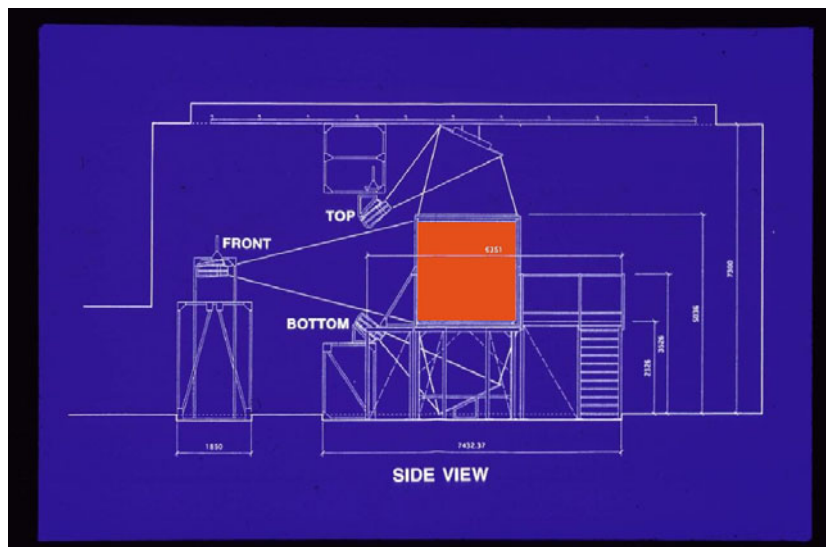


Figure 33. Le principe du "CAVE"™ vu en coupe de profil.

Ces diverses appellations commerciales désignent une catégorie d'affichage rétroprojeté complètement immersif délimitant un volume de plusieurs mètres, avec implémentation de la déformation de perspective et de la Stéréoscopie. La Figure 33 reproduit le principe du "CAVE"™ vu de profil, le volume immersif où prend place

l'observateur est figuré en rouge, chaque face du cube immersif est un écran affichant une rétroprojection directe ou par miroir, (Hirose, 2000).

La Figure 34 montre un exemple d'affichage à l'intérieur du "CAVE"TM en fonctionnement. On peut ici voir cinq faces d'affichage délimitées par les arêtes du cube.

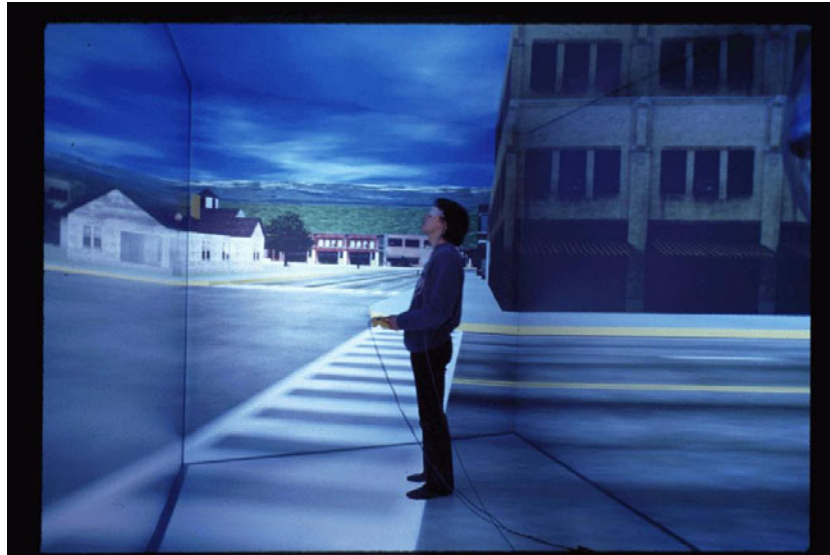


Figure 34. Exemple d'affichage à l'intérieur du "CAVE"TM en fonctionnement.

L'intérêt de ce type de dispositif d'affichage graphique réside en une surface de travail complètement immersive, dont une conséquence est l'absence de problème de "Window Violation". L'affichage reste cependant encore calculé pour un unique "viewPoint", donc, encore une fois, pour un seul observateur, éventuellement un groupe de quelques personnes rassemblées autour du "viewPoint". Le problème de l'accommodation subsiste car on doit accommoder en permanence à la distance aux parois-écrans (variable, cette fois, au gré des déplacements de l'observateur au sein du volume), avec des représentations d'Objets Graphiques 3D virtuellement affichés soit plus près, soit plus loin que les surfaces d'affichage.

Projections sur écran hémicylindrique

Il existe une solution de projection directe sur écran panoramique de forme quasiment hémicylindrique connue sous le nom de "Reality-Center", comportant plusieurs projecteurs synchronisés pour la stéréoscopie (bien que celle-ci perde de son efficacité avec la distance). Ce système d'affichage est destiné à une projection publique. La Figure 35 en illustre un exemple, l'écran se trouve sur la gauche et fait face à deux rangées de sièges (en rouge), surmontés par trois projecteurs, et derrière lesquels se trouve la plate-forme technique.

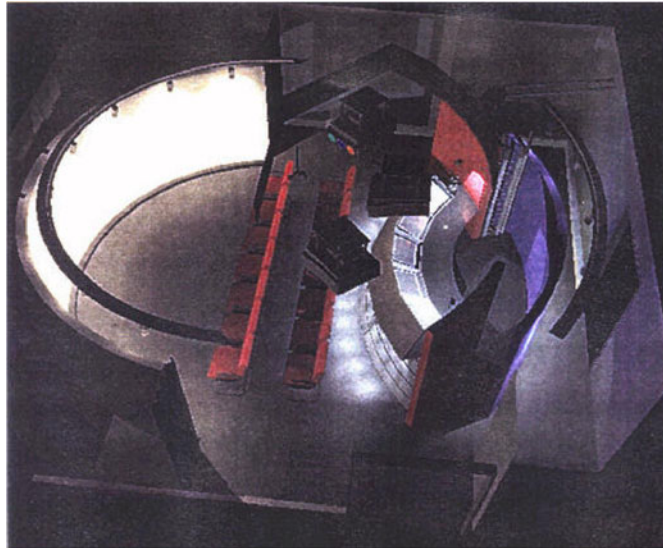


Figure 35. Exemple de "Reality-Center".

Le problème que poserait l'implémentation du "Motion Parallax" est résolu en regroupant les spectateurs autour du "viewpoint" sur des sièges.

Ce type de dispositif présente une grande Surface de travail panoramique, réduisant donc en proportion le problème du "*Cancellation Effect*". Par ailleurs, le problème de l'accommodation est notablement réduit tant que les Objets Graphiques représentés se situent à proximité ou au-delà de l'écran. Bien qu'il puisse réapparaître si l'on cherche à représenter des Objets Graphiques en deçà de l'écran. De plus, L'affichage n'est calculé que pour un unique point de vue ("*viewPoint*"), donc, pour un groupe de quelques personnes rassemblées autour du "*viewPoint*". D'où il s'ensuit l'absence de "*Motion Parallax*" déformant les perspectives de l'affichage au gré des mouvements du spectateur, fonctionnalité remplacée par une interaction avec l'affichage compliquée d'un opérateur manœuvrant le dispositif, auquel sont adressées des requêtes verbales.

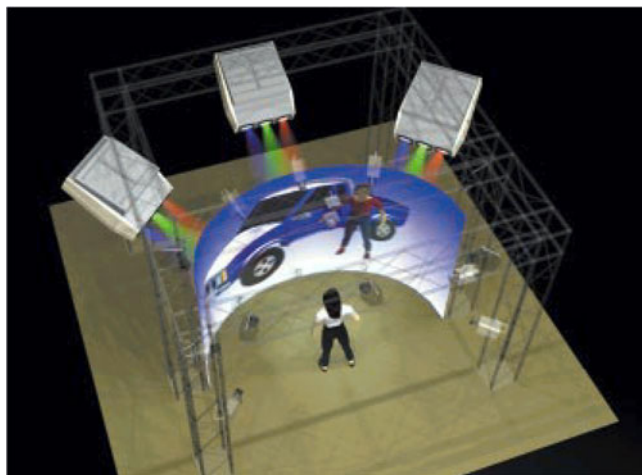


Figure 36. Exemple d'écran convexe rétro projeté.

Il n'existe pas d'exemple disponible au sujet d'un environnement hémicylindrique rétroprojeté. Un tel environnement présenterait l'avantage de la rétro projection,

(autorisant les déplacements de l'utilisateur qui n'interfererait pas avec le faisceau de projection). Le projet d'écran convexe rétroprojeté illustré Figure 36 (Gross & Stadt, 2001), implémenterait la stéréoscopie ainsi que la déformation de perspective.

Projections sur écran hémisphérique

Il existe au moins une solution de projection sphérique directe qui rappelle l'équivalent de l'effet obtenu au sein d'un planetarium.

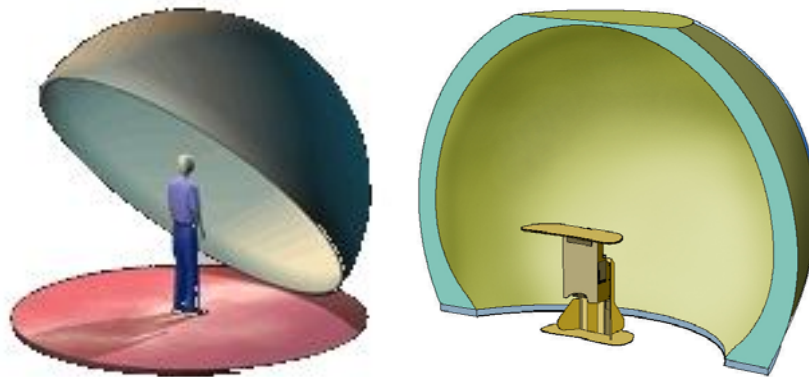


Figure 37. Exemple de coupole de projection.

Les coupoles telles qu'illustrée Figure 37 (Illustrations provenant du site de Elumens, <http://www.elumens.com/>, la figure de droite, colorée par nos soins, montre l'emplacement du projecteur) présentent une approche attractive de par l'étendue de l'affichage dont on y dispose, il s'agit là d'un hémisphère à projection directe, dont le centre est occupé par le projecteur derrière lequel l'observateur doit prendre place. La surface de travail y est relativement immersive, mais l'image étant projetée depuis la position du spectateur, il ne doit pas interférer avec le faisceau panoramique de projection par une quelconque gestuelle manuelle. D'autre part, la distance à l'écran étant réduite, le problème de l'accommodation peut réapparaître si l'on implémente la Stéréoscopie. Enfin, les prototypes existants ne disposent pas d'une éventuelle déformation de perspective au gré des mouvements de la tête du spectateur qui pourrait s'appliquer aux objets graphiques proches de l'observateur, c'est donc une navigation virtuelle au sein d'environnements plus vastes qui prend semble-t-il le relais à des fins d'exploration et de locomotion pour lesquels ce dispositif semble plus particulièrement destiné.

Il ne semble pas exister d'exemple de projection sphérique rétroprojetée, bien que certains projets artistiques envisagent à l'avenir des formes d'écrans hémisphériques voire sphériques rétroprojetés sous le nom de "Membranes". De telles projections sont encore peu ensisageables car il leur faudrait vraisemblablement surmonter des difficultés d'une part quant aux optiques propres à rendre une image sphérique sur une surface convexe, et d'autre part quant à l'ajustage des images projetées entre elles sur cette même surface.

"Head Mounted Displays"

Parfois aussi appelés "*Head-Worn Displays*" (HWD). Ces dispositifs comprennent deux écrans portés par l'utilisateur sont individualisés à chacun de ses yeux. L'implémentation du "Motion Parallax" dépend de la présence d'un système de tracking de la tête de l'utilisateur qui pourra alors se mouvoir par rapport à l'environnement présenté. La Stéréoscopie peut ne pas être appliquée. On distingue deux types de HMDs.

HMD Immersifs

L'utilisateur ne perçoit alors plus, visuellement, son environnement réel. Il est "Immergé" visuellement au sein d'un environnement virtuel distinct de la réalité.

HMD "See-Through"

L'utilisateur peut alors percevoir son environnement réel, soit par retransmission de l'environnement à l'aide d'une caméra vidéo, soit au travers de lunettes transparentes. On distingue par conséquent deux sous-catégories techniques de représentation de la réalité augmentée (AR), le "*AR-Optical-See-Through*" et le "*AR-Video-See-Through*".

Comme déjà mentionné, une autre approche présentée par la réalité augmentée dite immersive vue au travers de HMD "*seeThrough*" laisse entrevoir la possibilité non seulement de continuer à bénéficier des feedBacks visuels présentés par des dispositifs fixes d'affichage, (il doit être possible, à priori, concernant la mise en œuvre du stéréopsis, d'imaginer la possibilité de disposer aussi de la fonction "*shutters*" avec une paire de HMD "*seeThrough*"), mais également de présenter à l'utilisateur des Objets 3D situés dans différents lieux de l'espace réel, éventuellement de façon partagée, moyennant quelques détails supplémentaires à prendre en compte.

Les deux types de HMDs précédemment évoqués permettent un affichage "*see-through*" qu'on distingue soit par transparence (observation directe "*unmediated*") dite "*optical see-through*", soit par vision transmise par caméra baptisée "*video see-through*".

Le mode "*optical see-through*" (cf. Figure 38, sur laquelle la statue figure un objet du monde réel perçu au travers de surfaces semi-transparentes permettant aussi l'affichage d'autres informations venant se surimposer à la scène.). Ce mode autorise une vision directe de la réalité objective par transmission mais ne permet pas encore d'obtenir un affichage de qualité suffisante pour les objets virtuels en les laissant apparaître de façon fantomatique en environnement ambiant trop lumineux (ou en extérieur ensoleillé). Les objets virtuels sombres comme des plantes vertes (virtuelles) y apparaissent presque invisibles ou comme des fantômes sous une lumière ambiante trop importante, ce qui risque fort de se produire également avec un environnement extérieur ensoleillé, (cf. Tamura, 2000).

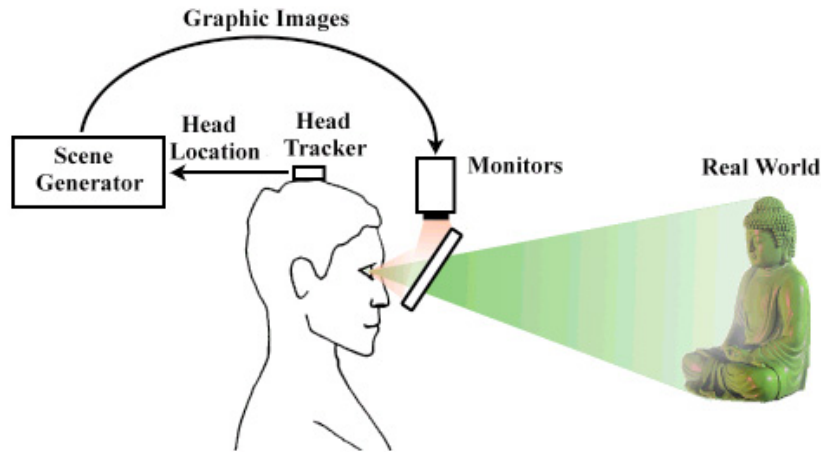


Figure 38. Illustration du "AR-optical-See-Through".

Le mode "*video see-through*" reproduit l'environnement objectif observé par une caméra et affiche une image de l'environnement (à augmenter) sur les afficheurs opaques des HMDs vidéo, la réalité objective y est donc retransmise par un médium technique.

Quelques autres systèmes expérimentaux.

Le cas particulier de la "Projection Rétinienne". (Virtual Retinal Display).

La "projection rétinienne" ("*Virtual Retinal Display*" ou VRD) reste à ce jour un cas particulier encore récent et très expérimental, ce qui est sans doute la raison pour

laquelle cette technique est rarement mentionnée dans la littérature. Son principe consiste à disposer devant l'œil de l'utilisateur, une source de lumière cohérente capable de balayer le tapis rétinien pour y projeter une image. Cette technique implique à priori des mouvements oculaires réduits de la part de l'utilisateur afin que la source reste en face de la pupille. Ce qui ne permettrait pas, par conséquent, des affichages d'Objets dans le champ de vision périphérique (Pryor *et al.*, 1998). La Figure 39 illustre la comparaison schématique entre l'illumination de la rétine produite par un écran cathodique et celle produite par projection rétinienne.

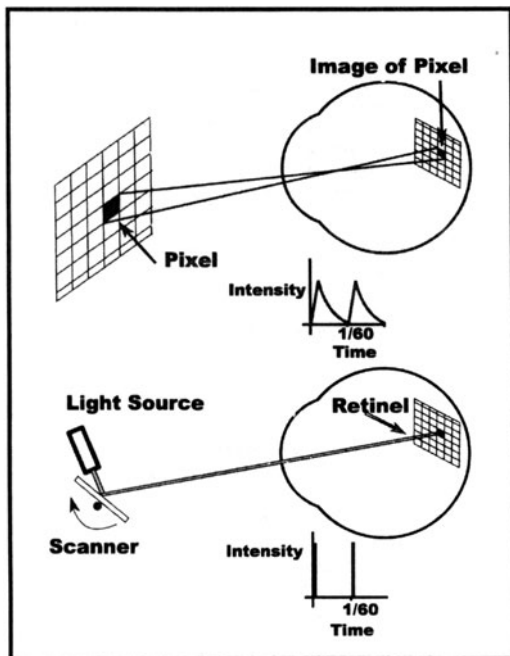


Figure 39 Comparaison schématique écran cathodique vs VRD.

Toutefois, si ces nouvelles techniques de projection sur le fond de l'œil peuvent être suffisamment miniaturisées, elles pourraient trouver un usage avec la proposition récente de Steve Mann d'intégrer la source de l'image synthétique à un verre de lunette tel

qu'illustré Figure 40 ci-dessous, (Mann, 2002). L'être humain ayant naturellement tendance à ajuster la position de sa tête pour faire face à l'objet observé, la source de l'image projetée se trouve ainsi en face de la pupille, à la discrétion du porteur de l'optique qui peut ne pas faire usage de l'augmentation produite comme le ferait un porteur de lunettes dotées de plusieurs foyers. Les avantages concernent à la fois une réalité augmentée "de tous les jours" (pour reprendre les termes de l'auteur), qui préserverait par ailleurs l'aspect social de la communication interpersonnelle avec un utilisateur dont on peut encore suivre le regard.



Figure 40 Le "Reality mediator".

Projection sur palet tournant

Autre type d'affichage, la projection volumique sur un mobile en rotation. Le mobile peut être constitué d'un hélicoïde ou, comme illustré ci-dessous (Figure 41), d'un palet tournant synchronisé à un balayage stroboscopique reproduisant une image en trois dimensions.

Cette approche est intéressante, mais ne saurait pour l'instant concerner que de petits volumes représentés à l'intérieur d'une enceinte fermée. D'autre part, elle nécessite vraisemblablement la connaissance de la position de l'utilisateur de la part du système d'affichage, et ne conviendrait par conséquent qu'à un unique spectateur.



Figure 41 Exemple de projection sur écran rotatif.

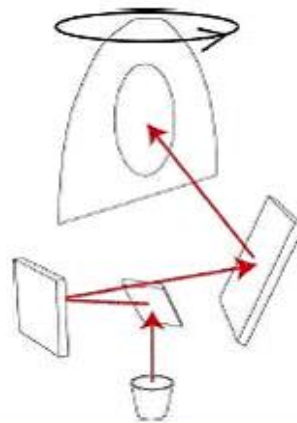


Figure 42. Principe de fonctionnement de la projection sur écran rotatif.

La Figure 41 montre un exemple de projection sur écran rotatif en fonctionnement, on peut distinguer, à l'intérieur du globe, le halo fantomatique résiduel produit par l'écran en rotation rapide. (Illustration extraite du site de Actuality-Systems,

<http://www.actuality-systems.com>). La Figure 42, (Favalora *et al.*, 2000), en illustre le principe de fonctionnement .

Les dispositifs d'affichage permettant la représentation d'objets graphiques en trois dimensions restent imparfaits, posent encore des problèmes techniques, et ne sauraient encore convenir à toutes les situations. La restitution de la profondeur est réalisable mais reste encore souvent incomplète à l'égard du système visuel de l'utilisateur. Les précédents paragraphes permettent de situer les contraintes les plus importantes imposées par l'usage d'un type de dispositif d'affichage particulier, qui influencera, voire limitera l'interaction utilisateur avec le système.

3.1.2. Autres Modalités de Sortie (non visuelles).

D'autres modalités de sortie sont envisageables pour le type d'application qui nous occupe. Les modalités relatives à l'audition seront surtout évoquées ici comme une option qui viendrait généralement compléter les modalités liées à l'affichage graphique pour mieux satisfaire une sensation de "Présence", sauf dans le cas particulier pour lequel l'objet de conception de l'environnement consisterait précisément en des "Objets Sonores 3D" n'ayant, à l'inverse, pas d'apparence graphique.

Les modalités liées au sens "tactilo-kinesthésique" ne seront citées ici que pour décrire la complexité et l'immaturation de ce domaine.

A leur suite sont évoqués quelques dispositifs de déplacement de l'utilisateur au sein de l'environnement affiché, mettant en œuvre la locomotion réelle.

Les modalités relatives à l'Audition

Un retour sonore peut compléter efficacement le feedback visuel pour augmenter la sensation de présence. On peut supposer que celui-ci n'est pas aussi utile que celui-là dans un environnement de conception de type statique (dans lequel les objets restent là où l'utilisateur les y a disposés), mais peut s'avérer plus souhaitable au sein d'un environnement dynamique dans lequel les objets peuvent par exemple subir une accélération simulée de la pesanteur donnant lieu à des collisions ou autres frottements, habituellement sonores dans notre environnement réel. On notera que la manifestation sonore des objets 3D évoquée ici ne concernerait que des sources identifiées comme ponctuelles et bien localisées dans l'espace. La question peut en effet se poser différemment si la *source est répartie* (ex. bruit des vagues sur la grève, bruit du vent, difficiles à localiser comme ponctuels, qu'il serait donc inadéquat de manifester comme tels, on aurait donc ainsi affaire à deux, voire trois, sous-ensembles de sources sonores, bruits de fond diffus, bruits de fond orientés (avec une direction générale), et bruits localisés ponctuels.

Mis à part ce type de signalisation sonore des collisions découlant de comportements de type réactif des objets disposés dans un tel environnement, d'autres usages peuvent être trouvés à un retour sonore, par exemple comme

modalité d'information vocale destinée à l'utilisateur de la part du système dans le cas où cette modalité de sortie serait utile (feedback de réception ou exécution d'une commande, ou encore sonification d'une action de l'utilisateur), communication avec un ou plusieurs autres utilisateurs, (d'ailleurs pas forcément présents en tant que participants directs à la scène).

L'approche qui semble la plus active dans ce domaine utilise la notion de "HRTF" ("*Head Related Transfer Functions*"), qui se constitue comme base théorique pour le développement de toolkits destinés à la simulation d'objets sonores dans l'espace tenant compte des contraintes imposées par les dispositifs de reproduction sonore existants.

La même dichotomie précédemment utilisée pour les affichages graphiques s'applique, ici encore, pour départager les dispositifs fixes (haut-parleurs et enceintes acoustiques disposés dans l'auditorium) des casques d'écoute.

Les dispositifs fixes permettent de constituer des rendus sonores calculés pour une position particulière au sein d'un auditorium équipé pour ce faire, position autour de laquelle doit se trouver l'auditeur, éventuellement un petit auditoire.

Les dispositifs portables tels que casques d'écoute dit "HiFi" illustrent davantage les problèmes posés par l'affichage sonore. Sans traitement particulier, un casque stéréophonique ne permet en effet qu'une localisation droite-gauche dans le plan horizontal par rapport au plan médian (*cf.* Figure 43), se compliquant par le fait que la source sonore est ressentie non pas autour de l'utilisateur mais à l'intérieur de sa tête.

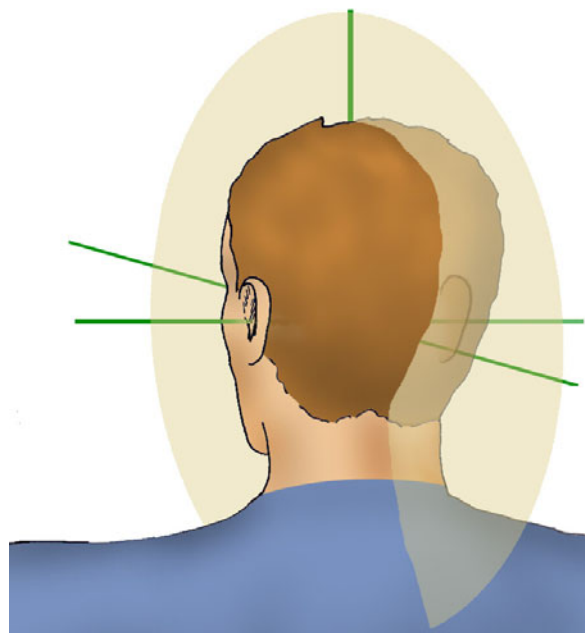


Figure 43. Le "plan médian".

Le traitement de l'azimut d'une source sonore peut reposer sur les intensité et moment d'affichage respectifs des signaux sonores de chacun des deux canaux droite-

gauche (respectivement ITD, "*interaural time difference*" et ILD, "*interaural level difference*"), auquel cas subsiste le problème du "cône de confusion" avant-arrière d'un signal sonore latéralisé. La simulation de l'élévation d'une source sonore pose un autre type de problème relatif cette fois à l'utilisateur, et en particulier à la forme du pavillon de son oreille qui lui est généralement propre. La sensation d'élévation d'une source sonore est en effet le résultat d'une modification du spectre sonore du signal initial, relativement à la forme du pavillon de l'oreille.

D'autres facteurs relatifs à la propagation que subit habituellement un signal sonore (réflexion, réfraction, réverbération, absorption éventuellement partielle, diffraction, difficulté croissante à localiser une source inversement proportionnelle à sa fréquence, etc.), interviennent également et font aussi l'objet de recherches actuelles.

Pour citer quelques exemples impliquant d'autres facteurs, Shilling & Shinn-Cunningham rapportent (2001) l'appellation de "*Cocktail Party Effect*" introduite par Yost en 1994. Cette appellation recouvre la notion selon laquelle la discrimination sonore par l'auditeur plongé au sein d'un environnement sonore multi-source est améliorée si les sources sont effectivement localisées dans l'espace.

Ce sujet, appliqué aux environnements virtuels apporte par ailleurs une conséquence intéressante directement l'éventuelle définition d'un objet 3D, qui se conçoit souvent (à première vue), comme un ensemble d'informations permettant seulement l'affichage graphique d'une forme éventuellement texturée. La notion de "*Cocktail Party Effect*" introduit la notion de manifestation audible d'un objet de la scène qui devient alors une information qui peut sinon entièrement, du moins en partie représenter l'objet. Il s'agira là en particulier du cas de l'objet émettant continuellement une manifestation sonore, -ex. le tic-tac d'un réveil, le bruit constant d'un mobile en déplacement, qui devrait donc subir entre autre l'effet Doppler-, dont la source sonore sera supposée accompagner les mouvements imposés à sa forme affichée.

Pour conclure quant au rendu sonore d'une scène 3D, on retiendra que si ce domaine peut s'avérer très utile à la simulation d'une scène 3D, il se révèle rapidement aussi complexe que la conception des affichages graphiques qui ont bénéficié de bien davantage de développements au cours des dernières années.

Les modalités relatives au sens tactilo-kinesthésique

Le retour haptique tactilo-kinesthésique reste encore un défi intéressant quant à l'enrichissement du spectre perceptuel de l'utilisateur plongé dans une simulation d'environnement 3D dont la vocation sera de lui autoriser une manipulation directe d'objets graphiques 3D. Les recherches sur le sujet semblent se concentrer néanmoins pour l'instant sur un retour concernant essentiellement la main, éventuellement le bras de l'utilisateur, et ne semblent guère envisageables de façon efficace à court terme, mis à part la simulation d'un contact avec l'objet, en terme de

rendu dynamique de texture sur un objet déformable (ex. l'évaluation tactile d'un tissu utilisé pour un vêtement).

Ce type de dispositif n'intéressant que la sensation de toucher ou de retour de force au niveau de la main ne saurait d'autre part concerner que le toucher d'objets de l'environnement disposés "à portée de main" de l'utilisateur. Ce serait par exemple le cas pour une application de peinture artistique au pinceau (Baxter *et al.*, 2001), telle qu'illustrée ci-dessous (Figure 44).



Figure 44. Exemple d'application d'un dispositif à retour d'effort pour la simulation de peinture artistique.

Notre propos se situant dans un rapport de communication entre l'utilisateur et le système lui présentant un environnement de conception, le domaine d'une haptique qui ne traiterait que la main et ses contacts avec des objets graphiques 3D resterait d'un intérêt assez périphérique, raison pour laquelle le sujet ne sera pas traité ici.

Nous allons toutefois, afin de laisser entrevoir la complexité ainsi que les difficultés rencontrées dans ce domaine naissant, fournir quelques exemples représentatifs du domaine.

Dans un exemple datant de 1990, (Iwata, 1990), le principal objectif de recherche visé concerne l'implémentation du retour d'effort au cours d'une manipulation d'un objet virtuel. Un manipulateur fut développé en tant qu'entrée tactile sur un ordinateur de bureau (*cf.* Figure 45), lequel pouvait générer un retour d'effort sur la paume et les doigts de l'utilisateur.



Figure 45. Le manipulateur à 9DOF.

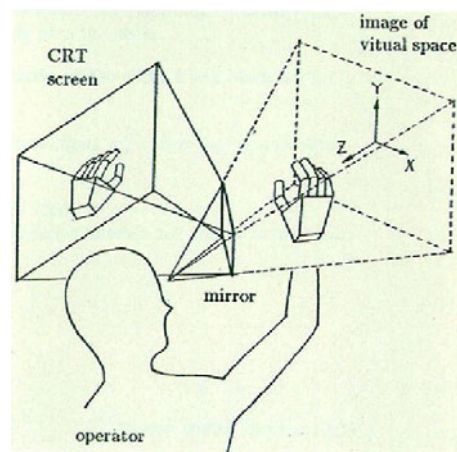


Figure 46. Combinaison du retour d'effort sur la main avec l'affichage d'une main virtuelle.

Afin d'obtenir un rendu plus réaliste de la scène, l'affichage permettait de représenter un modèle de la main de l'utilisateur en même temps que l'objet de la

manipulation (cf. Figure 46 et Figure 47). Les applications visées concernaient les CAD et logiciels d'animation 3D.

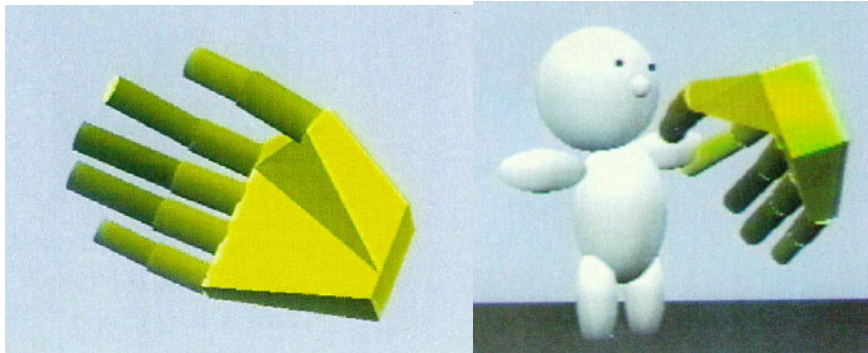


Figure 47. Détail de la main virtuelle et de son utilisation.

Outre le fait que la main y est considérée comme unique modalité d'entrée dans le cadre d'une interaction directe tactile avec les objets 3D virtuels, on notera, malgré des progrès réalisés dans ce domaine, la complexité et l'encombrement du système mécanique ainsi que le manque de liberté permise à l'utilisateur.

Autre difficulté inhérente à la complexité du domaine, le rendu haptique de formes de surfaces, dont un exemple exposé par Iwata *et al.* en 2001 est illustré ci-après (Figure 48). On y voit là encore un encombrement qui en limite le spectre d'utilisation.



Figure 48. Exemple de dispositif fournissant un rendu haptique de formes de surfaces.

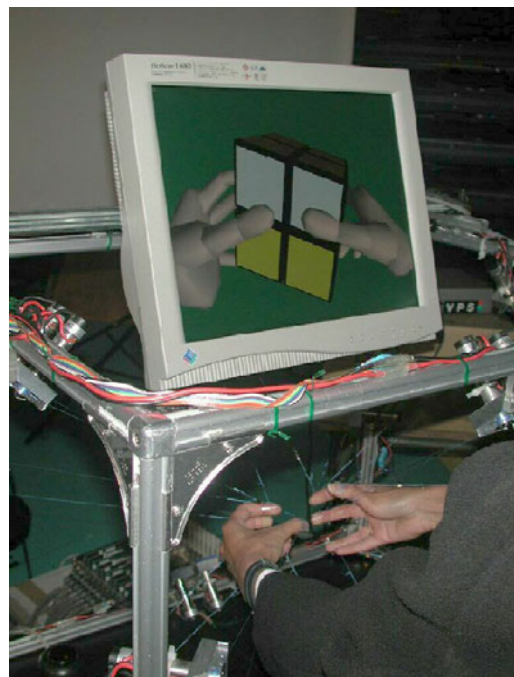


Figure 49. Manipulation avec le "Spider".

Un autre exemple représentant une interaction bimanuelle déportée est le "Spider" de M. Sato (Sato, 2001, ci-dessous,

Figure 49). La manipulation pour l'application représentée y semble praticable en très peu de temps, le champ d'application d'un tel dispositif semble cependant devoir se restreindre à des exercices du même type et reste déroutant en n'appliquant des retours d'efforts que sur les extrémités des doigts.

Un dernier exemple récent présenté sur le salon Imagina 2004 est le « DigiHaptic » de G. Casiez (Figure 50).



Figure 50. Dispositif « DigiHaptic » de contrôle de locomotion ou de manipulation.

Ce dispositif permet en effet de percevoir de manière kinesthésique les collisions avec les obstacles rencontrés au cours d'une locomotion au sein d'un environnement 3D. Il permet également, à l'aide d'un mode alternatif, la manipulation d'un objet sélectionné au sein de son environnement.

Ces derniers exemples illustrant quelques-uns des efforts actuellement réalisés dans le vaste domaine des recherches visant à obtenir un rendu tactilo-kinesthésique indique la diversité des solutions mises en oeuvre. Ils permettent dans certains cas à l'utilisateur d'effectuer une autre tâche de manière alternative, voire d'accéder facilement à une tâche non informatisée (rédiger une note manuscrite, répondre au téléphone, etc.). D'une manière générale, ces techniques restent cependant dédiées aux catégories de tâches pour lesquelles elles sont conçues, par exemple la locomotion au sein de l'environnement, ou la manipulation, modification ou création d'un objet 3D. Les dispositifs résultants ne semblent pas, en entrée du système, permettre un accès aisé à l'ensemble des catégories de tâches qu'impliquerait un environnement 3D. Ceci suppose, en entrée du système et pour accéder à un ensemble de catégories de tâches utilisateur impliquées par une activité, la mise en place de plusieurs dispositifs correspondants auxquels l'utilisateur devra alternativement faire appel au cours de son activité.

3.2 Modalités d'entrée

3.2.1. les "gestuelles"

Définitions générales

Depuis le "Put-That-There" de R. Bolt (1980), de nombreuses recherches ont mis en oeuvre diverses formes de gestuelles généralement restreintes à certaines parties du corps les plus expressives de l'utilisateur. Les plus connues à titre d'exemple concernent habituellement le gant de données, qui ne renseigne pourtant pratiquement que sur les mouvements des doigts de la main, focalisant les chercheurs qui en disposent sur les seuls mouvements de cet appendice pour en extraire une information exploitable dans le cadre d'une IHM. D'un point de vue théorique, "La gestuelle" est souvent elle aussi considérée isolément des autres modalités de communication humaine, ainsi que du contexte dans lequel a lieu l'expression de l'utilisateur, en se concentrant essentiellement sur les indices fournis par ses mains.

Selon F. K.H. Quek, (1994), la définition de la "gestuelle" n'est pas très claire dans la littérature. Il rapporte notamment que la communauté concernée par la "réalité virtuelle" a désigné la gestuelle comme une modalité d'entrée de choix en estimant cependant que la définition d'une "*gesture-based computer input*" n'a été que partiellement effectuée. Une table ronde en 87 qu'il évoque au sujet de la gestuelle se serait ainsi essentiellement attachée à interpréter les positions et mouvements de stylos sur les tablettes graphiques.

L'auteur brosse également un historique des expérimentations du domaine fin des années-80 - début des années 90 avec des systèmes de « *Hand-Tracking*⁴ » prenant généralement la forme de gants, en faisant remarquer assez justement que la plupart de ces systèmes dédient davantage la main de l'utilisateur à l'interface que l'interface à l'utilisateur, notamment à cause de leur encombrement qui gêne d'autres types d'interaction manuelle, que l'utilisateur doit pouvoir effectuer sans gêne et sans avoir à se débarrasser du système de "*Hand-Tracking*", par exemple, afin de rédiger une note à l'aide d'un stylo, voire, simplement prendre une communication téléphonique.

La "*Taxonomie des gestes*" proposée par Nespoulous & Lecours en 1986 (cf. Figure 51), départage "*Symbol Gestures*" et "*Act Gestures*". Les gestes "*Symbol*" sont des mouvements de la main prenant un rôle linguistique (ex. langages de signes). Dans les "*Act Gestures*", les mouvements réalisés sont directement relatifs à l'interprétation visée.

⁴ Hand-Tracking : dispositif de suivi des mouvements de la main dans l'espace, fournissant trois coordonnées de position par rapport à un point de référence, ainsi que trois angles signalant une orientation du dispositif relativement à une orientation de référence.

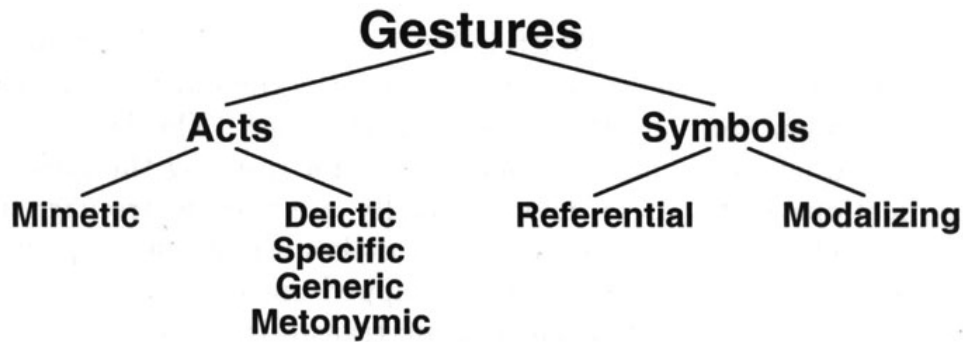


Figure 51 "Taxonomy of gestures" de Nespoulous & Lecours (1986).

Les gestes symboliques y sont classés en *référentiel* et *modalisant* ("*referential and modalizing*"). Les premiers opèrent indépendamment pour désigner des objets ou des concepts. Frottant le pouce sur l'index de façon circulaire, par exemple, fait référence à la notion d'argent (le référent). Les gestes modalisants s'utilisent en conjugaison avec quelqu'autre moyen de communication (ex. le langage) pour indiquer l'opinion du communicant. Un autre exemple est le "*common continuation symbol*" (mouvement circulaire répétitif de la main avec index pointant vers le haut) qui peut signifier de continuer à effectuer une opération en cours relative au contexte et accompagnant la communication effectuée via d'autres modes.

Dans cette même "*Taxonomy of gestures*", les "*Act Gestures*" distinguent deux classes: *mimétiques* et *déictiques*.

Les "*Mimetic act gestures*" se caractérisent par leur iconicité. Ils sont généralement réalisés comme pantomimes du référent. Par exemple, un fumeur mimant l'action d'allumer une cigarette qu'il porte aux lèvres indiquera ce faisant qu'il a besoin de feu. De tels gestes sont habituellement générés "*on-the-fly*" sans convention préalable.

Les "*Deictic act gestures*" (ou gestes de désignation) peuvent se subdiviser en "*specific*", "*generic*" et "*metonymic deictic act gestures*".

Les "*specific deictic act gestures*" apparaissent quand quelqu'un désigne un objet ou endroit particulier.

Les "*generic deictic act gestures*" désignent une classe d'objets par désignation d'un unique objet de cette classe.

Les "*metonymic deictic act gestures*" se manifestent lorsqu'on pointe un objet pour désigner quelqu'autre entité qui lui est relative (ex. pointer du doigt une représentation de la tour Eiffel pour désigner Paris).

Francis K.H. Quek, (1994), dans le cadre de la reconnaissance de la gestuelle (réduite) des mains de l'utilisateur en utilisant des caméras pour la vision par ordinateur, définit quatre classes d'actes gestuels à reconnaître pour une machine :

"*Locative*", pointage d'un endroit particulier, d'une direction, ou d'un objet.

"*Orientational*", positionnement d'objets ou de viewpoints par des rotations de la main.

"*Spatial pantomime*", relevé global des mouvements de la main pour déterminer une forme.

"*Relative spatial*", relation spatiale telle que plus près, plus loin, plus à droite, plus à gauche, etc.

Dans son article intitulé "*Le geste, canal de communication homme/machine; la communication « instrumentale »*", Cadoz (1994), définit des fonctions du geste parmi lesquelles on distingue des gestuelles *ergotique*, *épistémique* et *sémiotique*.

" *Associées à la main, on peut considérer trois fonctions différentes mais complémentaires et imbriquées : une fonction d'action matérielle, modification, transformation de l'environnement, que nous appellerons ici ergotique, une fonction épistémique, c'est-à-dire de connaissance de l'environnement et enfin une fonction sémiotique, c'est-à-dire d'émission d'information à destination de l'environnement. Chacune de ces trois fonctions prise séparément fait intervenir à des degrés variables chacune des deux autres.*"

Outre sa relation au sens exploratoire du "*toucher*" détectant le contact et l'état de surface d'un objet, la fonction épistémique tient compte de considérations *Tactilo-kinesthésique* (on dit aussi haptique) et *Proprioceptive*.

La fonction *ergotique* du geste illustre l'action mécanique par laquelle un utilisateur peut modifier l'environnement ou des objets présents.

La fonction *sémiotique* enfin, est davantage relative à la communication, Cadoz y distinguent les catégories suivantes.

- les *bâtons* : mouvements qui accentuent ou donnent de l'importance à un mot particulier ou une phrase.
- les *idéographes* : mouvements qui esquissent le chemin ou la direction de la pensée.
- les mouvements *déictiques*: qui indiquent un objet, un lieu ou un événement.
- les mouvements *spatiaux* : qui décrivent une relation spatiale.
- les mouvements *rythmiques* : qui décrivent le rythme ou le repos d'un événement.
- les *kinétographes* : mouvements qui décrivent une action du corps ou quelque action non humaine.
- les *pictographes* : qui dessinent une figure ou la forme du référent.
- et les mouvements *emblématiques* qui sont des ensembles utilisés pour illustrer un propos verbal, répétant ou se substituant à un mot ou une phrase.

Axel Mulder, (1996), divise les mouvements de la main en deux principales catégories, l'une impliquant la *communication* (tels que les mouvements avec les mains vides), l'autre impliquant *manipulation* et *préhension*. Avec quelque part entre les deux, des mouvements de la main identifiés comme des actions d'*exploration haptique*. Des considérations similaires devraient selon lui avoir conduit C. Cadoz en 1994, à classer les mouvements de la main en terme de fonction *sémiotique*, (pour communiquer des informations à l'environnement), *ergotique*, (associé à la notion de travail et à la capacité de l'être humain à manipuler des objets du monde réel, de créer des outils), et *épistémique* (permettant aux êtres humains de pratiquer une exploration haptique et tactile de l'environnement).

La majorité des taxonomies relatives à "la gestuelle" nous semble rester partielle en se restreignant à certaines catégories de mouvements observables chez l'utilisateur. Le point commun des diverses définitions concernant toute gestuelle prenant la forme d'un mouvement observable chez l'homme, notre point de départ dans un cadre plus général, sera la prise en compte de toute sorte de mouvements observables pouvant fournir une information ayant un rapport avec le sujet de l'étude, mouvements qui resteront à organiser en diverses catégories.

Rappel sur les modalités liées aux mouvements

Des "modalités gestuelles" propres à transmettre des informations qui seraient exploitables dans un environnement graphique 3D nous semblent impliquer des catégories complémentaires de médias transmettant aussi des signaux redondants.

Le regard apparaissant par exemple comme un préalable à la déictique manuelle pour situer ce que l'on compte indiquer du doigt, (du menton, du coude, etc.). Ces deux types de mouvements paraissent redondants, avec toutefois des différences telles que l'avantage de meilleure stabilité et précision du regard, comparés à la visée approximative du doigt, menton, coude, etc. qui apparaît alors davantage comme une confirmation de désignation si la direction du regard est perceptible par le destinataire du message.

Deux catégories principales de mouvements porteurs d'information utile apparaissent donc, les *directions oculaires*, et les *mouvements des mains*. Chacune de ces catégories implique, selon le contexte d'exploitation, d'autres catégories de mouvements dont elles dépendent sans que celles-ci ne constituent forcément des canaux de modalités informant une gestuelle.

Les directions oculaires, manifestant un indice préalable à une gestuelle déictique d'une autre partie du corps ou à l'énonciation d'un acte verbal, dépendent ainsi des mouvements de la tête qui elle-même est située par rapport au reste du corps. Les informations fournies par les mouvements de la tête peuvent éventuellement constituer par exemple une gestuelle déictique (le mouvement du menton), et il en sera parfois de même pour les déplacements de l'ensemble du corps au sein de l'environnement, mais ce genre d'éventualité n'en fait pas des catégories principales.

Si toutefois ces médias se trouvent mis en oeuvre, il pourrait être concevable d'étendre l'extraction de gestes déictiques (ou autres) à d'autres médias renseignant sur les mouvements d'autres parties du corps.

De même que pour les directions oculaires, il pourra être utile de situer les mouvements des mains par rapport au reste du corps via les mouvements des épaules et des bras.

Diverses modalités d'expression utilisent les mains au nombre desquelles on retiendra des mouvements *déictiques* (qui peuvent se manifester par diverses 'poses'), des paramètres dimensionnels via généralement des 'poses' statiques, ainsi que des descriptions de formes (les "*pantomime spatiales*" de Francis Quek) ou encore des invocations de déplacements divers (translations ou rotations, les "*Relative spatial gestures*" de F. Quek que l'on retrouve parmi les "*modalizing symbol gestures*" de Nespoulous & Lecours sous l'appellation de "*common continuation symbol gesture*" cités plus haut), que l'on rangera dans la catégorie des gestes mimétiques.

Une autre façon d'aborder le sujet consiste, et c'est peut-être là le plus important, à partir d'une connaissance des tâches que l'utilisateur va effectuer, à déterminer quelles seraient les modalités à prendre en compte pour ensuite seulement déterminer les médias les plus importants à renseigner concernant les mouvements observables de l'utilisateur.

A titre d'exemple, nous avons évoqué ci-dessus une modalité déictique servant entre autres à la désignation d'un objet au sein de l'environnement. En adoptant à ce sujet la complémentarité de la direction du regard et d'un mouvement d'une partie du corps confirmant une forme de sélection, nous aurions, outre la mise en oeuvre du Gaze-tracking, à choisir une ou plusieurs des catégories de mouvements corporels renseignant sur une telle déictique.

Distinction des dispositifs "embarqués" ou portables

Cet aspect a priori technique en concerne d'autres que nous allons évoquer ici et qui nous ont amenés à les distinguer.

La mobilité de l'utilisateur au sein de différents éventuels environnements augmentés pouvant concerner l'application envisagée peut s'avérer importante. Il existe d'ailleurs déjà, dans le même ordre d'idée, des applications "extérieures" ("*outdoor*") qui pourraient bénéficier d'une telle mobilité, (ex. le résultat du projet européen "ArcheoGuide", dédié au tourisme, mais dont les techniques seraient sans doute exploitables en conception architecturale sur le terrain).

La notion de "station de travail" fixée à un environnement physique réduit se conçoit de facto lorsqu'il s'agit de la mise en oeuvre de dispositifs encombrants tels que "mur de projection", "workbench" ou autre "reality center", ou encore pour des raisons de coût lorsqu'on aura affaire à un environnement réduit (ex. Mulder & van

Liere, 2002). Les premiers cherchent à adresser plusieurs utilisateurs en leur présentant un environnement strictement virtuel sans les "instrumenter", on notera cependant que si l'on cherche à permettre à l'un de ces utilisateurs d'user de modalités d'entrée telles que définies plus haut, il faudra sans doute l'instrumenter tout de même (Gaze-tracking, microphone, tracking de mouvement sur la tête, les mains...).

Une autre notion désigne les dispositifs "portables" sous le terme de "*wearable computers*" qui encombrant l'utilisateur pour le rendre davantage mobile et indépendant d'un lieu éventuellement instrumenté. Ces techniques apportent de nouvelles possibilités tant pour ce qui est des médias de sortie que des médias d'entrée du système en personnalisant notamment l'interaction de l'utilisateur avec le système.

Nous apporterons par la suite la distinction entre matériels fixes et portables concernant les dispositifs de capture de mouvements.

Distinctions de catégories de mouvements corporels

Nous assimilerons donc, pour notre part, les modalités gestuelles en entrée comme provenant de la prise en compte de mouvements observables de l'utilisateur par le système, renseignant sur ses intentions en situation de conception.

Dans une décomposition de mouvements du point de vue le plus global au plus particulier, l'ensemble des mouvements que l'on peut observer de la part d'un acteur va concerner ses déplacements dans l'espace ainsi qu'éventuellement des variations de position de l'ensemble du corps. On distinguera ensuite les mouvements des membres, ceux des mains et des doigts, les mouvements de la tête, puis ceux des yeux.

Les déplacements de l'utilisateur

Ce type de mouvement peut trouver une signification dans certains cas particuliers sous une forme par exemple chorégraphique. Les cas dans lesquels les positions ou déplacements du corps portent une signification sont cependant rares. Toutefois, la position d'un utilisateur reste une information dont on devra tenir compte selon l'environnement d'affichage mis en oeuvre, (graphique, sonore...), ce dernier cas ne constituant pas forcément, exprimé ainsi, une modalité d'entrée mais plutôt une donnée nécessaire à l'affichage graphique ou sonore par le système.

Cette question, qui semble de prime abord intéresser essentiellement la réalité augmentée représentée sur HMD, intéresse en fait tous les dispositifs d'affichage à des degrés divers de précision ou d'extension au sein de l'environnement réel de l'utilisateur, dès lors que la position du "viewpoint" utilisateur, (et "*earpoint*", le cas échéant), peut être modifiée par les propres déplacements de celui-ci (Reality Center mis à part pour ce qui concerne le seul aspect visualisation). Le tracking des position et orientation de l'utilisateur au sein de l'environnement physique réel intéresse en

effet les dispositifs fixes d'affichage sur lesquels on applique la déformation de perspective, laquelle nécessite la connaissance du "*viewpoint*" de l'utilisateur (que ce soit avec ou sans stéréoscopie). Ce tracking utilisateur est aussi nécessaire dans le cas d'une Réalité Augmentée sur HMDs, non seulement pour des affichages locaux dans un réduit, mais aussi en extérieur, (comme c'est par exemple le cas pour le projet Européen "ArcheoGuide"). Les techniques disponibles sont peu nombreuses, apportent nombre de contraintes supplémentaires et ne conviennent pas, considérées isolément, pour différents cas de figures de déambulation de l'utilisateur au sein de l'environnement réel.

Autre sujet de cette section, l'usage d'une locomotion "sur-place" de la part de l'utilisateur pour naviguer au sein de l'environnement de synthèse qui sera évoquée plus loin.

Techniques de capture des position, orientation et déplacements de l'utilisateur

Comme le signale R. Azuma en 1999 au sujet du challenge que pose une AR nomade en milieu extérieur, "... the biggest challenge lies in accurate tracking outdoors: determining the user's position and orientation with sufficient accuracy to avoid significant registration errors."

Dans cet article, il commente quelques techniques telles que

- GPS
- Inertial and dead reckoning
- Active sources
- Passive optical
- Electronic compass and tilt sensors

Cette liste n'est pas exhaustive et concerne des applications nomades, mais constitue un bon aperçu que nous allons maintenant détailler.

Le *GPS* consiste en un équipement qui s'est suffisamment miniaturisé pour être maintenant porté par l'utilisateur (il est aussi devenu d'un coût très abordable). Ce type d'équipement ne fonctionne pas en intérieur, et il ne fournit qu'une position approximative de l'émetteur avec une imprécision de l'ordre de la dizaine de mètres, il ne fournit pas son orientation, même si celle-ci est parfois affichée par certains systèmes GPS, qui est en réalité calculée comme un cap à partir d'un relevé de position précédent (ce qui se conçoit pour un véhicule). Il a récemment été utilisé avec succès par le projet européen "ArcheoGuide" pour déterminer la position approximative de l'utilisateur sur le site.

Les *centrales inertielles*, qui peuvent fournir une orientation sur trois degrés de liberté en rotation, ont été elles aussi miniaturisées au point d'être portables, elles ont aussi gagné en précision, mais conservent cependant leur principal défaut d'accumulation d'erreur qui a toutefois été considérablement réduit. Elles sont par

exemple utilisées dans les exosquelettes de post-production multimédia servant à animer quelque personnage de synthèse.

Les *accéléromètres* peuvent, sur trois degrés en translation, fournir une indication de déplacement du dispositif de façon approximative en raison de l'erreur qui peut s'accumuler. On peut les utiliser plus efficacement pour fournir une mesure d'inclinaison par rapport à la verticale (champ de pesanteur, inclinomètre) en l'absence de mouvement (ou plutôt, d'accélération notable). Les dimensions modestes de certains modèles les autorisent à être transportés. Ils sont plus économiques que les centrales inertielles, mais aussi moins précis en proportion pour leur être associés.

Les *boussoles* ou autres *compas électroniques* fonctionnant par mesure de champ magnétique deviennent de plus en plus précis (sans toutefois encore atteindre la résolution visuelle de l'ordre de la minute d'angle que perçoit l'œil humain), mais restent sensibles aux variations de champs électro-magnétiques que génèrent de nombreux appareils munis de transformateurs ou encore aux masses métalliques trop importantes et contenant des matériaux ferreux. Leurs dimensions leur permettent d'être transportés. Un tel équipement est utilisé dans le projet "ArcheoGuide" pour préciser la direction approximative de la tête de l'utilisateur sur le site d'Olympie.

L'analyse de *modification de champ électrique*. La position de la main de l'utilisateur au sein d'un champ électrique peut être mesurée, (Zimmerman *et al.*, 1995 ; Smith, 1996 ; Paradiso & Gershenfeld, 1996), ce type de technique est sans doute perfectible, mais ne saurait fonctionner que dans un environnement réduit excluant la présence de masses métalliques.

L'*active badge*" consiste en un émetteur radio ou infraRouge de séquence identifiée, portable, qui est détecté par un réseau de capteurs Radio (ex. Starner *et al.*, 2000) ou IR (Mynatt *et al.*, 1998), disposés en différents lieux d'un intérieur aménagé à cet effet. Ce type d'équipement portable permet une localisation approximative de la position de l'utilisateur. Une version miniaturisée utilise les transpondeurs détectables au passage de portes aménagées en portiques.

Le *tracking magnétique* (MT) constitue une technique très précise de positionnement absolu tant en translation qu'en rotation, elle est cependant sensible à toute modification de l'environnement par apport de masses métallique, elle reste à réserver à des volumes de tracking de l'ordre du mètre, et nécessite de fréquents calibrages encore longs et fastidieux.

Le *tracking Ultra-sonore*, légèrement moins précis que le tracking magnétique, peut être utilisé sur un volume aménagé plus important (plusieurs mètres) et ne présente pas de sensibilité aux modifications de l'environnement par des objets métalliques. Il est cependant sujet aux problèmes d'occlusion. (ex. Zimmerman *et al.* 1987 ; Foxlin *et al.*, 1998).

Les *cameras d'observation* de l'utilisateur existent sous diverses formes et donnent lieu à diverses applications de "*body-tracking*", (ex. Ghidary *et al.*, 2000), elles sont généralement fixées dans l'environnement et calibrées. Certaines techniques de post-production habillent l'utilisateur de façon à ne pas absorber la lumière ou lui accolent des "marqueurs" détectables par analyse d'image. D'autres vont faire appel à des marqueurs IR (LEDs, ex. Foxlin *et al.*, 1998) pour relever les position et orientation de la tête de l'utilisateur coiffée d'un casque ainsi marqué. Bien que ces techniques progressent, elles restent cantonnées à un périmètre généralement bien défini et gèrent difficilement plus qu'un unique utilisateur présent sur la scène.

Le même principe que ci-dessus s'illustre aussi non plus avec des caméras d'observation mais des *microphones* en stéréophonie pour déterminer la direction d'un orateur. (Pingali *et al.*, 1999]).

La *reconnaissance active (outward looking)*, de l'environnement à l'aide d'une camera portée par l'utilisateur réalise le procédé inverse. Malgré les progrès réalisés dans le cadre du "*machine vision*", un marquage de l'environnement est encore nécessaire pour obtenir une reconnaissance exploitable du *view-point* ("*registration*", ex. Rekimoto, 1998; López de Ipiña & Lo, 2001). Toutefois, on trouve quelques méthodes de reconnaissance de l'environnement sans marquage de celui-ci, mais par relevé d'une forme de signature (analyse de fréquences) comparée à des signatures répertoriées pour un site défini et enregistrées sur base de données accessibles par radio, (cas du projet "ArcheoGuide").

On peut citer aussi d'éventuelles applications utilisant la détection d'*infraRouge passif* par des capteurs auto-orientés (capteurs de poursuite) en direction de la source que constitue l'utilisateur, (l'application connue consiste en un système d'alarme s'orientant en direction de l'intrus pour l'asperger de gaz lacrymogène à intervalles de quelques secondes). Ce type de technique ne serait encore cependant à utiliser qu'avec un unique utilisateur.

Plus exotiques, les *tapis sensibles* aux masses pesantes de type "*Smart Floor*", (Orr & Abowd, 2000), permettent, outre le tracking de plusieurs utilisateurs présents sur le même sol aménagé à cet effet, de les identifier en fonction de leur poids, voire, de leur démarche.

Application à la navigation utilisant la locomotion de l'utilisateur

Ce type de simulation mécanique d'un "déroulement" du terrain au gré du déplacement "sur place" de l'utilisateur est, au même titre que ce qui concerne le retour d'effort sur, par exemple, la main de l'utilisateur, une technique souffrant d'immaturation et destinée sans doute encore pour quelques temps davantage à des environnements de transfert de savoir à mettre en œuvre (ou encore à des installations d'entraînement sportif) qu'à des environnements de conception d'objets 3D simulés.

"Virtual Treadmill" et Tapis roulants

Tenter d'exploiter la démarche que nous utilisons quotidiennement pour nous déplacer afin d'en tirer une modalité de navigation au sein d'un environnement virtuel immersif a donné lieu à quelques réalisations. La grande difficulté semble résider dans le tracking exact du centre de gravité de l'utilisateur afin d'asservir la réponse du dispositif aux mouvements des pieds, ce qui semble difficile à obtenir avec un système mécanique qui présenterait un "déroulement du terrain" simultanément aux deux pieds de l'utilisateur de façon homogène. Il ne semble pas exister de réalisation qui traiterait chacun des pieds de l'utilisateur de façon indépendante.



Figure 52. Virtual motion controller (VMC).

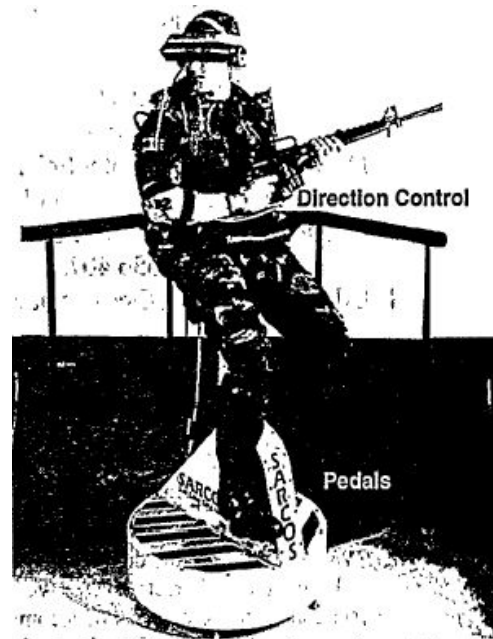


Figure 1: Soldier on UniPort
Figure 53. "Uniport", un dispositif de locomotion à pédalier.

Il existe des approches ne mettant en œuvre que certains indices de la démarche, comme le fait de marcher sur place face à la direction de déplacement voulue ou encore le déplacement minimal sur une surface réduite indiquant une direction de déplacement virtuel ainsi qu'une vitesse ou une accélération proportionnelle au déplacement local effectivement réalisé (cf. le "VMC", Figure 52, ci-dessous, de Peterson *et al.*, 1998).

D'autres réalisations ont cherché à utiliser une démarche plus naturelle sur quelque "tapis-2D" (ex. "L'OmniDirectional Treadmill" de Darken *et al.*, -1997-), sous la forme de deux couches de rouleaux rotatifs superposées de façon perpendiculaire, mais posent des problèmes encore mal résolus d'asservissement à la démarche de l'utilisateur.

On trouve aussi des compromis exploitant d'autres moyens de locomotion au travers de la mise en œuvre d'un pédalier à l'instar de ce que l'on fait avec une bicyclette, tel le "Uniport" de Sarcos, (cf. Figure 53).

Enfin, une autre possibilité (disponible à plus large échelle) consiste en l'usage d'un tapis roulant d'entraînement sportif, tel que par exemple mis en oeuvre pour le projet SNCF "SOFI" (cf. Figure 54, Illustration extraite de la plaquette de présentation, VRIC 2002).

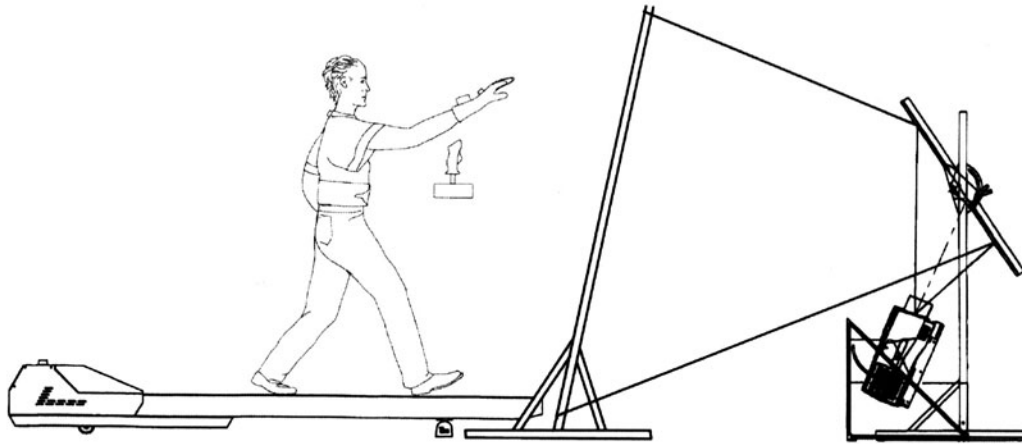


Figure 54. Exemple de mise en oeuvre d'un tapis roulant.

Il pourra sembler curieux d'évoquer ce cas particulier au chapitre des "*modalités en entrée du système*". Dans le principe, il s'agit bien d'une modalité d'entrée qui serait supposée impliquer une "locomotion" au sein de l'environnement de synthèse asservie aux déplacements effectifs réalisés par l'utilisateur sur le tapis pour obtenir ce type de déplacement.

Dans le cas particulier du tapis roulant, le dispositif mis en oeuvre présente des difficultés d'asservissement de la vitesse de rotation des rouleaux d'entraînement à celle du déplacement réalisé par l'utilisateur. Un compromis peut alors être réalisé en imposant une vitesse constante de déroulement du tapis (donc en sortie système), impliquant de la part de l'utilisateur une adaptation de sa propre vitesse de marche, ce qui en fait alors davantage un dispositif de sortie si l'asservissement en entrée se limite à la seule commande "Marche-Arrêt" de la rotation des rouleaux d'entraînement. Il n'est par ailleurs guère possible de changer de direction de progression sans l'usage d'une modalité d'entrée supplémentaire qui va compliquer cette "locomotion" par un pilotage de l'environnement virtuel en rotation de lacet autour de l'utilisateur, ceci afin de lui permettre de sélectionner une nouvelle direction virtuelle de progression.

On constate, avec cet aperçu de diverses techniques de suivi des mouvements de l'utilisateur, une disparité importante qui s'explique en partie par les contraintes qu'apporte chaque technique, qu'il s'agisse du poids de l'équipement, de son encombrement, des limites de l'espace de travail, etc.

Les mouvements des membres

Cette catégorie concerne d'abord les jambes dont les mouvements sont le plus souvent une conséquence du déplacement de l'utilisateur dont le but n'est pas forcément de bouger les jambes pour signifier quoi que ce soit, mais seulement de se déplacer au sein de son environnement (il peut arriver que l'on désigne une direction à l'aide d'une jambe et de son pied alors qu'on a les mains occupées, mais le cas est tellement rare et malcommode qu'il mérite à peine d'être cité ici). Elle concerne ensuite et donc surtout les *mouvements des bras*. A part certains cas particuliers dans lesquels un utilisateur les agite pour attirer l'attention d'une personne distante ou pour diriger un orchestre, ceux-ci ne sont pas forcément la source d'une signification particulière, mais participent cependant bien souvent à une signification en situant les mains qui les terminent et qui, elles, peuvent exprimer une indication par leurs mouvements. Les hypothétiques mouvements d'épaules qui indiqueraient eux aussi quelque chose entrent aussi dans cette catégorie.

De la même façon que pour les déplacements du corps, les mouvements des bras constituent une information pouvant s'avérer nécessaire au système, il ne s'agit pas, là encore, d'une modalité d'entrée, mais du respect d'un affichage visuel cohérent traitant les occlusions (des bras, en l'occurrence). Il semble à cet égard que la majorité des systèmes mis en oeuvre reconstituent les mouvements des bras de l'utilisateur par cinématique inverse à partir des positions et orientations absolues de la tête et des mains de l'utilisateur fournis par quelque dispositif de Tracking fixé à l'environnement physique.

Pratiquement chaque méthode de tracking par capteur pourrait être utilisable au niveau du coude et de l'épaule pour signaler les mouvements (tracker magnétique, ultra-sonore, centrale inertielle...).



Figure 55. Partie scapulaire d'un exosquelette d'animation.

Les méthodes précédemment citées de tracking par *cameras d'observation* de l'utilisateur affublé de "marqueurs" sont là aussi mises en oeuvre pour les post-productions multimédias et sont à réserver pour des volumes circonscrits au champ couvert par les optiques des cameras.

Une autre technique consiste en un exosquelette d'animation dont les mouvements suivent ceux du corps de l'utilisateur (ex. la gamme "Gypsy", "Animazoo", Imagina et VRIC 2002, Figure 55). Cette technique reste encombrante et contraignante (pour

s'asseoir notamment), et s'avère moins précise que les positionnements absolus que

fournit par exemple le tracking magnétique, elle constitue cependant la seule technique nomade qui paraît améliorable au point de s'intégrer peut-être un jour aux vêtements de l'utilisateur.

Citons aussi les mesures par Electro-myographie, (Marrin Nakra, 2000, *cf.* Figure 56). Des électrodes fixées aux vêtements du chef d'orchestre transmettent les myogrammes relevés pour les muscles concernés par les mouvements étudiés (biceps, trapèze, etc.). Cette technique signale bien les mouvements de façon qualitative, mais pas de façon quantitative.

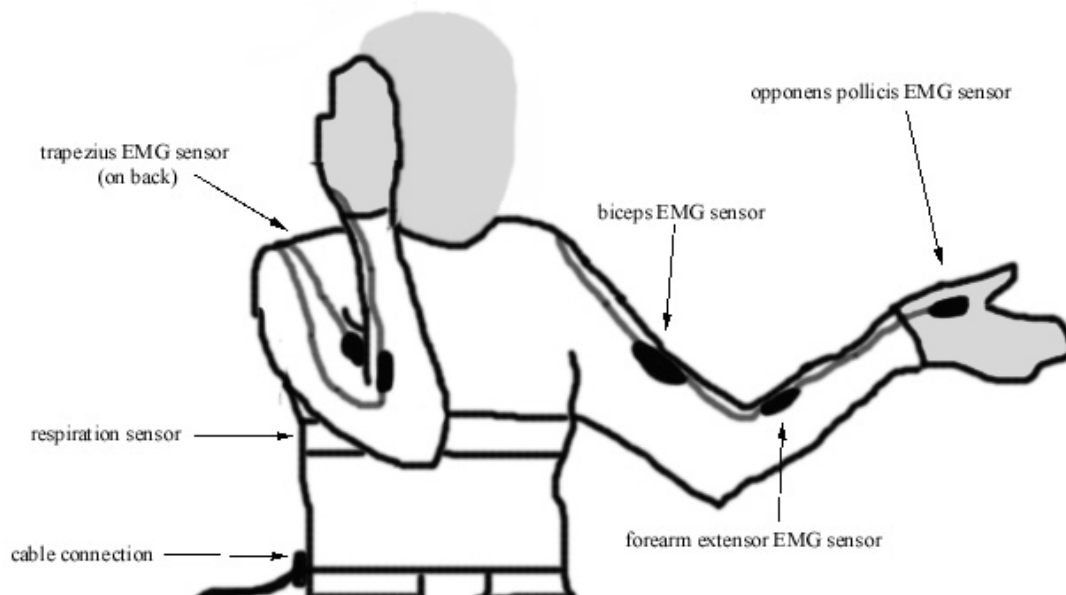


Figure 56. Principe de la "veste du chef d'orchestre" fonctionnant à l'aide de capteurs Electro-Myographiques.

Les mouvements des mains

Cette catégorie est sans doute la plus détaillée dans la littérature et fournit, sous son aspect sémiotique, quantité d'informations présentant une relation évidente avec une activité de conception. On y distinguera essentiellement d'abord les gestes déictiques (les "*déictiques spécifiques*" de Nespoulous & Lecours ou gestes "*Locative*" de Francis Quek). Puis les "*mimétiques*" décrivant soit une forme spatiale ou indiquant des paramètres dimensionnels, soit des descriptions de mouvements à réaliser ou à imposer à un objet ou à son propre viewpoint.

De même que pour les déplacements de l'utilisateur et les mouvements de ses bras, les informations sur les mouvements des doigts sont aussi à prendre en compte pour un affichage visuel en sortie qui aurait à traiter les occlusions.

T.G. Zimmerman *et al.* développèrent en 1987 chez VPL Research un "Gant de Données" (*cf.* Figure 57) dont la vocation était de permettre à l'utilisateur de manipuler des objets graphiques virtuels générés par la machine comme s'ils existaient réellement (sans retour tactilo-kinesthésique), et de permettre d'autre part l'interprétation de signes alphabétiques pour malentendants, l'évaluation clinique de

troubles fonctionnels de la main, ainsi que son utilisation comme interface avec un langage de programmation visuelle.

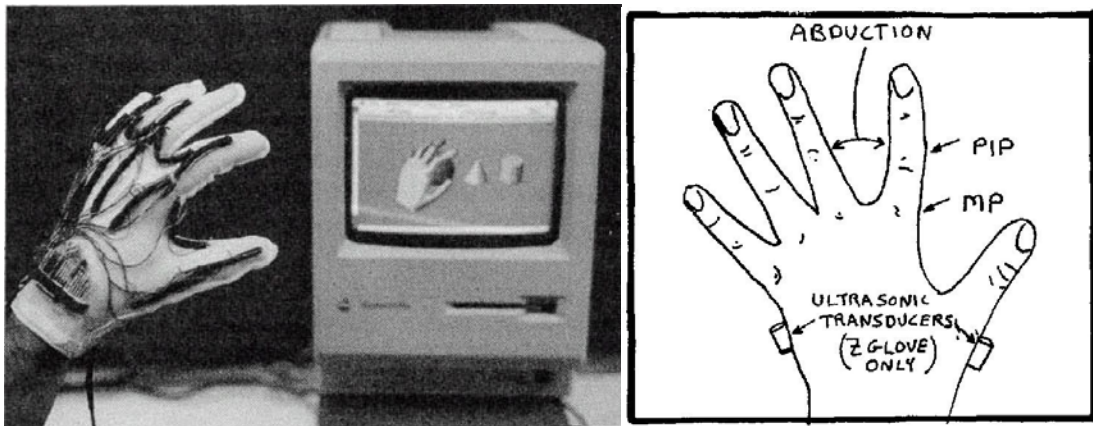


Figure 57. Le "Gant de Données".

Il s'agit d'une forme d'exosquelette captant les mouvements des doigts, par diverses techniques possibles (potentiomètres sur jointures, effet Hall, jauges de contraintes, déperdition optique dans une fibre, etc...), ce qui en fait une technique de capture centrée utilisateur. La position de la main est relevée généralement par quelque technique exocentrique consistant en un dispositif de mesure supplémentaire placé de façon fixe dans l'environnement physique de l'utilisateur, déterminant à distance les variations d'un principe physique, tel qu'un champ magnétique, dues au déplacement de transducteurs fixés, eux, sur le Gant de Données (Tracker Magnétique, Tracker UltraSonore, etc., éventuellement mécanique par rapport à une position fixe). Ce qui en fait généralement une technique mixte cumulant à la fois les contraintes d'une capture centrée utilisateur encore encombrante (port du gant, gêne des mouvements) ainsi que celles d'une capture exocentrique (utilisable uniquement à proximité de l'aire de capture).

Plusieurs technologies sont donc présentes que l'on peut maintenant, d'un point de vue utilisateur, classer essentiellement en techniques exocentriques ("*Camera based*", Quek, 1994), vision based, Tracking Magnétique, dont le point commun reste l'observation et le suivi de la gestuelle utilisateur à distance et par rapport à un repère fixe de l'environnement physique), et techniques davantage centrées utilisateur, (Exosquelettes mécaniques utilisant différentes techniques de capture de flexion des articulations, Capture EMG).

Les mouvements de la Tête

On trouvera d'abord, dans cette catégorie, le même type de relation avec les mouvements oculaires (évoqués ci-après) que celui qui existe entre les bras et les mains précédemment évoqué. Les mouvements de la tête présentent parfois une signification comme c'est par exemple le cas lorsqu'on indique une direction du "menton" (qui s'accompagnera d'ailleurs généralement d'un regard soutenu), ou encore par un mouvement latéral d'inclinaison rapide de la tête pour indiquer une

direction générale relative à l'objet du discours, autre que la direction observée du regard.

En dehors de cela, l'essentiel des mouvements de la tête sert à rendre un objet accessible au regard.

Les informations de position et orientation de la tête sont parfois considérées à elles seules comme représentatives d'une direction approximative du regard de l'utilisateur (on ne peut en effet observer que vers l'avant de la tête, avec une tendance naturelle à centrer le sujet d'observation dans le champ de vision quand c'est possible par un mouvement de la tête, l'amenant à faire face au sujet d'intérêt observé). Ces informations sont aussi, par ailleurs, nécessaires à certains affichages visuels qui simulent un champ de vision asservi aux mouvements de la tête de l'utilisateur (cas des HMD).

Tout comme pour le suivi des mouvements des bras ou des mains, celui des mouvements de la tête fait souvent appel aux techniques classiques de tracking magnétique ou ultra-sonore mentionnées ci-dessus, parfois par caméras disposées au sein de l'environnement, et plus rarement de façon mécanique (Ware *et al.*, 1993). Dans le cadre du "*wearable computing*", d'autres principes sont mis en œuvre et installés sur un casque porté par l'utilisateur pour devenir compatibles avec des applications nomades, tels que GPS, Compas électronique, centrales inertielles et reconnaissance visuelle active de l'environnement par caméra (cas du projet "ArcheoGuide").

Les mouvements des yeux et ce qui en fait une forme supplémentaire de gestuelle

Pour reprendre les termes cités par Eibl-Eibesfeldt (1984):

" Notre répertoire de mouvements d'expression est très riche, nous pouvons signaler avec nos yeux. (M. v. Cranach, 1971 et p. 190-191). L'évolution de globes oculaires clairs, qui contrastent bien avec l'iris, déjà amorcée chez les singes anthropomorphes, se développa chez l'homme au point que les mouvements des yeux ont une fonction de signalisation. "

Irenaus Eibl-Eibesfeldt, " Ethologie - Biologie du Comportement ", Naturalia et Biologia, 1984 (3^{me} édition), p.617.

Ce type de mouvement ne reflète pas, généralement, de mouvements volontaires. C'est sans doute la principale raison qui n'en fait pas une "gestuelle" à part entière dans la littérature. Il existe de rares cas de la vie courante dans lesquels on utilise la direction du regard pour indiquer une direction ou même des mouvements de paupière pour signaler une connivence, mais d'une manière générale, l'oeil n'est pas à considérer comme un organe de commande. Toutefois, les mouvements involontaires produits par les yeux sont porteurs d'indices de par la nature même de l'usage quotidien que nous faisons du sens visuel.

Il est même possible de déterminer la distance approximative d'observation d'un utilisateur par le simple aperçu de ses deux yeux dont la position des iris révèle la convergence.

Selon Murphy et Duchowski, (2001), l'usage de l'Eye-Tracking distingue deux principales catégories d'applications, les applications de "*diagnostic*" et les applications "*interactives*", (cf. Figure 58).

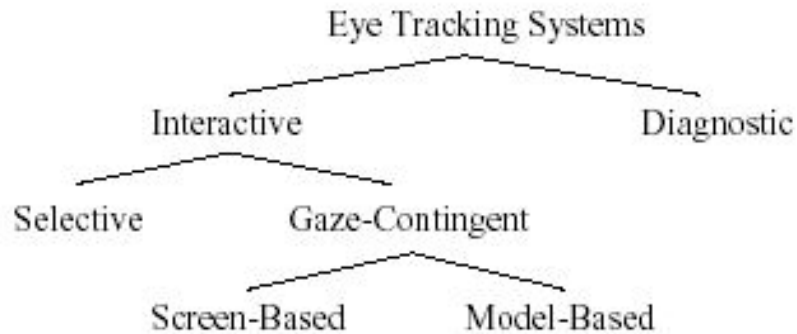


Figure 58 Hiérarchie des applications d'Eye-Tracking.

Les applications dites de "*diagnostic*" impliquent un enregistrement des mouvements oculaires au cours du temps en vue d'obtenir un relevé des trajectoires observées par l'utilisateur (connues sous le nom de "*Scanpath*"), et ne requièrent généralement pas de réaction au regard de l'utilisateur de la part du système qui serait reflétée par l'affichage.

Les applications "*interactives*" se caractérisent par une réaction de l'affichage simultanément à la direction du regard de l'utilisateur. L'auteur y distingue deux sous-types d'applications, les Applications dites "*Selective*" et celles baptisées "*Gaze-Contingent*". Les "*Selective Applications*" mettent en scène l'Eye-Tracker comme dispositif d'entrée alternatif tel que par exemple pour le pointage (ex. Tanriverdi & Jacob, 2000). Leur propos se situant par rapport au "*Perceptually-Adaptive Computer Graphics*" (PACG), les auteurs précisent que la motivation principale des "*Gaze-Contingent systems*" est de diminuer la bande passante requise par l'affichage de façon à l'adapter aux variations d'acuité de la vision humaine au sein du champ de vision. Une haute résolution à l'affichage ne serait ainsi appliquée qu'à la zone couverte par ce qu'ils appellent la "*foveal Region Of Interest*" (ROI), cependant qu'une moindre résolution serait appliquée au champ de vision périphérique (en réduisant par exemple le spectre de couleurs auquel le champ de vision périphérique est moins sensible chez l'Homme du fait d'une moindre concentration en cônes rétiniens).

Murphy et Duchowski, (2001), distinguent là encore deux catégories d'approches traitant le problème, les approches "*screen-based*" et "*model-based*".

L'approche de type "*screen-based*" implique la manipulation du contenu du Frame-Buffer juste avant l'affichage, par des techniques de traitement d'image.

L'approche dite "*model-based*", moins courante au gré de l'auteur, cherche à manipuler la résolution au niveau du modèle géométrique à afficher avant l'étape de rendering.

Quelques applications choisies dans la littérature

Pour aborder maintenant le sujet des applications que l'on trouve dans la littérature, Ware & Mikaelian, présentent en 1987 trois méthodes de sélection d'un objet fixé par le regard sur un moniteur, (baptisées "*dwelling*", "*Screen*" et "*hardware*"). La première méthode, ("*dwelling-time*"), consiste en une fixation prolongée avec le regard pour signaler à la fois pointage et sélection (par fixation de plus de 0.4s). La méthode dite "*Screen*" permet la sélection d'un item dans une barre de menu verticale (à droite sur la reproduction Figure 59), puis confirmation de validation de la sélection par fixation consécutive de l'image d'un "Bouton" de validation situé à côté du menu, (à gauche sur la reproduction Figure 59). La dernière méthode expérimentée, ("*hardware*"), consiste à valider la sélection d'un élément du menu par bouton poussoir physique de validation, sur clavier.

Les résultats semblent montrer que les méthodes "*Dwell*" et "*Hardware*" présentent un avantage d'environ 15 % sur la méthode "*Screen*", avec un léger avantage supplémentaire pour la méthode "*Hardware*" qui semble moins sensible à la distance du point initial à la cible que la méthode "*Dwell*". (cela suggère la possibilité d'une éventuelle surcharge oculaire de la méthode "*Dwell*", et/ou un gain dû à l'emploi d'une seconde modalité de validation, le gaze-tracking ne constituant dans ce cas qu'un pointage en attente de validation pour obtenir l'opération complète de sélection). Les résultats montrent qu'un eye-tracker peut être utilisé comme un moyen de sélection d'item sur un moniteur, de façon plus rapide que les pointeurs manuels (approximativement 60%) pourvu que la taille de la cible ne soit pas "trop petite".

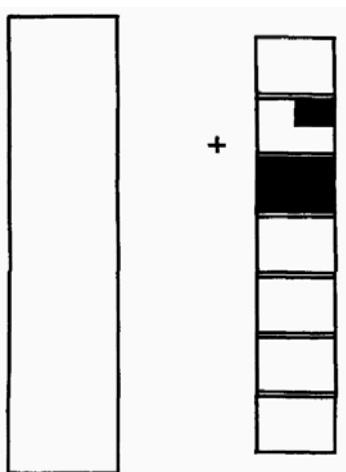


Figure 59 "Screen method".

On notera que dans tous les cas présentés, la sélection réalisée pourrait se passer de clavier ou de souris, et que la sélection y est réalisée plus rapidement qu'avec un pointeur manuel. En revanche, ces méthodes sont contraignantes en mobilisant un contrôle du regard afin d'en obtenir un organe de commande (ce pour quoi l'oeil n'est pas apte, sauf en de rares occasions). D'autre part, la méthode "*Dwell*" présente l'inconvénient d'éventuellement provoquer une activation involontaire de la sélection d'un élément sur lequel l'utilisateur ne fait par exemple que s'interroger en cherchant à l'identifier (aussi connu sous le terme de "*Midas Touch Effect*").

Starker & Bolt (1990), tentent d'implémenter de

façon précoce une interaction de type "Story-telling" ajustant les comportements narratifs en réponse au regard de l'utilisateur qui est alors supposé refléter son intérêt. Les auteurs en concluent qu'apparemment, le regard de l'utilisateur reflète effectivement son intérêt, du moins dans le contexte particulier décrit qui relève plutôt du multimédia interactif et qui préfigure, ou du moins comporte beaucoup de points communs avec un environnement d'agents réactifs du domaine de l'Intelligence Artificielle.

Jacob (1991), développe une étude qui adresse l'usage de l'Eye-tracker par une approche générale comme intervenant possible dans l'IHM en tant que canal d'entrée, sans réellement aborder les problèmes de comparaison avec d'autres médias.

L'auteur explique les diverses technologies existantes pour le suivi du regard, les difficultés inhérentes à ce qui est mesuré, évoque aussi vaguement les dispositions algorithmiques à adopter pour qu'un suivi du regard fonctionne. Il commente par ailleurs les aléas de fonctionnement et le manque de maturité de la technologie (encombrement, bruit, imprécision, limites de fonctionnement).

Plus récemment, Zhai *et al.*, (1998), ont utilisé la technique de Gaze-Tracking pour améliorer une interaction manuelle déjà existante avec la souris. (Ceci afin de tenir compte de l'imprécision de la position fournie par l'Eye Tracker). Ce qui semble effectivement apporter une amélioration pour les cibles distantes de l'origine, pour le pointage desquelles la loi de Fitts devient prépondérante lors du pointage manuel. Le pointage par Eye-Tracker, même imprécis, apporte la possibilité d'une sorte de "prépointage" grossier (warp) "précisé" par une autre technique d'interaction (en l'occurrence ici, un pointeur manuel comme la souris, mais rien à priori n'exclut d'envisager une autre modalité précisant la cible par un autre moyen).

Ils évoquent aussi leur position concernant d'une part le fait que l'oeil est essentiellement un "*primary perceptual device*", et qu'il n'a pas évolué pour devenir un organe de contrôle. Raison pour laquelle ni la sélection par "*dwelling time*" ni celle par "*blinking*" (*cf.* Ware & Mikaelian, 1987), ne sont envisagées, l'oeil doit selon eux rester une modalité passive.

Pour citer les auteurs;

"Furthermore, dwell time can only substitute for one mouse click. There are often two steps to target activation. A single click selects the target (e.g., an application icon) and a double click (or a different physical button click) opens the icon (e.g., launches an application). To perform both steps with dwell time is even more difficult".

On retrouve donc ici l'idée d'un signal de validation faisant appel à une autre modalité pour signifier la sélection.

En 1999, Salvucci expérimente l'usage d'un Eye-Tracker pour étudier la saisie de caractères entrés par scrutation volontaire d'un clavier virtuel affiché à l'écran. Cet

article traite essentiellement encore des problèmes techniques (en particulier d'algorithmiques et de traitement des signaux fournis par l'appareillage) que rencontre la technologie d'Eye-tracking, et l'auteur conclut qu'un suivi du regard par chaînes de Markov améliore la possibilité d'utiliser l'Eye-tracking (dans ce contexte).

Tanriverdi & Jacob, (2000), semblent être les premiers à tester un "head-mounted eye tracker in a virtual environment display". La Figure 60 illustre cette interaction à l'aide du regard, l'image du haut montre l'environnement, celle en bas à gauche est un grossissement montrant l'objet mauve cible du regard, on voit en bas à droite l'affichage par transparence du contenu de la cible. Le but annoncé consiste à tester l'hypothèse selon laquelle une interaction en environnement virtuel basée sur les mouvements oculaires serait plus performante que les autres types d'interaction naturels. Le résultat de leurs investigations réside en ce que le suivi du regard est effectivement une méthode rapide de pointage, (Mais, accessoirement, que son usage unique révélerait a priori un affaiblissement de la mémorisation spatiale.). On retiendra que l'opération de pointage et sélection d'objets 3D rapprochés par suivi du regard en environnement virtuel est aussi rapide que le pointage manuel, et que le premier est par contre plus rapide pour les objets distants (à relativiser toutefois au profit de l'Eye-Tracking compte tenu des différences de maturité technique respective des dispositifs). Il semblerait que l'Eye-Tracking, de par un "engagement cognitif de l'utilisateur" plus réduit qu'avec la sélection manuelle, du moins serait-ce l'explication des auteurs, affaiblisse la mémorisation spatiale des objets et de leur contenu...

On notera l'absence d'information concernant l'influence comparée de ces techniques sur la fatigue de l'usager.

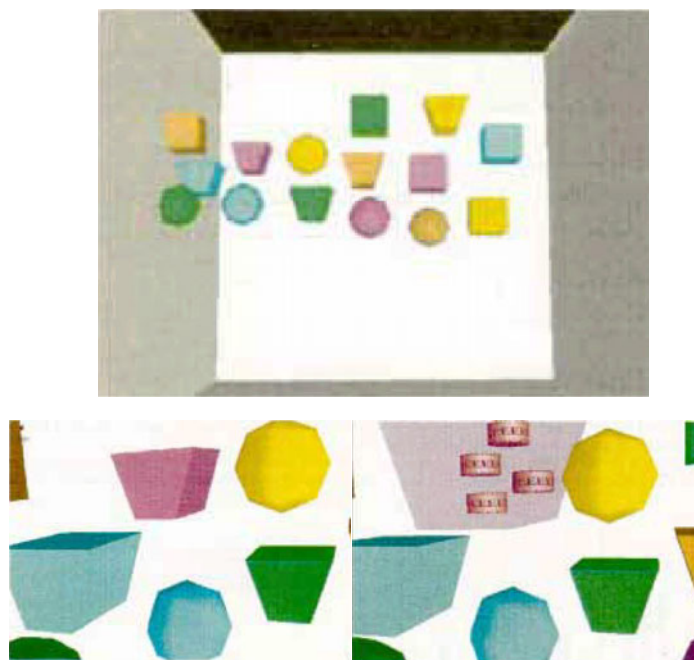


Figure 60 Interaction par le regard avec un environnement virtuel.

Salvucci & Anderson, (2000), expérimentent l'ajout de la direction du regard à un WIMP⁵ standard en tant qu'entrée alternative utilisée au gré de l'utilisateur, baptisé IGO, (Intelligent Gaze-added OS), expliqué à travers ses trois composants :

Basic non-gaze interpretation.

Gaze-added interpretation.

Intelligent gaze interpretation.

L'Eye-tracking mis en oeuvre agit comme pointeur alternatif dans le deuxième composant, et à la fois, dans le troisième composant, comme pointeur et actuateur - *eg.* un "*gaze button*", c'est à dire un bouton physique et non pas un bouton visuel de validation-, et adresse les opérations du GUI⁶ telles que "drag & drop", "window opening", " window closing", etc. comme au sein d'un WIMP standard.

Les conclusions évoquées permettent d'envisager les opérations d'un WIMP (d'une part à défaut de souris, d'autre part comme canal d'entrée additionnel au choix de l'utilisateur), comme cela pourrait être envisageable par la suite au sein d'un environnement 3D (encore que le "*Drag & Drop*" fut par exemple en particulier dénoncé par plusieurs utilisateurs comme difficile). Mais les auteurs, qui établissent que ces mêmes utilisateurs réservaient alors le gaze-tracking pour les "actions simples" en l'évitant pour les "actions complexes", ne concluent pas que c'est éventuellement une action motrice active du regard qui pouvait peut-être s'avérer gênante, indiquant alors peut-être que les utilisateurs cherchaient à tirer profit uniquement de ce qu'apporte l'action passive résultante du suivi du regard. A noter d'autre part qu'apparemment, les données concernant 3 utilisateurs sur 13 ne furent pas prises en compte pour des raisons d'un "*extreme noise in their eye-tracking data*", ce qui est une mention sous-entendue supplémentaire à l'évocation précédente de la "variabilité des utilisateurs". D'où il faut peut-être retenir un besoin auto-adaptatif individualisé à chaque utilisateur dans le traitement logiciel à appliquer aux signaux issus du dispositif.

On retiendra ici que l'on peut, au choix, pointer avec le regard et valider la sélection d'une cible en 2D à l'aide d'un bouton, (mais c'est extensible à d'éventuelles icônes 3D), pourvu que l'interface soit conçue de telle sorte que les dites cibles soient suffisamment distantes les unes des autres eu égard au manque de précision (temporaire) des dispositifs d'Eye-Tracking existants, mais aussi au manque de précision du regard.

Sibert & Jacob, (2000), rapportent les résultats de comparaisons des durées de sélection à l'écran pour deux tâches de sélection avec l'Eye-Tracking d'une part, et avec la souris d'autre part.

⁵ « WIMP » : "*Windows-Icon-Menu-Pointing*", ou pointage sur fenêtres présentant icônes et menus.

⁶ « GUI » : "*Graphical User Interface*", ou Interface Graphique Utilisateur.

Ils sont semble-t-il les premiers à évoquer la possibilité d'utiliser le suivi du regard pour "*libérer les mains de l'utilisateur pour d'autres tâches*". Aucun exemple n'est fourni en la matière par les auteurs, mais on peut cependant imaginer un apport pour des tâches composites.

Les auteurs concluent d'autre part que l'Eye tracking se conçoit comme un canal d'entrée supplémentaire, qui non seulement se révèle plus rapide que le pointage/sélection par la souris pour des tâches simples, mais qui de plus permet (implicitement à l'idée de libérer les mains pour d'autres tâches) de se passer de dispositif d'entrée "manuelle" de pointage/sélection.

Raison pour laquelle ils suggèrent aux concepteurs d'interfaces utilisateur du futur de prévoir ce canal d'entrée, même en tant qu'option (sous-entendu).

On retiendra qu'il devient ainsi possible de se passer de dispositif d'entrée "manuelle" de pointage/sélection tel que la souris pour pointer/sélectionner un objet visible dans l'environnement d'affichage, si l'on utilise la technique d'eye-Tracking.

A noter enfin que selon les mêmes auteurs, " With tasks that are well-structured and speeded, research has shown that people look at what they are working on " (Just, & Carpenter, 1976). Ce texte cité en référence traite en effet de l'influence de quelques représentations internes des objets de la tâche chez l'utilisateur sur les fixations oculaires (ex. rotation mentale d'une figure tridimensionnelle), et rapporte entre autres que le locus of fixation n'est pas toujours synonyme de direction d'attention (p.474). Argument qui militerait en faveur de l'usage d'une modalité complémentaire validant la sélection éventuelle de quelque objet de la tâche dans l'environnement de travail de l'utilisateur.

Par ailleurs, pour ce qui concerne l'information de profondeur de scrutation de l'utilisateur, il existe au moins un autre exemple (au sein de la littérature), mettant en scène des *HMDs intégrant une paire d'Eye-Trackers* qui permettent ainsi la mesure du point de convergence du regard utilisateur, (Danforth *et al.*, 2000). Les auteurs mentionnent d'autre part l'indication de Mel Slater (1997), selon laquelle l'immersion en environnement virtuel requiert une représentation de soi-même (le "*Virtual Body*"), représentant à la fois une partie de l'environnement perçu ainsi qu'une image de celui qui perçoit, en ajoutant qu'à cette date, aucune expérience formelle n'a encore été conduite pour tester ce type de spécification proprioceptive. Les auteurs estiment néanmoins qu'une telle représentation de soi-même au sein de l'environnement virtuel augmentera les performances utilisateur pour nombre de tâches.

A cette fin, ils font appel à la technique décrite par Van Pernis en 1999. Cette technique consiste en une méthode de post-production impliquant une caméra montée sur les HMDs et pointant vers l'avant de la tête de l'utilisateur. Elle implique aussi un environnement physique de l'utilisateur uniformément bleu afin d'en extraire par "Chroma-Key" tout ce qui n'est pas bleu dans l'image captée par la caméra (en particulier, ses mains, ses bras, éventuellement ses jambes et ses pieds).

Ces portions d'images sont ensuite ajoutées par "alpha-blending" au contenu du Frame-Buffer contenant les images synthétiques.

La technique revient à un placage d'images, sans modèle 3D qui permettrait au système de gérer les occlusions. Pour ce faire, les auteurs font appel à un tracker magnétique qui, fixé sur la main, permettrait de reconstituer un pseudo-modèle 3D (l'occlusion n'est pas encore implémentée à la date de parution de cet article).

Le rapport avec les recherches sur la dégradation du niveau de détail à l'affichage en périphérie du champ de vision, ("level of detail", LOD), réside en ce que l'approche "*Model Based*" -mentionnée plus haut- permettrait au système, dans une configuration [HMD + Eye-Tracker binoculaire intégré], de déterminer directement quel est l'objet visé par le regard de l'utilisateur. Cet objet graphique 3D résidant modélisé en mémoire système, ses caractéristiques sont connues du système, en particulier sa distance virtuelle au "view-point" de l'utilisateur, ce qui permettrait, à l'aide d'une optique adéquate, d'appliquer une correction oculaire évitant le défaut d'accommodation à l'affichage (source, rappelons-le, de conflit visuel).

Le rapport entre le mouvement des yeux de l'utilisateur et une "gestuelle"

La direction du regard est observée par le suivi du mouvement des yeux. Sauf dans certains cas particuliers (clin d'oeil, déictique oculaire signifiant volontairement une direction avec exagération de l'ouverture des paupières), ce mouvement est généralement davantage, de la part de l'utilisateur, une conséquence de l'observation visuelle de l'environnement qu'un mouvement volontaire exprimant une information.

On trouve, dans la littérature, (concernant "*L'Adressage des Objets 3D, Pointage 3D et Sélection*"), deux techniques particulières de sélection baptisées respectivement "*Aperture Based Selection*" et "*Image Plane*" impliquant toutes deux une gestuelle volontaire de la main dominante de l'utilisateur pour définir l'angle solide d'un volume conique de sélection dont la position de l'apex est confondue avec le "*viewpoint*" de l'utilisateur.

La première technique évoquée implique un dispositif matériel de pointage tenu en main, dont les position et orientation dans l'espace sont connues du système (via quelque méthode de tracking). Le positionnement de l'extrémité du pointeur dans l'espace sera généralement superposé par l'utilisateur à l'objet d'intérêt à sélectionner vu depuis son oeil dominant (qui sera aussi le "viewpoint" pour cet oeil), cependant que l'éloignement ou rapprochement du dispositif aura pour effet de faire varier l'angle solide du cône, (*cf.* Figure 74 p. 92).

La seconde technique évoquée, "*Image Plane*", fonctionne selon le même principe en remplaçant le dispositif de pointage matériel par un dispositif non seulement de tracking de la main de l'utilisateur, mais aussi de traitement des mouvements des doigts de la même main dont un geste particulier initiera le processus de sélection au sein d'un volume conique dont l'angle solide varie au gré de l'écartement du pouce et

de l'index ainsi que de la distance de la main à l'oeil dominant (le viewpoint). (cf. Figure 75 p. 92).

La direction du regard de l'utilisateur, (idéalement matérialisée en vert sur le schéma modifié Figure 75 p. 92), apparaît finalement, après superposition par l'utilisateur du mécanisme de pointage sur l'objet graphique 3D représenté de son choix, contenu au sein du volume conique ainsi déterminé.

On peut définir trois informations nécessaires à la mise en oeuvre de ces techniques de sélection, la *position de l'apex du cône de sélection*, (reste fixée au viewpoint assimilé à l'oeil dominant de l'utilisateur), puis l'*orientation du cône* défini par une direction issue du tracking de position du dispositif manipulable de pointage ou de la main de l'utilisateur, et enfin la *modification du paramètre réglant l'ouverture du cône*.

Remplacer l'orientation du cône ainsi effectuée "*à la main*" par la direction du regard que fournirait un Eye-Tracker reviendrait à remplacer le geste manuel d'orientation par un "*geste oculaire*" (involontaire). Cette seule opération n'impliquerait plus l'usage des mains de l'utilisateur. On notera que la direction du regard ainsi employée ne permettrait plus dans ce cas de modifier aussi à elle seule le paramètre réglant l'ouverture du cône comme cela est permis dans les deux techniques évoquées, mais qu'elle ne nécessiterait cependant plus l'usage de la main dominante dont elle remplace le geste d'orientation.

Techniques de capture des mouvements des yeux

Les anciennes techniques myographiques contraignantes pour l'utilisateur ayant été remplacées par l'usage des InfraRouge réfléchis sur la cornée, les techniques qui se répandent maintenant pour le suivi du mouvement des yeux mettent en oeuvre des caméras IR distantes ("*remote Eye-Tracking*") ou portables sous la forme de casques ou de lunettes.

Le "*Remote Eye-Tracker*" (observation distante par caméra) est à réserver pour le repérage d'une position visée par l'un des yeux de l'utilisateur sur un écran de taille modeste en raison de l'obligation de placer la caméra en bordure d'écran et de l'angle maximal de l'optique de la camera (on ne pourra par exemple pas l'utiliser avec l'affichage offert par le WorkBench).

Comme déjà signalé, il existe au moins un exemple mettant en scène des *HMDs intégrant une paire d'Eye-Trackers* qui permettent ainsi par ailleurs la mesure du point de convergence du regard utilisateur (Danforth *et al.* 2000). Les auteurs indiquent que la seule mention précédente à leur connaissance est celle d'Ohshima *et al.*, (1996). Une mesure de convergence sur les yeux permet de déterminer de façon approximative la distance à laquelle regarde l'utilisateur.

Cette intégration du principe du Gaze-Tracking à des lunettes se révèle d'intérêt non seulement dans le cas de l'usage d'une paire de HMD, mais également

concernant les paires de Shutters utilisées pour les affichages par projection sur grands écrans (cas, par exemple, du WorkBench).

Extension des mouvements de la main, les gestuelles instrumentées.

Ce sujet aurait pu être abordé à la suite du paragraphe traitant du mouvement des mains, tellement il nous paraît naturel d'utiliser cet appendice préhensile de façon quasiment exclusive pour l'usage d'outils dédiés à diverses tâches. L'idée d'instrumenter une partie du corps afin de réaliser une tâche particulière ne devrait cependant pas concerner uniquement cet appendice. Par ailleurs, la dimension supplémentaire apportée par l'usage d'outils préhensibles interchangeables multiplie le nombre de tâches réalisables, ce qui contribue à en faire une catégorie particulière.

Comme signalé précédemment, l'une des premières définitions de "la gestuelle" en 1987 adressa les mouvements d'un stylos sur tablette graphique. Un tel stylos trouve aujourd'hui des extensions en environnement 3D sous diverses formes telles que "stylos 3D" (équipant par exemple le "workbench") ou autre souris 3D pour le même type d'interaction utilisateur (généralement, quelque forme de manipulation directe).

Pour Cadoz, Le geste instrumental n'est pas une catégorie de geste supplémentaire, mais une "modalité de communication" complémentaire au "geste à main nue", cf. "*Le geste, canal de communication homme/machine; la communication « instrumentale »*" (Cadoz, 1994).

Ce type de gestuelles complémentaires qui vient enrichir le "geste à main nue" ne sera pas exploré ici davantage que par un exemple significatif. Il sera sans doute traité ultérieurement en tant que prolongement sémiotique des mouvements de la main à laquelle il fournit des fonctionnalités supplémentaires.

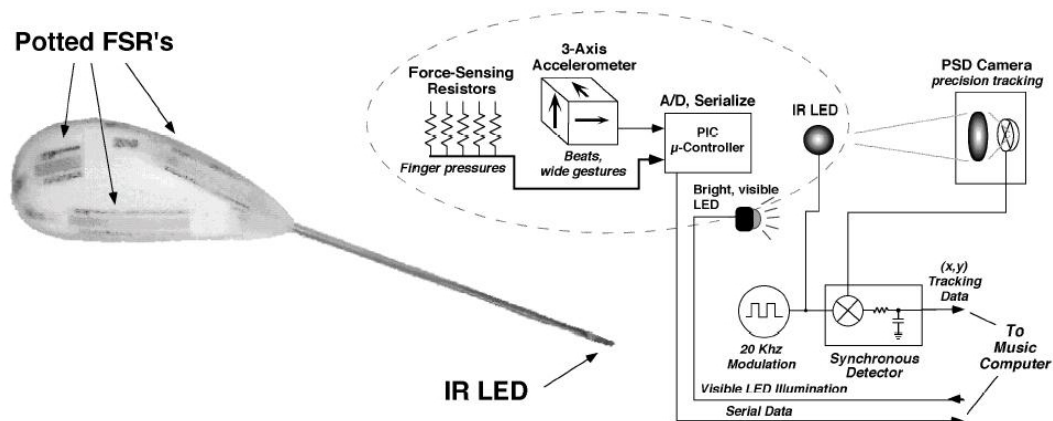


Figure 61 Le "Digital Baton".

Un exemple d'implémentation d'une telle gestuelle instrumentée consiste en une version augmentée de la baguette du chef d'orchestre, le "*Digital Baton*", (Paradiso & Sparacino, 1997. cf. Figure 61). Celle-ci contient des accéléromètres sur les trois axes ainsi qu'une LED InfraRouge dont la position est relevée par des caméras de façon

exocentrique. L'ensemble permet au système de suivre les mouvements directeurs du chef d'orchestre au cours de sa prestation (cf. Figure 56 p. 67).

Signalons pour terminer qu'il existe de nombreux autres dispositifs de manipulation de pointage 3D (souris et autres stylos 3D) ou d'appariement d'un objet sélectionné de la scène, voire de la scène toute entière (ex. "*World In Miniature*", cf. Stoakley *et al.*, 1995, évoqué plus loin), ou encore un autre principe relevant de la poupée vaudou et prenant le nom de "*props*" (ex. "*The FingerBall*", Zhai *et al.*, 1996), dispositifs que nous évoquerons au chapitre suivant avec les métaphores d'interaction.

3.2.2. La parole

Les actes de langages

En 1995, Ferber présente, au sujet de la communication au sein des systèmes multi-agents (Ferber, 1995, chapt.6), une modification de la classification des principales catégories d'actes de langage de Searle (1979) et Vanderveken (1988), "*assertifs, directifs, promissifs, expressifs, déclaratifs*" qu'il juge trop générale au point de ne pouvoir être utilisée telle quelle pour réaliser des systèmes multi-agents. Il propose une classification plus fine qui s'effectue en deux temps. D'une part, les directifs sont décomposés en interrogatifs qui servent à poser une question pour obtenir des informations et en exercitifs qui demandent à un agent d'accomplir une action dans le monde. D'autre part, il raffine son analyse en définissant des "sous-actes" de langage en détaillant l'objet (ou l'action) sur lequel porte l'acte et, surtout, en caractérisant la structure conversationnelle dans laquelle il s'insère. Il conserve les trois composantes des actes de langage de Austin (1962) et surtout Searle (1969), considérées comme des actes élémentaires (*composantes locutoire, illocutoire, perlocutoire*), et considère, lorsqu'il est question d'actes de langage, faire implicitement référence à l'ensemble de ces trois aspects et non pas seulement aux actes illocutoires comme le fait Searle.

Une telle adaptation de la communication aux agents aboutit ainsi à un modèle qui semble posséder, du moins dans ses grandes lignes, les mêmes caractéristiques que celui qui mettrait en jeu un locuteur humain avec un destinataire machine ou Objets (Agents) de l'environnement adressé. Les caractéristiques des divers types de communication envisagés et expérimentés entre agents pourrait ainsi éventuellement être directement mis à profit dans le cadre d'une interaction Homme-Machine (ou Homme-Objet/Agent) orientée Communication.

Malheureusement, bien qu'on y trouve des éléments intéressants sur notre sujet, il semble que la source principale d'exemples dans ce domaine (l'interaction avec un acteur agent-virtuel s'inspirant des actes communicatifs) mette surtout l'accent sur la reproduction des indices de communication chez l'acteur virtuel dans le cadre d'une interaction utilisateur apparentée à une situation de dialogue à statut social équivalent.

Le domaine qui nous intéresse ici plus particulièrement relève davantage d'un dialogue dans lequel l'utilisateur est le siège du "*Locus of Control*" et le système, l'exécutant des ordres de l'utilisateur.

L'un des rares exemples d'une telle situation décrite dans la littérature est celui présenté par Chalmé *et al.*, (1999) lors de la conférence DSC'99. Cet exemple étudie la relation entre pilote et copilote en vue d'implémenter une aide à la navigation routière. Cet exemple, bien qu'intéressant, figure malheureusement la situation inverse de celle qui nous intéresse, car lorsqu'il est sollicité, c'est le copilote informatique qui "dirige la manoeuvre". Dans le cas qui nous occupe, transposé à cette situation, l'utilisateur aurait plutôt la place du copilote, et le système, celui du chauffeur.

Une communication inspirée de la théorie des "actes de langage"

La reconnaissance syntaxique "d'actes d'expression utilisateur", ("*Communicative acts*" selon C. Pelachaud et I. Poggi), acheminés par les modalités d'entrée vocale et/ou gestuelle corporelle et/ou "gestuelle" oculaire permet d'envisager l'identification d'une syntaxe de commande système destinée par exemple à un environnement CAD à défaut du système lui-même. Cela resterait somme toute une maigre évolution sur la ligne de commande entrée au clavier si l'utilisateur reste ainsi contraint à une syntaxe prédéfinie de commande.

La transposition *visuelle* d'une commande par *pointeur* propre à la "manipulation directe" d'icônes-objets et d'icônes-programmes sur l'interface graphique 2D du GUI a fait le succès du WIMP, grâce, entre autres, à une pseudo-reconnaissance de la commande utilisateur en fonction du contexte de la séquence initiée par l'utilisateur (ex. Drag & Drop d'une icône-objet dans une icône-programme ou dans un répertoire, l'un pour signifier l'ouverture du document par le programme, l'autre pour signifier le déplacement ou copie du document).

Une telle "interprétation" de la séquence de commandes graphiques utilisateur reste semble-t-il à découvrir concernant l'usage des nouvelles modalités d'entrée considérées, en particulier l'interprétation, même syntaxique, de la séquence produite par l'utilisateur au travers des modalités alternatives considérées (vocal + gestuelles corporelles + gestuelle oculaire), afin d'aboutir à une suite de commandes admissibles (compréhensibles, ou plutôt "pertinentes") pour le système ou programme considéré.

Dans un article datant de fin 2001 intitulé "Universal Speech Interfaces", Rosenfeld, Olsen & Rudnický (2001) discutent trois approches possibles concernant une exploitation de la reconnaissance vocale en entrée.

La première est l'usage du "*Command and Control*" dont on peut se faire une idée en se référant aux entrées de type "ligne de commande" obéissant à une syntaxe stricte et rigide que l'utilisateur se doit d'apprendre pour chaque application afin de pouvoir utiliser le système.

A l'autre extrémité est l'implémentation du *langage naturel* que l'on pourrait supposer aisément utilisable, mais requérant une très lourde charge de travail initial pour le concepteur qui doit prévoir toutes les possibilités entrées par l'utilisateur final.

Entre ces deux extrêmes se trouve ce que les auteurs qualifient de "*Dialog Trees*". Cette solution aurait l'avantage de réduire la complexité de la reconnaissance en réduisant cette activité au domaine de l'application en une série de termes sélectionnés dont un utilisateur pourrait aussi bien faire usage au sein d'un ensemble d'alternatives possibles que répondre verbalement à une question spécifique (tel qu'un nom ou une quantité). Les inconvénients de tels systèmes, du point de vue de l'utilisateur, résident en l'impossibilité d'accéder directement à différentes parties d'un domaine qui sont d'intérêt immédiat ou de court-circuiter la séquence d'interaction implémentée par le concepteur. Un espace avec de nombreuses alternatives requiert nécessairement le passage au travers de nombreuses couches de "*Dialog Trees*", car le nombre de choix possibles à chaque étape doit nécessairement être restreint. Du point de vue du concepteur, de tels systèmes sont difficiles à construire car ils impliquent d'être capable de modéliser l'activité en une arborescence de dialogues; leur maintenance est d'autre part compliquée du fait qu'il faut reconstruire entièrement l'arborescence de dialogue en cas d'ajout de nouvelles fonctionnalités. Cependant, même si ce principe de "*Dialog Trees*" peut s'avérer frustrant à utiliser et difficile à faire évoluer, il simplifie l'interaction et minimise l'apprentissage utilisateur. Ces trois approches sont résumées sur la Figure 62.

Speech interface approach	Unconstrained Natural Language	Dialog Trees	Command and Control
User's effort	low	moderate	high (moderate with use)
Developer's effort	very high	moderate	moderate
User training	none	none	required for each application
Supports discovery ?	no	yes, inefficiently	possibly
Scales to complex tasks ?	unknown	no	no
Scales to hundreds of applications ?	yes, but effort not amortized	yes	no
Stress on speech recognizer & parser	high	low	moderate

Figure 62. Comparaison des trois approches possibles pour une exploitation de la reconnaissance vocale en entrée.

Comme déjà signalé, (cf. § 2.2.4.), une autre approche existe sous l'appellation de "*Declarative Modelling*" (Coyne & Sproat, 2001), dont un exemple illustrera le fonctionnement. L'utilisateur produit des énoncés (au clavier dans une fenêtre de saisie), descriptifs de la scène à composer, cf. exemple ci-après, Figure 63, pour obtenir une composition résultante de la scène parmi plusieurs interprétations possibles rendues par le système.



Figure 63. Exemple illustrant l'approche du "Declarative Modelling".

Énoncé entré par l'utilisateur au clavier:
"John uses the crossbow. He rides the horse by the store.
The store is under the large willow.
The small allosaurus is in front of the horse.
The dinosaur faces John. A gigantic teacup is in front of the store.
The dinosaur is in front of the horse.
The gigantic mushroom is in the teacup.
The castle is to the right of the store."

Enfin, d'autres travaux visent à établir une communication Homme-Agent de type directif à partir d'actes directifs, un bras-robot simulé interprétera ainsi des directives utilisateur concernant la tâche en cours. Le pilotage d'un tel système est assuré par une architecture hybride gérant des comportements sur la base d'un système délibératif, (Strippgen & Peters, 1999).

Cette rapide évocation des quelques méthodes et directions de recherche permet de constater l'existence de nombreuses possibilités de recherche et de mise en œuvre qui restent ouvertes en ce qui concerne une interaction vocale avec le système en entrée.

Chapitre 4. Métaphores d'interaction 3D et implémentation logicielle

Les métaphores liées aux environnements 3D

La majorité des actions considérées comme requises (mais non suffisantes) dont doit disposer un utilisateur au sein d'un environnement graphique de synthèse va ici être évoquée par catégories adressant respectivement, au sein de ces environnements, la "*navigation*" qui permet un déplacement virtuel de l'utilisateur, les "*pointage* et *sélection*" préalablement nécessaires à toute "*manipulation*" dont résultera une modification de position ou d'orientation.

Ces dernières "manipulations" seront ensuite distinguées de la "*création*" de formes libres et des techniques de "*modification*" d'objets graphiques.

4.1.1. La "Navigation"

Patrick *et al.*, comparent, en 2000, les résultats concernant les différences relevées dans l'apprentissage du "*Spatial Knowledge*" par 48 utilisateurs en environnement virtuel selon trois types d'affichage :

1. Head-Mounted Display (HMD)
2. Large Projection Screen
3. Desk-Top Monitor

Et ne trouvent pas de différences significatives entre l'usage du HMD et celui de systèmes de projection pour cette tâche.

Une technique de navigation par modification d'échelle de l'environnement à des fins de "survol", baptisée "*Flying, scaling and grabbing in virtual worlds*" fut exposée en 1992 par Robinett et Holloway (Robinett & Holloway, 1992)

Butterworth *et al.*, évoquent, en 1992, un programme de modélisation de surfaces en 3D qui tire ses techniques de manipulation de modèles à la fois des CAD et de programmes de dessin 2D en les appliquant au "modeling" de façon intuitive. 3DM utilise un head-mounted display (HMD) pour simplifier le problème de manipulation et compréhension d'un modèle 3D. La *navigation* y comporte deux aspects, l'utilisateur peut d'abord se déplacer physiquement de quelque pas au sein de l'application (qu'on imagine délimitée d'une manière ou d'une autre). Comme ce type de navigation ne suffit pas à nombre d'applications, le "*Flying*" est aussi implémenté, lequel "*Flying*" est défini ici comme une "fonction déplaçant l'utilisateur au sein de l'EV dans la direction vers laquelle pointe le curseur". "*Flying*" déplace donc l'aire physiquement accessible matérialisée par le "tapis magique" qui délimite l'aire de fonctionnement du système de tracking.

Sans qu'il s'agisse d'une technique de navigation à part entière, une méthode d'exploration locale appelée "*bird's eye*" (Fukatsu *et al.*, 1998) consiste à disposer

d'une "vue aérienne" exocentrée de l'environnement dans lequel se trouve immergé l'utilisateur. Celui-ci dispose d'une vue immersive locale à l'échelle, ainsi que d'une vue aérienne réduite, sorte de "carte" de l'environnement lui indiquant sa position virtuelle.

SOFI (Simulateur Opérationnel de Formation Individuelle) constitue un exemple intéressant d'implémentation d'une interaction utilisateur distante avec un affichage basé sur un mur de projection dans un environnement pédagogique développé pour la SNCF (et décrit dans sa thèse par Domitile Lourdeaux, 2001). La navigation utilisateur au sein de l'environnement est pilotée par un tapis roulant équipé de deux manches de direction. L'utilisateur dispose par ailleurs d'un gant de données pilotant une main virtuelle distante avec laquelle une manipulation d'objets 3D est possible par contrôle visuel uniquement.

Autre méthode relevant plutôt de la "visualisation", la méthode de "*Head-tracked orbital viewing*", qui fut exposée par Koller, Mine & Hudson (1996), reste une exception et constitue aussi une technique d'exploration exocentrique (objet-centrée, en fait) destinée à être utilisée au sein d'un environnement virtuel immersif affiché sur HMD (donc isolant de l'environnement réel). Cette technique utilise un Head-Tracker pour capturer les position et orientation de la tête de l'utilisateur afin de contrôler un déplacement orbital du point de vue utilisateur ("*viewpoint*") relativement à la position qu'occupe l'objet 3D sélectionné. L'objet 3D est situé au centre d'une sphère virtuelle, et les mouvements de la tête de l'utilisateur contrôlent une révolution du "*viewpoint*" inversement proportionnelle à l'amplitude de ces mouvements à la surface de cette sphère virtuelle, de sorte que l'Objet 3D semble effectuer un déplacement relatif en occupant constamment le centre du champ de vision de l'utilisateur tout en conservant son orientation propre, ce qui a pour résultat de permettre d'examiner l'objet sous tous les angles souhaités en ne bougeant que la tête.

Il s'agit d'avantage d'une technique d'affichage permettant d'explorer l'aspect d'un objet que d'une méthode de sélection ou de véritable "manipulation", on pourrait la considérer comme une méthode de "manipulation" apparente de l'objet par les mouvements de la tête, sachant que l'objet conserve en réalité sa position dans le système de coordonnées de l'environnement virtuel (donc, il n'est en fait pas déplacé). La technique peut éventuellement s'avérer intéressante dans le cas particulier d'une immersion concernant l'affichage d'un unique objet, par l'éventuelle évocation d'un mode métaphorique.

Récemment, Tan, Robertson & Czerwinski (2001), présentèrent une "*task-based taxonomy of navigation techniques for 3D virtual environments*", ainsi qu'une nouvelle technique de Navigation baptisée "*Speed-coupled Flying with Orbiting*". (Reproduite Figure 64).

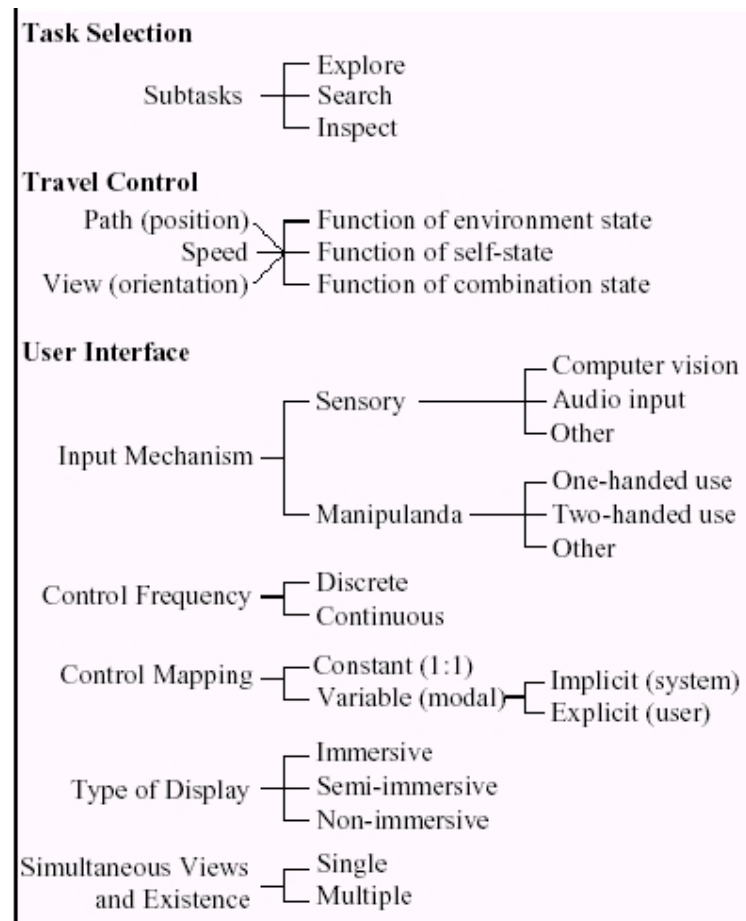


Figure 64. La taxonomie de techniques de navigation de Tan, Robertson & Czerwinski, 2001.



Figure 65. Vue locale de la scène. (Lorsque l'utilisateur se tient immobile ou se déplace lentement).

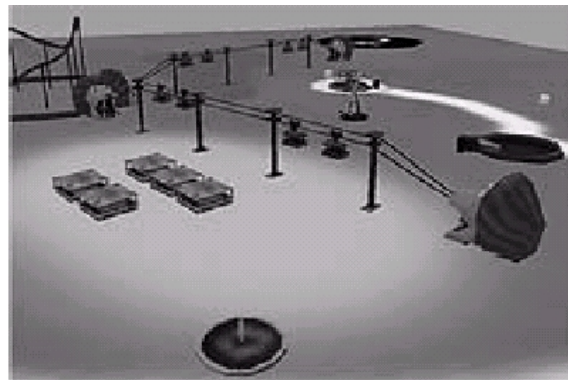


Figure 66. Aperçu de la scène lorsque l'utilisateur se déplace rapidement.

Cette technique fait appel à un dispositif de contrôle tel qu'une souris déplacée par l'utilisateur dans la direction de navigation souhaitée, avec une vitesse de déplacement du dispositif qui détermine la "hauteur" du saut effectué au sein de l'environnement virtuel pour "voler" d'un point à un autre (*cf.* Figure 65-Figure 66).

4.1.2. L'Adressage des Objets 3D ; Pointage 3D et Sélection

Dans un domaine plus particulièrement restreint aux métaphores de "manipulation" applicables en environnement 3D et à leurs techniques que l'on décrira ci-après, Poupyrev *et al.* (1998), présentent une taxonomie des techniques existantes de "manipulation" (de "sélection" d'objets, en fait, reproduite Figure 67) utilisant les seuls mouvements de la ou des mains en tant que technique d'interaction pour la manipulation immersive d'Objets 3D. Les auteurs y distinguent des métaphores elles aussi exocentriques ou centrées utilisateur au sein d'une classification des environnements virtuels tenant compte des techniques disponibles d'interaction utilisateur-environnement.

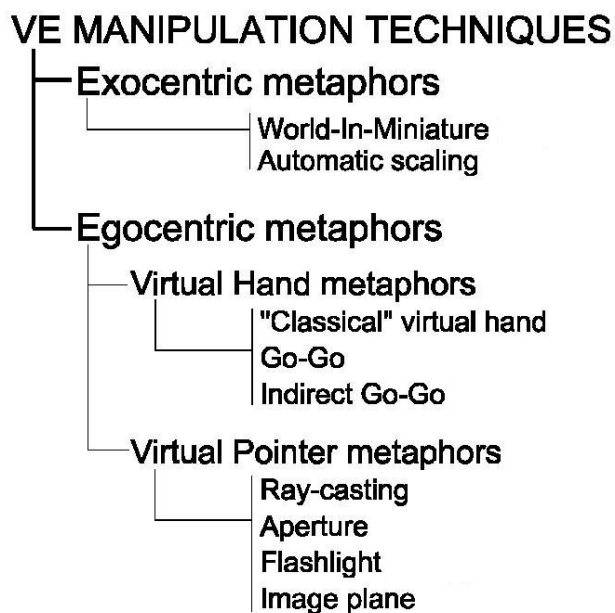


Figure 67 Classification des techniques de "manipulation" en environnement virtuel.

La plupart de ces techniques (*cf.* Figure 67 ci-dessus) impliquent une forme ou une autre de manipulation "manuelle".

Pour reprendre la classification de Poupyrev *et al.* dans le même ordre, Stoakley *et al.* présentent en 1995 les "World in Miniature" (illustrés Figure 68), dans lesquels les Objets graphiques se présentent à portée de main devant l'utilisateur sous la forme de copies à échelle réduite du reste de l'environnement. La manipulation de la copie y

implique un dispositif d'entrée particulier baptisé "props" qui, tenu en main par l'utilisateur, permet d'en modifier diverses caractéristiques.

Sur la Figure 68, en haut à gauche, la miniature représentant le décor d'arrière plan en échelle plus réduite. En haut à droite, le "Props" tenu en main par l'utilisateur pour manipuler la miniature. En bas, un résultat de modification d'inclinaison de la miniature.

Le "Automatic scaling" (Mine *et al.*, 1997) est une technique visant à rapporter les Objets graphiques à portée de main de telle sorte qu'ils deviennent "manipulables" par l'usage de la proprioception, laquelle consistera en la perception qu'a l'utilisateur des positions et orientations de son corps ainsi que de ses membres (*cf.* exemple Figure 69).

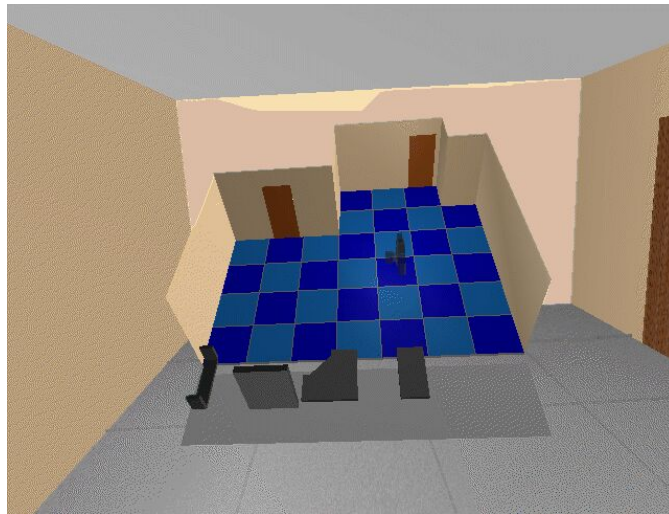
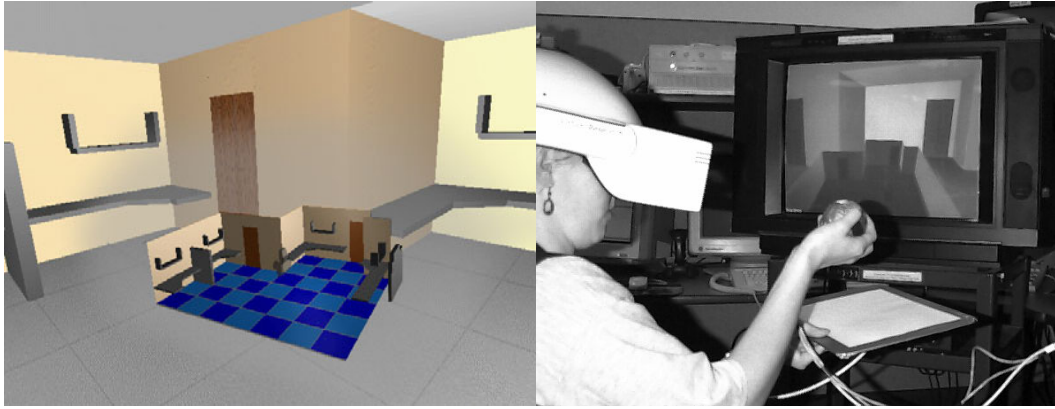


Figure 68. Exemple de "WIM".

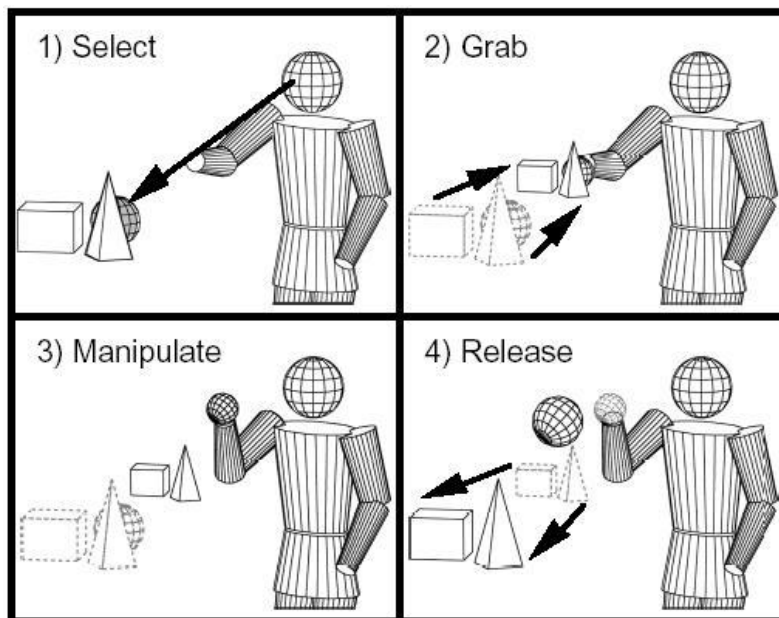


Figure 69. Exemple impliquant un "Automatic Scaling".

Le "*Physical intersection*" (action de "toucher" un objet 3D avec une main virtuelle ou un pointeur de forme quelconque piloté par la main de l'utilisateur (cf. Figure 70, Poupyrev *et al.*, 1998), et le "*laser pointing*" figurent parmi les premières techniques

qui furent utilisées par les chercheurs pour la sélection en environnements virtuels (Hinckley *et al.*, 1994).

Avec la "virtual hand", (*One-to-One Virtual Hand, Go-Go, Indirect Go-Go*), l'utilisateur peut attraper et repositionner un objet graphique 3D à l'aide d'une représentation virtuelle de sa main.

Le principal facteur qui distingue ces techniques de "Virtual Hands" concerne la correspondance de position et d'orientation de la main réelle de l'utilisateur avec la main virtuelle représentée.

Une "Classical Virtual Hand" semblera (autant que possible) effectuer les mêmes mouvements que la main réelle de l'utilisateur, alors qu'avec la méthode dite "Go-Go" (*cf. Figure 71, Poupyrev et al.*, 1998), au-delà d'un seuil prédéfini d'extension de la main réelle de l'utilisateur, la main virtuelle semblera s'éloigner proportionnellement davantage. La méthode de "Physical intersection" présentant l'inconvénient de ne permettre que la sélection d'objets à portée de main, une main virtuelle terminant un bras "extensible" constitue une première solution envisageable.

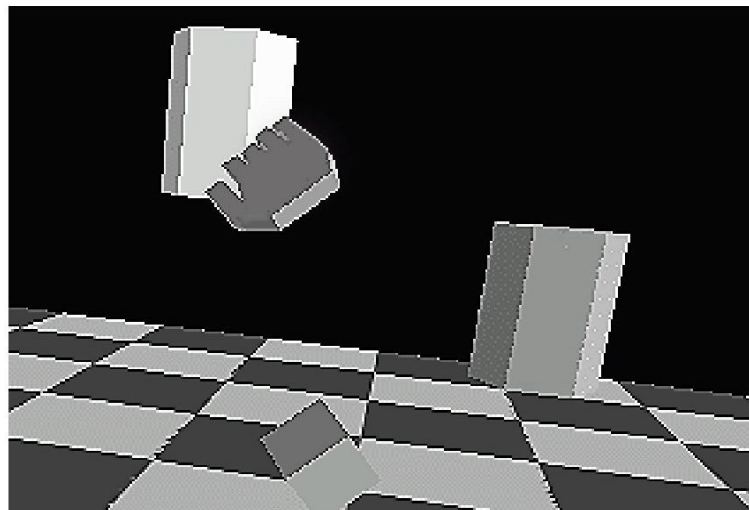


Figure 70 "Go-Go" technique illustrant le "physical intersection".

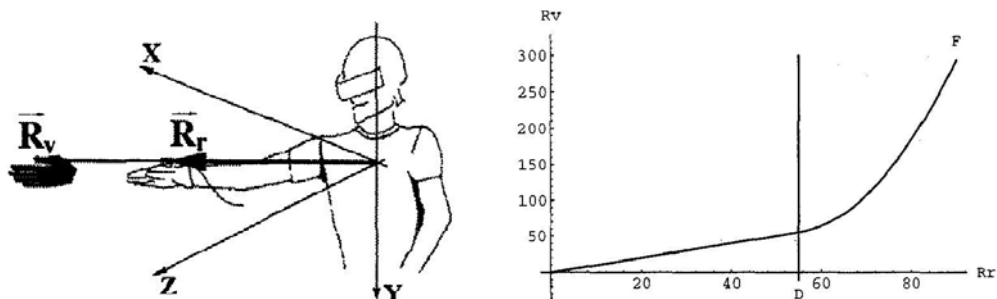


Figure 71 Illustration schématique de la technique dite "Go-Go" (à gauche). A droite, est représentée la fonction non linéaire d'extension de la main virtuelle (RV en ordonnée) par rapport à celle de la main réelle de l'utilisateur (Rr en abscisse).

Une seconde solution possible consiste par exemple à considérer trois régions concentriques auxquelles peut accéder l'utilisateur avec sa main, sachant que celle-ci se trouve généralement dans la région intermédiaire. Si l'utilisateur étend la main vers une certaine direction dans la zone concentrique la plus extérieure (par exemple en étendant complètement le bras), la main virtuelle s'éloignera à vitesse constante dans la même direction que la main réelle et au-delà de celle-ci jusqu'à ce que l'utilisateur ramène sa main réelle à l'intérieur de la zone intermédiaire. Si l'utilisateur rapproche pour finir sa main en sorte de l'amener à l'intérieur de la première zone concentrique la plus proche de lui (en repliant par exemple complètement le bras), la main virtuelle se rapprochera cette fois à vitesse constante.

Dans une étude comparative sur ces techniques permettant de saisir et manipuler des Objets graphiques 3D, Bowman & Hodges (1997b) considèrent que, d'une manière générale, les techniques d'extension du bras rendent la manipulation simple car l'utilisateur déplace et tourne les objets avec des gestes qui lui sont naturels mais que, cependant, se saisir des objets ("*grabbing*") est souvent moins aisé du fait qu'il faille positionner la main virtuelle en contact ou partiellement à l'intérieur de l'objet, du fait de sa taille qui peut être inférieure à celle de la main virtuelle et de la distance qui peut augmenter l'imprécision.

Avec la métaphore du *Pointeur Virtuel* (*Ray-Casting*, *Aperture Based Selection*, *Flash-Light*, *Image-Plane*), l'utilisateur sélectionne et "manipule" des objets en pointant dans leur direction. Lorsque le vecteur émanant du pointeur virtuel rencontre un objet, celui-ci peut être sélectionné et manipulé. Les principaux aspects de conception qui distinguent les techniques basées sur cette métaphore sont la direction du pointeur virtuel, la forme du pointeur (volume de sélection), et les méthodes de spécification de l'objet que l'utilisateur veut sélectionner. Dans le cas le plus simple, la direction du pointeur virtuel est définie par l'orientation de la main virtuelle (cf. Figure 72), et le pointeur se manifeste sous l'apparence d'un "rayon laser" qui traverserait une atmosphère chargée de colloïde ou de poussières.

Certaines techniques définissent la direction du pointeur virtuel en relevant deux positions, celle de l'œil dominant de l'utilisateur ainsi que celle du Tracker qu'il manipule.

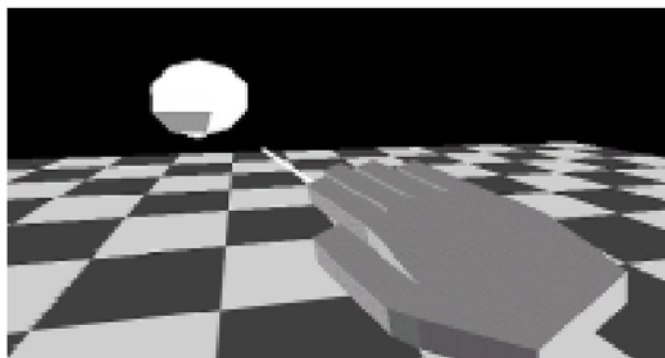


Figure 72 Principe de la méthode de "Ray-Casting".

Le "Laser pointing" contourne donc en partie le problème du rayon d'action de l'utilisateur, mais devient peu efficace pour sélectionner des petits objets à grande distance du fait que de faibles mouvements incontrôlés de la main se traduisent par un balayage angulaire apportant une erreur proportionnelle à la distance (sans compter le bruit ajoutant de l'imprécision au sein des systèmes de capture). Selon Forsberg *et al.*1996, le tracker magnétique appliqué à la méthode de sélection par "Laser pointing" implique un bruit de +/- 5° lorsque l'utilisateur cherche à sélectionner un objet.

Par ailleurs, dans leur étude comparative déjà citée, Bowman & Hodges (1997b) considèrent que les méthodes de "Ray-casting" ne permettent pas facilement d'imposer une rotation sur place de l'objet sélectionné, un seul degré de rotation peut être contrôlé de façon indépendante autour de l'axe du rayon lui-même. Tous les autres degrés de liberté impliquent alors des métaphores supplémentaires qui en compliquent l'usage.

Pour ce qui est des techniques de "Virtual-Hand" et de "Virtual Pointer" évoquées jusqu'à présent, Bowman & Hodges (1997b) rapportent dans leur étude comparative que leur plus grande découverte fut que le "grabbing" et la manipulation doivent être considérés séparément pour une Utilisabilité d'ensemble.

Bien que seulement 20% des 56 utilisateurs préférèrent la méthode de "Ray-casting" aux autres techniques, presque tous trouvèrent qu'il est plus facile d'attraper un objet avec le "Ray-casting" qu'avec n'importe quelle technique d'extension du bras. Cette opération n'implique aucune extension du bras de l'utilisateur qui pointe simplement un rayon et relâche un bouton. Avec les techniques d'extension du bras, on doit placer la main virtuelle à l'intérieur de l'objet, ce qui peut s'avérer délicat pour des objets petits ou trop distants. D'un autre côté, aucun utilisateur ne préféra quelque technique de Ray-Casting pour une manipulation d'objet, au moins du fait qu'une rotation arbitraire sur les trois axes de coordonnées est quasiment impossible. Avec les techniques d'extension du bras, les objets peuvent se voir imprimer une rotation arbitraire dans leur propre système de coordonnées, et leur position peut aussi être contrôlée. Il en résulte qu'aucune des techniques envisagées ne peut paraître universelle.

Technique	Characteristic	Nb Subjects
Go-go	finite range	7
Go-go	imprecise grabbing	8
Ray-casting	difficult rotations	11
Ray-casting	can't move objects in/out	11
Ray-casting	ease of grabbing	10
Arm-Extension	ease of manipulation	9

Le tableau ci-dessus rapporte, sur un total de 11 utilisateurs, le nombre d'entre eux qui fournirent le commentaire qualitatif correspondant pour chacune des techniques testées (Bowman & Hodges, 1997b).

Pour résoudre le problème d'imprécision du pointage virtuel par "*Ray-casting*", la technique du "*Spotlight*" apparaît comme une variante du "*Laser pointing*" (elle ne figure cependant pas, ainsi que la suivante, dans la classification établie par Poupyrev *et al.* en 1998). Cette technique consiste en un volume de sélection conique permettant de sélectionner tout objet compris dans le volume conique défini comme celui d'une torche électrique. Elle nécessite cependant une méthode supplémentaire afin de permettre au système de distinguer quel objet l'utilisateur veut sélectionner lorsque plusieurs objets se trouvent dans le volume de sélection.

Le "*Silk Cursor*" de Zhai *et al.*, (1994), apparaît quant à lui comme une variante de la méthode du "*Physical intersection*", ou action de "toucher" un objet 3D avec une main virtuelle ou un pointeur de forme quelconque piloté par la main de l'utilisateur (*cf.* Figure 73). Il permet de sélectionner des objets devenant contenus au sein d'un volume cubique semi-transparent (le "*cursor*") positionné par l'utilisateur. Cette

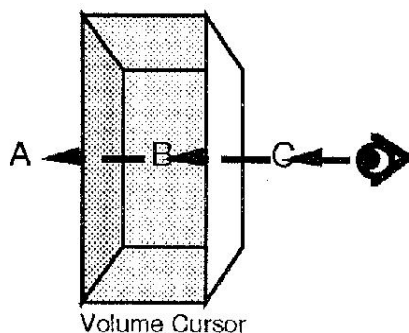


Figure 73. Principe du "Silk Cursor".

méthode autorise une plus grande marge d'erreur de la part de celui-ci lors d'un positionnement du volume de sélection, et le curseur semi-transparent informe visuellement quant à la profondeur de visée, l'objet à sélectionner se trouvant derrière, devant, ou à l'intérieur du volume cubique de sélection. La sélection de la taille du volume est parfois relatée comme un problème, car la sélection d'un unique objet entouré de voisins au sein du même volume pose le même problème que pour la méthode du "*spotlight*".

La méthode de sélection par "*Aperture*" (ouverture), dite encore "*Aperture Based Selection*" développée par Forsberg *et al.* en 1996 (*cf.* Figure 74) est une modification de la technique de sélection par "*spotlight*". Avec cette technique, l'apex du volume conique de sélection (le point de départ) est situé au niveau de l'oeil dominant de l'utilisateur, et le vecteur-direction du cône est celui qui relie l'oeil dominant à la position manipulée d'un tracker magnétique à six degrés de liberté. La position du tracker est affichée sous la forme d'un curseur 3D par exemple circulaire. L'ouverture du curseur (ou "*aperture*") est montrée sous la forme d'un cercle de rayon fixe et d'un réticule ("*CrossHair*") aligné sur le plan normal à la direction du cône de sélection. La taille du volume de sélection (ou l'angle solide du cône) est déterminée par la distance séparant l'apex du cône (l'oeil dominant de l'utilisateur) de l'ouverture du curseur dont le diamètre est constant, mais que l'utilisateur peut rapprocher ou éloigner du

"viewpoint" (l'apex), ce qui aura respectivement pour effet d'augmenter ou de resserrer le volume de sélection (l'angle solide du cône).

L'utilisateur peut donc ainsi ajuster l'étendue du volume de sélection en éloignant ou rapprochant le curseur de son oeil.

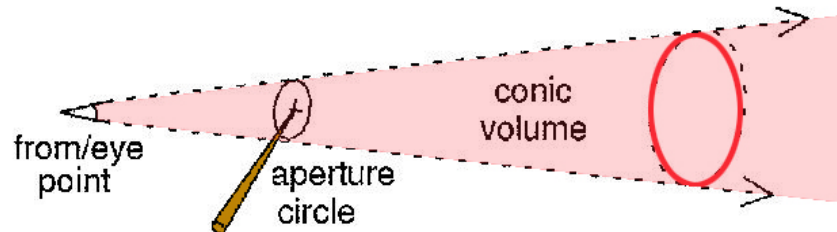


Figure 74. Principe de la technique "Aperture Based Selection".

Comme déjà mentionné au sujet de ce qui fait de la direction du regard une gestuelle particulière, la technique "*Image Plane*" développée par Pierce *et al.* en 1997 (cf. Figure 75) semble fonctionner selon le même principe que la méthode "*Aperture Based Selection*" en remplaçant le dispositif de pointage matériel par un dispositif non seulement de tracking de la main de l'utilisateur, mais aussi de traitement des mouvements des doigts de la même main dont un geste particulier initierait le processus de sélection au sein d'un volume conique dont l'angle solide varie au gré de l'écartement du pouce et de l'index ainsi que de la distance de la main à l'oeil dominant (le "*viewpoint*"). La différence technique réside en le fait que l'interaction s'effectue avec la projection de l'image de l'objet graphique sur le plan d'affichage.

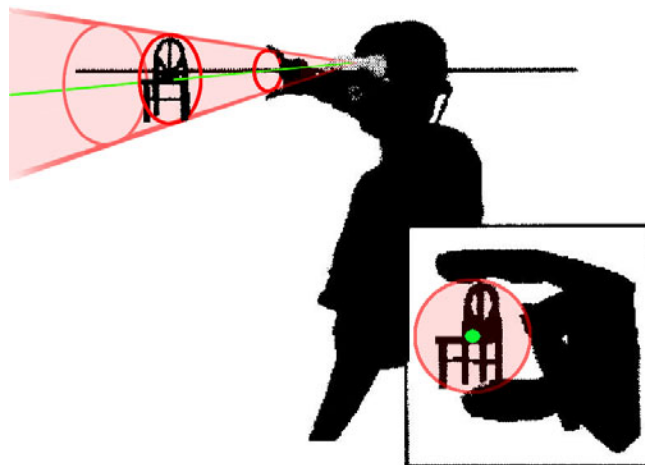


Figure 75. Principe de "l'image plane interaction".

4.1.3. La "Manipulation" d'objets 3D

Par "Manipulation d'Objet", on semble souvent sous-entendre (au fil des lectures) les seuls déplacement et orientation de l'objet à quelques rares exceptions telles que

l'édition/modification de courbes directement sur un modèle 3D "fil de fer" par modification directe ou camouflée des positions des points de contrôle.

On trouvera sans doute alors utile de distinguer non seulement les opérations de *pointage* et de *sélection d'objet* (évoquées précédemment) des "*manipulations*" proprement dites en tant que modification des seules position et orientation d'un objet, mais aussi des *modifications*, ou *éditions*, d'une ou plusieurs propriétés internes de l'objet graphique.

Butterworth *et al.* implémentent, en 1992, au sein de leur programme de modélisation de surfaces en 3D, la fonction "*Grabbing the world*" qui s'y définit par la "possibilité qu'a l'utilisateur d'attacher l'espace de travail au curseur de façon à le déplacer et le faire tourner selon les mouvements de la main". Les fonctions complémentaires de modification d'échelle baptisées ici "*Shrinking*" et "*Growing*" n'affectent pas le modèle 3D, mais plutôt la taille relative de l'utilisateur au sein de l'environnement virtuel (sans modification, semble-t-il, de l'échelle du rayon d'action). Les fonctions d'affichage telles que les "rayons élastiques", extrusion, etc. sont présentes ainsi que d'autres fonctions communes aux logiciels de CAO lors de la définition de contours d'une nouvelle surface. Le vertex le plus proche du curseur est identifié par "*highlighting*" sur action d'un bouton de la souris.

L'un des aspects intéressants de cet article est la mention des auteurs concernant une liste de trois classes d'outils les plus courants à implémenter, lesquels peuvent s'identifier aux "sous-tâches utilisateur", quoiqu'encore de très bas niveau puisque directement importées des logiciels de CAO 2D. Ces trois classes consistent en

Outils de Création de surfaces,

Outils d'Édition,

Outils de Hiérarchisation.

Bien que la façon dont s'effectue la sélection d'une icône dans la palette toolbox n'est pas détaillée, on notera que les auteurs manipulent ici un pointeur directement importé des logiciels de CAO 2D.

SOFI (Simulateur Opérationnel de Formation Individuelle) constitue un exemple déjà mentionné d'implémentation d'une interaction utilisateur distante avec un affichage basé sur un mur de projection dans un environnement pédagogique. Outre la navigation au sein de l'environnement qui est pilotée par un tapis roulant équipé de deux manches de direction, l'utilisateur dispose d'un gant de données pilotant une main virtuelle distante avec laquelle une manipulation d'objets 3D est possible par retour visuel uniquement. Les problèmes constatés par l'auteur consistent en "des difficultés éprouvées par les utilisateurs pour saisir les objets", ces difficultés étant dues selon l'auteur au problème d'estimation des distances et au temps de latence du gant de données.

4.1.4. La "Création" d'Objets 3D

Une hypothétique "création" d'objet se distinguerait de la modification d'un objet importé au sein de l'environnement en tant qu'étape initiale de définition d'une forme libre ("*Free-form*"). Les exemples sont rares dans la littérature. Il existe toutefois une technique présentée par Schkolne *et. al* en 2001 qui présente de telles caractéristiques en permettant à l'utilisateur de définir des surfaces curvilignes au sein de l'environnement virtuel à l'aide de mouvements de la main (*cf.* Figure 76- Figure 77). Le mouvement de la main décrit un "coup de pinceau" dans l'espace lorsque l'utilisateur rapproche le pouce de la paume afin de fermer un contact signalant l'action de dessin dans l'espace de travail au système.

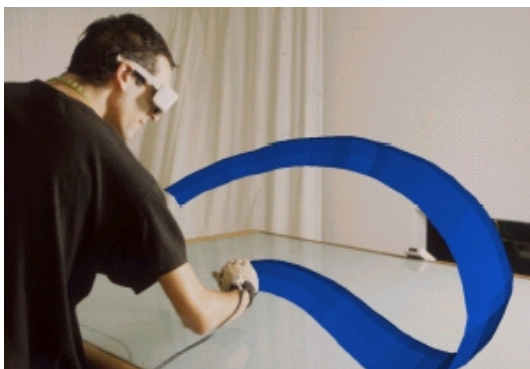


Figure 76. Illustration d'une technique de conception de forme libre.

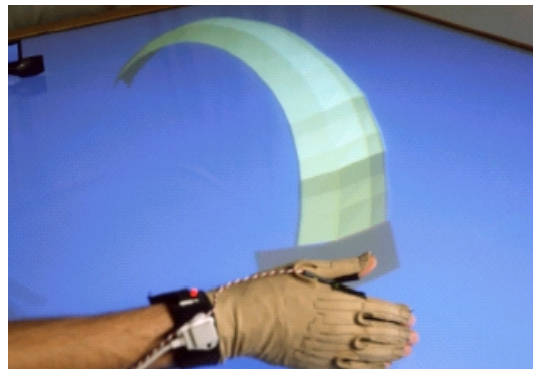


Figure 77. Détail de fonctionnement.

4.1.5. La "Modification" d'Objets 3D

Les "techniques de modification" désignent d'une manière générale les méthodes d'édition qui permettent notamment de modifier des caractéristiques propres aux "Objets 3D" au sein d'un environnement virtuel. En première considération, on entend jusqu'à présent par "Objet 3D" une forme essentiellement graphique, dont, par conséquent, on modifiera surtout la forme apparente, éventuellement des formes internes s'il en existe, ainsi que les caractéristiques de surface telles que texture, couleur, etc.

La notion "d'Objet 3D" n'étant pas particulièrement bien définie dans la littérature, on a toutefois évoqué au cours des chapitres précédents qu'en fonction des modalités installées en sortie et accessibles au système, adressant divers sens de perceptions de l'utilisateur, un "Objet 3D" pouvait se manifester non seulement sous forme graphique, mais aussi sonore, ou encore tactile, pour ne considérer que les principales manifestations qui nous intéressent ici.

Dans cette perspective, une éventuelle liste de toutes les catégories possibles d'édition de caractéristiques diverses relatives aux "Objets 3D" adressant par exemple pour chaque type de modalité en sortie les "comportements" possibles d'un objet

dynamique (mouvements, viscosité, couleurs, réactions à l'environnement, relations aux autres objets, etc. sans évoquer par ailleurs les méthodes classiques de bases de données pouvant adresser d'autres types d'informations relatives à un tel objet) déborderait du cadre de ce document.

Nous nous limiterons donc ci-après à quelques exemples intéressant uniquement la déformation graphique de la forme d'un objet 3D.

Quelques Techniques de Modification d 'Objets 3D

Historiquement, l'une des premières méthodes d'édition de la forme d'un objet 3D consiste en celle de courbes Splines telle qu'évoquée par Krueger en 1993, effectuée dans ce cas en identifiant les extrémités des ombres des doigts aux points de contrôle de la courbe et qui figure une interaction bimanuelle libre de tout instrument, (*cf.* Figure 78).

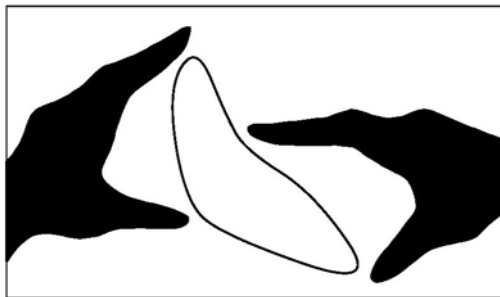


Figure 78. Edition de courbe Splines.

Une technique nettement plus récente présentée par Biermann *et al.*, (2001), semble préfigurer des méthodes simples de modification d'objets graphiques 3D à l'aide d'opérations booléennes telles qu'union, intersection, exclusion de plusieurs objets 3D entre eux. Ce type de méthode semblerait convenir non seulement pour des primitives graphiques typiques du "*solid modelling*", mais aussi pour la modification de formes libres (*cf.* Figure 79).

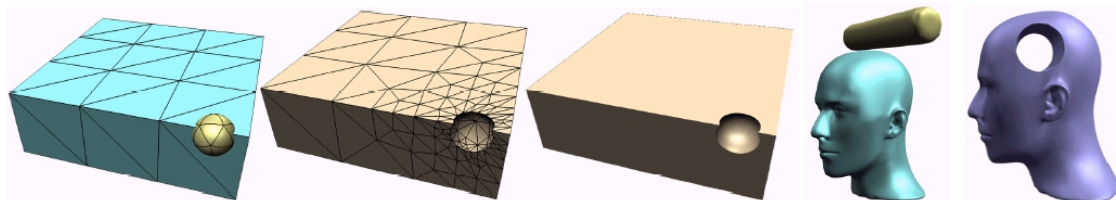


Figure 79. Opérations booléennes d'exclusion d'un volume sur un autre.

Dans une autre catégorie faisant appel cette fois aux "props" que peut manipuler l'utilisateur, Murakami et Nakajima (2000), présentent une technique de déformation d'un objet 3D proportionnelle aux contraintes physiques imposées par l'utilisateur sur un objet réel similaire doté de capteurs physiques (*cf.* Figure 80). Bien que ce type de modification n'adresse dans ce cas qu'un unique objet 3D à la forme ramassée et proche d'une primitive (*cf.* Figure 81 à Figure 84.), toutes sortes de formes de props se rapprochant d'autres formes d'objets 3D sont imaginables.

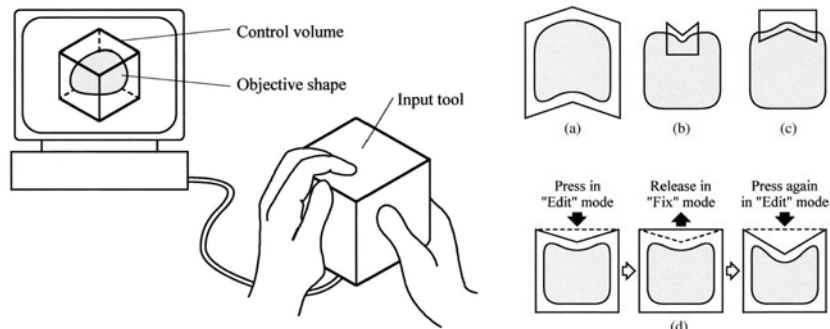


Figure 80 Principe de la déformation physique d'un "props" pour obtenir une déformation proportionnelle d'un Objet 3D.

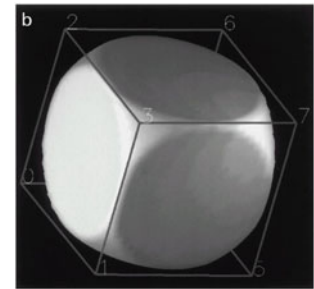


Figure 81. La forme de base devant subir les déformations.

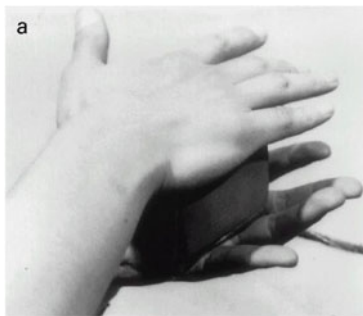


Figure 82 Pressage de deux faces.

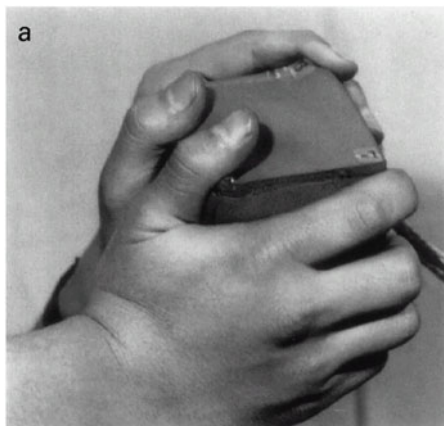
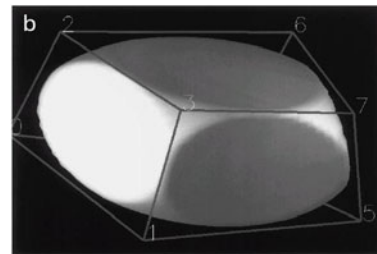


Figure 83 Torsion.

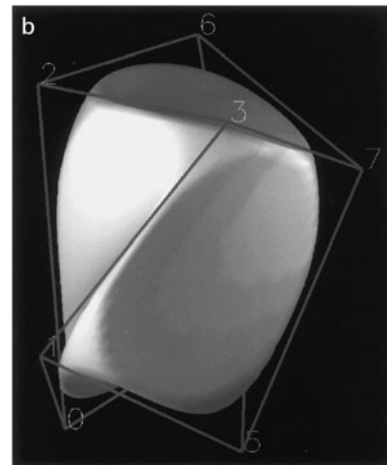
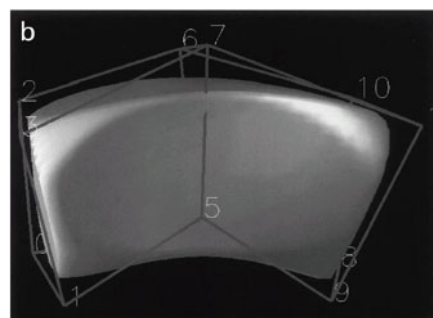


Figure 84 Flexion.



Les déformations résultant des contraintes imposées au "prop".

Pour citer maintenant une technique de déformation de maillage, R. Szeliski and D. Tonnesen (1991) proposent une modélisation de surface basée sur des particules orientées selon des lois de la physique Newtonienne. Ce modèle permet différentes opérations dites "*Free Form Modelling*" de la part de l'utilisateur telles que la découpe d'un "*mesh*" avec un objet contondant virtuel comme on le ferait d'un tissu dans la réalité avec une lame, (cf. Figure 85), ou encore la déformation dynamique d'un réseau maillé au gré des manipulations d'un outil de déformation virtuel par l'utilisateur (cf. Figure 86-Figure 87, pour plus de détail, voir la thèse de D. Tonnesen, 1998). Cette dernière fonctionnalité fut portée sur le WorkBench de l'INRIA à Rocquencourt (Figure 87), et présentée en fin d'année 2000 au cours d'un séminaire sur les "environnements virtuels à base de projections sur grands écrans", (Séminaire INRIA-Rocquencourt, 2000, GEOCLAY Demo).

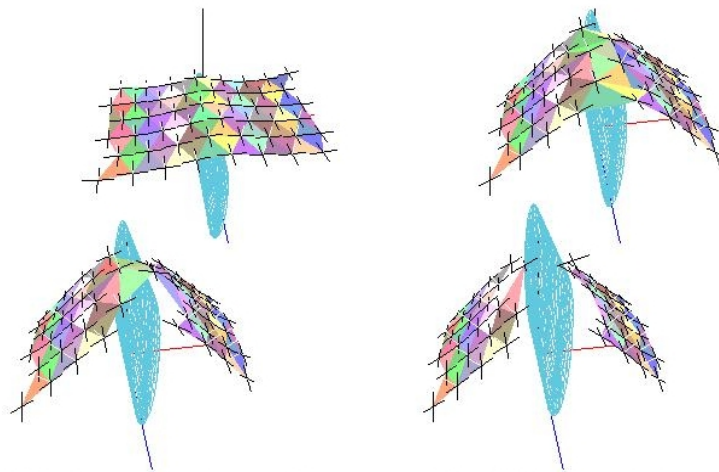


Figure 85 Exemple de "découpe" d'une surface par rupture du réseau de particules.

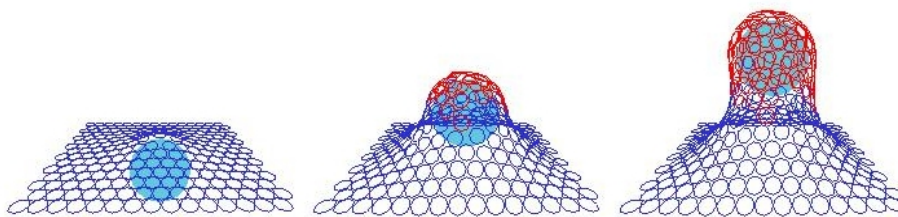


Figure 86 Principe de déformation du réseau de particules dont résulte une technique de "Free-Form Modelling".



Figure 87. Illustration de ce que permet cette méthode de déformation sur le WorkBench.

Les quelques techniques de déformation évoquées ci-dessus se limitent aux formes graphiques d'un objet 3D. Elles font appel à des dispositifs d'entrée manipulables permettant de sélectionner, déplacer, contraindre des objets graphiques 3D.

De nombreuses autres techniques existent, qui font par exemple appel au retour d'effort mécanique permettant de contraindre un objet 3D "visqueux" au point d'être déformable. Ces techniques de déformation intéressent en fait non seulement les opérations disponibles au sein de l'application, mais aussi les modalités disponibles en entrée qui seront accessibles à l'utilisateur.

4.2 Conclusion

Dans le contexte d'activités de conception (initialement dans le domaine de la confection vestimentaire), la revue de question qui fait l'objet de cette première partie a recensé, décrit et organisé les travaux récents en matière de modalités et de dispositifs d'interaction avec des objets graphiques 3D.

Nous avons, dans le cadre de cette revue de question, examiné la littérature disponible dans des domaines tels que l'IHM, l'ergonomie, l'infographie, les environnements virtuels, la multimodalité, etc., afin de situer les problèmes les plus importants qui limitent encore l'usage de ces techniques. Nous avons par ailleurs tenté, ce faisant, de définir les relations envisageables entre l'utilisateur et le système sur la base des expressions que celui-ci peut mettre en oeuvre. Cette double progression nous a permis de définir les caractéristiques générales d'une interaction orientée communication utilisateur avec le système.

La description et l'exploration du domaine nous ont tout d'abord amené à clarifier certaines notions qui nous semblent essentielles, et à retenir, au long du document, un vocabulaire de travail. D'un point de vue centré sur l'utilisateur, et à des fins de simplification, les "environnements virtuels" nous ont paru pouvoir, sur la base de taxonomies existantes, être classés en termes de capacité croissante d'action permise à l'utilisateur.

Nous avons ensuite considéré les modalités d'entrée-sortie du système et leurs contraintes en commençant par la présentation visuelle. Divers dispositifs d'affichage sont évoqués avec les contraintes qu'ils apportent à leur mise en œuvre.

D'autres modalités en sortie sont ensuite évoquées plus brièvement. Parmi celles-ci, le constat de l'immaturité des techniques adressant le sens tactilo-kinesthésique suggère encore provisoirement une orientation vers un style d'interaction utilisateur n'impliquant pas de simulation d'un contact physique avec un objet graphique 3D.

Ce premier constat nous permet de limiter l'étendue du domaine concerné par les modalités d'entrée à celles convoyant plus particulièrement des informations de nature communicative. Les deux catégories de modalités retenues en entrée concernent ainsi le vocal et diverses gestuelles rapportées à des mouvements observables sur l'utilisateur.

Il nous semble qu'à l'avenir, le domaine devrait porter davantage d'efforts sur l'expérimentation ainsi que les tests utilisateur en ce qui concerne notamment des interactions diversifiant les modalités en entrée. En effet, les nombreuses directions de recherche évoquées ne rapportent encore que trop rarement des résultats de tests utilisateurs pour le domaine de l'interaction avec des environnements 3D.

Pour notre part, nous envisageons de tester certaines configurations pour lesquelles on a retenu la reconnaissance vocale, la direction du regard ou celle de la tête, ainsi que certains mouvements de la main. Dans cette perspective nous allons approfondir l'étude des combinaisons de modalités en regard des diverses tâches utilisateurs possibles, notamment dans le domaine de la conception.

Partie II

Etude expérimentale

Chapitre 5. Manipulation dans un environnement virtuel 3D : Tâches et commandes élémentaires

A ce stade de ce travail à la fin de l'année 2002, le partenaire industriel fournisseur des terrains d'étude dans le domaine du stylisme-modélisme s'est désengagé du projet. L'accès attendu de sa part à une application informatisée relative à ce domaine de conception ainsi qu'à une population de stylistes-modélistes n'est alors plus d'actualité. Afin de poursuivre néanmoins ce travail en vue d'aboutir à des résultats utiles, l'objet de l'étude est resitué dans le cadre d'une interaction utilisateur plus générale à expérimenter avec une application informatique simplifiée, réalisée par nos soins. Pour ce faire, nous conservons le modèle d'interactions entre les acteurs de la conception vestimentaire, résultant des observations de terrain. Toutefois, parmi les conséquences de ce retrait du partenaire industriel, nous retiendrons la fin des relations avec les acteurs du domaine qui ne permet plus d'expérimentation possible avec cette population de concepteurs. Nous verrons plus loin au sujet de l'expérimentation effectuée que ceci a contribué à considérer la participation d'utilisateurs ne pratiquant pas une activité de conception particulière.

Nous avons à traiter les relations existant entre les tâches à effectuer, les modalités d'entrée permettant de les mettre en oeuvre, et les dispositifs qui devront permettre l'emploi de ces modalités.

Les tâches que doit effectuer l'utilisateur concepteur confronté à un environnement permettant d'agir sur les objets graphiques 3D sont décomposées en sorte d'obtenir une arborescence dont le plus bas niveau est occupé par les tâches élémentaires réalisables au sein de l'environnement de présentation. Les tâches que l'utilisateur doit pouvoir mener à bien ne s'arrêtent pas aux seules opérations réalisables sur les objets graphiques présentés par le système. Celui-ci présente en effet par exemple non seulement les objets graphiques "objets" de la tâche (de pointage, de sélection, de déplacement, etc.), mais également un environnement de présentation au sein duquel l'utilisateur devra pouvoir éventuellement mener à bien d'autres tâches que celles concernant uniquement un objet graphique (relations entre objets, stockage, requêtes, caractéristiques d'attribut des objets telles que textures, etc.).

A titre d'exemple, l'une des tâches utilisateur consistera à "sélectionner" un objet graphique devant subir quelque opération, cette "sélection" se décompose elle-même en un "pointage" qui se trouve "validé" d'une manière ou d'une autre. Un tel "pointage" est une opération élémentaire préalable que doit pouvoir effectuer l'utilisateur afin de "sélectionner" effectivement un objet graphique 3D. Par ailleurs, l'utilisateur devra pouvoir effectuer d'autres tâches telles que recherches d'information, exploration, opérations système, etc.

5.1 Classification de tâches

Selon Bowman (2002), parmi la diversité des tâches utilisateurs possibles, il existe un ensemble de tâches universelles pour les interfaces 3D, qui se définissent comme des tâches simples que l'on retrouvera dans la plupart des applications, et qui peuvent être combinées afin de réaliser des tâches plus complexes.

Ces tâches utilisateur « universelles » pour les environnements virtuels ont motivé diverses taxonomies distinguant des sous-ensembles de tâches tels que « Navigation », « Sélection » et « Manipulation ».

Une étude générale réalisée par Gabbard et Hix en 1997 offre une telle taxonomie de « tâches utilisateurs » dérivée d'une classification antérieure présentée par Esposito en 1996, dans laquelle ce dernier aurait identifié cinq types de tâches,

« *Navigation* »,

« *Object Query* »,

« *Object Manipulation* »,

« *Object Creation and Modification* »,

« *Application Environment Query and Modification* ».

Considérant la création et modification d'objet comme une forme spécialisée de la manipulation d'objets, Gabbard et Hix simplifient leur taxonomie à trois principales catégories d'activité,

« *Navigation/Locomotion* »,

« *Sélection d'objets* » et

« *Manipulation/Modification/Exploration d'objet* », (Gabbard et Hix, 1997, p.46).

Depuis lors, d'autres études furent réalisées sur des sous-ensembles de telles taxonomies, par exemple au sujet des techniques de « Locomotion » (Arns et Cruz-Neira, 2000), des techniques de « Manipulation » (Poupyrev et al., 1998), etc.

L'importante place descriptive que prennent les divers dispositifs d'entrée-sortie dans ces classifications sous-tend l'idée générale qu'un système d'interaction 3D implique des dispositifs d'entrée à manipuler mécaniquement par l'utilisateur de façon à obtenir un effet sur un objet de l'environnement virtuel.

La plupart de ces « tâches universelles » se définissent alors par rapport à la nature de l'équipement du système en termes de dispositifs d'entrées-sorties qui réduisent la communication aux « techniques d'interaction » qui les mettent en oeuvre entre l'utilisateur et un système interactif 3D.

En considérant un contexte de communication interpersonnelle dans lequel l'un des participants exprime une commande à exécuter par un tiers sur un objet plutôt que de chercher à réaliser lui-même l'opération, nous constatons que la plupart des classifications disponibles ne s'appliquent que partiellement à notre problème. Par

exemple, la « Sélection » est souvent décrite sur la base d'un retour visuel de quelque « pointeur » ponctuel 3D pouvant se présenter sous diverses formes pour lesquelles nous n'admettons qu'une efficacité toute relative en ce qui concerne notamment l'appréciation de sa distance de la part de l'utilisateur.

Pour un environnement d'interaction orienté « communication », il est concevable qu'il n'y ait aucun retour visuel d'un tel pointeur 3D qui serait précisément positionné au sein d'un système de coordonnées 3D sur une image plate représentant l'environnement, à l'aide d'un dispositif d'affichage incomplètement adapté à la vision humaine.

Dans une première approche pour considérer une interaction orientée « communication » avec un système 3D, et en particulier pour les tâches utilisateur relatives aux objets au sein de la scène 3D, nous distinguerons d'abord trois catégories de tâches se situant à des niveaux différents :

les « pointage, visée et présélection »,

la « Sélection »,

les « Mouvements d'objets ».

5.1.1. Pointage, Visée et Présélection au sein d'un Environnement Virtuel.

Il s'agit ici de l'indication de cibles par l'utilisateur. L'importation de la notion de pointeur largement développée pour cet usage avec les systèmes WIMP (windows-Icon-Menu-Pointing), apporte une solution permettant entre autre un pointage précis (et "ponctuel", par conséquent) qui peut trouver de nombreux usages en environnement 3D mais qui s'avère cependant délicat et laborieux à utiliser de manière générale du fait de la difficulté pour l'utilisateur d'en apprécier la distance de façon précise.

En rapport avec ce qui a été mentionné en introduction au sujet des classifications de tâches, la notion de « pointage » sous-entend bien souvent celle de quelque dispositif mis en œuvre de façon discontinue par l'utilisateur (à l'image de ce que l'on obtient avec une souris sur un plan). Ainsi, les diverses techniques de « Virtual Hand » ou encore le "Silk Cursor" de Zhai et al. (1994) représentent-elles des extensions au sein de l'environnement virtuel de la notion de pointeur au sens d'une position spatiale virtuelle généralement signalée par quelque retour visuel.

Cette prééminence de la notion de « pointeur » dans les classifications des techniques d'interaction semble ainsi éclipser d'éventuelles techniques alternatives pour lesquelles une visée, ou plus généralement, l'indication d'une direction, plus atomique pour notre part, résulterait d'un acte involontaire éventuellement continu de l'utilisateur.

Cependant, les exemples de la littérature semblent arrêter la manipulation d'une direction aux seuls actes volontaires que peut produire l'utilisateur. Ainsi, la méthode

de sélection par "Aperture" (ouverture), dite encore "Aperture Based Selection" développée par Forsberg et al. en 1996, ou encore la technique "Image Plane" développée par Pierce et al. en 1997, qui représentent pourtant des exemples de manipulation d'une direction dans laquelle se trouve l'objet d'intérêt, sous-entendent-elles, de la part de l'utilisateur, la manipulation d'un dispositif manuel supplémentaire à celui qui détermine le point d'origine de la visée qui est ainsi réalisée.

Par conséquent, nous distinguerons ici, dans la même catégorie d'opérations dont l'objectif est l'identification d'une cible comme préalable nécessaire à la sélection, la notion préalable de visée qui, plutôt qu'identifier un point dans l'espace, met plutôt en oeuvre l'indication d'une direction prenant son origine à la position de l'observateur et bornée à la distance maximale perceptible (ou représentable) au sein de l'environnement.

Ainsi, la spécification d'un point particulier de l'espace devient-elle possible, de façon centrée utilisateur, d'abord sur la base d'une tâche « universelle » d'indication d'une direction, puis sur l'indication d'une distance additionnelle pour définir une position dans un système de coordonnées polaires.

La notion de « *présélection* » admet donc, de la part de l'utilisateur, un acte éventuellement continu, involontaire mais ajustable, révélant éventuellement un objet « d'intérêt », qualité qui se verra alors confirmée par la manifestation d'un acte volontaire produisant la sélection de l'objet d'intérêt.

Nous concevons deux principales tâches de « *présélection* » que nous qualifierons d'atomiques :

Spécification implicite d'une *Direction*, orientée à partir d'une position courante (celle de l'utilisateur).

Spécification implicite d'un *objet de la scène* dans la direction spécifiée.

De telles "*spécifications implicites*" ne précisent pas le *moyen* utilisé pour convoier cette information au système. Nous préférons pour notre part définir des *fonctions communicatives* qui pourront transiter via diverses *Modalités* correspondant à l'information à transmettre. Nous admettons cependant que ces spécifications implicites peuvent être le résultat permanent d'un acte involontaire mais continu de la part de l'utilisateur.

5.1.2. Sélection au sein de l'Environnement Virtuel.

"Action de faire en sorte que l'Objet préalablement visé, ou présélectionné, devienne l'Objet courant de la tâche".

Ainsi énoncée, la Sélection apparaît, en ce qui concerne l'objet de la tâche, comme la somme de deux conditions,

[Visée] & [Validation].

La validation constitue un acte utilisateur volontaire qui confirme son intérêt pour l'objet de la visée et qui détermine une sélection. En tenant compte de ce qu'on a précédemment admis concernant l'aspect éventuellement continu de la source des spécifications implicites, cette formulation implique a priori une source d'événements ponctuels, du moins de courte durée produisant une interprétation immédiate, mise en œuvre de façon volontaire, éventuellement motrice, pour effectuer une validation.

Différentes façons de "valider" un pointage sont concevables en fonction des modalités d'expression utilisées au travers de quelque dispositif mis en œuvre pour les convoier.

Nous ne devons cependant pas perdre de vue l'éventualité selon laquelle la notion de sélection ne s'applique pas exclusivement à l'objet de la tâche. Certains actes des utilisateurs peuvent en effet présenter une information de commande dont la validation n'est soit pas forcément nécessaire, soit difficilement réalisable sans affecter l'efficacité de l'interaction. Nous parlerons alors de « sélection implicite », par distinction avec la « sélection explicite » qui implique, comme précédemment, une forme de « présélection » produisant un état modifiable de façon contrôlée et maintenue jusqu'à ce qu'une « validation » apparaisse afin de produire une « sélection ».

Par exemple, la sélection d'une direction de la part de l'utilisateur peut se voir effectuée via une gestuelle déictique (Regard, pointage manuel de l'index), elle pourra aussi être le produit d'un énoncé (dactylographique, vocal, sélection d'une valeur discrète).

Pour résumer ces caractéristiques générales distinguant les sources possibles de présélection et de sélection, nous allons les reporter dans le Tableau 1 ci-dessous.

Présélection	implicite	Indication éventuellement continue d'un état modifiable et mesurable, réduisant l'ambiguïté. Éventuellement involontaire.	
	(explicite)	(Réponse du système au cours d'un dialogue de niveau élevé.)	
Sélection	implicite	Information de désignation non ambiguë (présélection), ET Acte volontaire ponctuel.	
	explicite	Toute combinaison de deux modalités d'entrée	Présélection implicite OU Présélection explicite ET Acte volontaire ponctuel.

Tableau 1. Caractéristiques générales distinguant les sources de présélection et de sélection.

Par présélection « explicite » reportée en deuxième ligne dans ce tableau, mais qui ne sera plus abordée par la suite, nous entendons, à un niveau de dialogue plus élevé avec le système, une interprétation de l'expression incomplète de l'utilisateur pour exécuter une opération de « déplacement », autorisant toutefois la production d'un retour visuel à l'indication d'une direction dite « exocentrée » (au chapitre suivant), permettant à l'utilisateur un ajustement visuel de cette indication de direction.

Nous pouvons maintenant, sur la base de ce que permet la présélection, en conclure au moins deux premières catégories de « Sélections » :

Spécification explicite d'un point destination ou position de la part de l'utilisateur, à partir de l'indication explicite (sélection) d'une direction, puis d'une distance, le tout constituant une position en coordonnées polaires par rapport à la position virtuellement occupée par l'utilisateur.

Spécification explicite d'un ensemble de points au sein de l'environnement, il pourra s'agir d'un Objet Graphique 3D, situé dans la direction implicitement indiquée par l'utilisateur, et d'un acte volontaire confirmant la sélection de l'objet 3D.

D'autres sélections explicites sont possibles, et la notion de « sélection implicite » laisse présager des actes volontaires désignant sans ambiguïté tout ou partie d'une commande, que nous réexaminerons lorsque nous aborderons plus loin les modalités et dispositifs d'entrée-sortie.

5.1.3. Mouvements d'objets 3D au sein de l'environnement.

Un « déplacement » peut être défini par la modification des position et orientation d'un Objet quelconque sur 6 degrés de liberté (translation et orientation). Chaque groupe de degrés de liberté devant pouvoir être spécifié pour le déplacement (translation et/ou orientation, voire éventuellement une composition de ceux-ci avec la position occupée par l'utilisateur, ex. le parcours par l'objet 3D sélectionné d'une trajectoire orbitale autour de l'utilisateur avec ou sans révolution de l'objet 3D).

Afin de réduire l'étendue de l'étude, nous nous limiterons ici aux résultats que l'on peut obtenir avec des déplacements finis, les mouvements continus représentant une autre catégorie de déplacements qui adressent la notion de trajectoire et concernent davantage les environnements dynamiques.

Le terme de « déplacement » caractérisant davantage un changement de position, nous retiendrons la catégorie générale des « mouvements » d'objets 3D pour y distinguer les changements de position des mouvements giratoires.

En ce qui concerne les changements de position, nous constaterons rapidement que la notion générale de « déplacement » peut prêter à confusion lorsque nous distinguerons par la suite la spécification d'un « placement » d'un objet à une position particulière référencée par rapport à celle de l'utilisateur, de celle d'un « déplacement » de l'objet par rapport à sa propre position initiale.

Nous préférons alors, pour notre part, en conservant la définition du dictionnaire « petit Larousse »⁷, réunir sous le terme de « translations », les deux notions distinctes de « positionnement » et de « déplacement ».

La notion de « positionnement » prenant pour origine la position de l'utilisateur, à laquelle ce dernier ajoute l'indication d'une direction et d'une distance pour définir une position destinataire sans rapport avec la position initiale de l'objet.

La notion de « déplacement » conservant quant à elle la position initiale de l'objet de la tâche comme origine, à laquelle l'utilisateur ajoute l'indication d'une direction et d'une distance pour définir une position de destination.

La première opération évidente de mouvement que l'on peut établir sur la base de ce que nous venons de mentionner est celle qui consiste par exemple pour l'utilisateur à exprimer un « placement » d'un objet 3D à une position particulière de l'espace.

[Sélection (Opération de positionnement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (Valeur)]
& [Sélection (Unité)]
& [Sélection (Direction)]

On en déduit immédiatement un second exemple qui consistera cette fois à « déplacer » un objet 3D d'une certaine distance dans une certaine direction à partir de sa position courante.

[Sélection (Opération de déplacement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (Valeur)]
& [Sélection (Unité)]
& [Sélection (Direction)]

On constate déjà que pour la réalisation de ces deux tâches citées en exemple, les directions à indiquer pour les deux opérations mises en oeuvre présentent des différences dans chacune de ces deux expressions de commandes.

La direction indiquée par l'utilisateur pour un « positionnement » se définit en effet à partir de la position qu'il occupe virtuellement par rapport à la scène, alors que pour un « déplacement », l'origine de cette direction pointant vers la position de destination se situe à la position initiale de l'objet 3D.

Nous devons donc, par la suite, distinguer sans ambiguïté une direction jusqu'alors « *egocentrée* » d'une direction pour ainsi dire « *exocentrée* ».

Cette nouvelle spécification de direction d'origine localisée à l'objet de la commande constitue par conséquent, du point de vue de l'utilisateur, une nouvelle tâche atomique « *universelle* » à laquelle devra correspondre une spécification qui lui sera propre.

⁷ Le dictionnaire "Petit Larousse" définit la translation issue du latin *translatio*, transfert, comme "l'action de déplacer quelque chose ou quelqu'un d'un lieu dans un autre".

Ayant maintenant établi la possibilité de situer l'origine d'une direction indiquée par l'utilisateur de façon *exocentrée*, nous pouvons ajouter celle qui consiste à étendre les « *positionnements* » non plus par rapport à l'utilisateur, mais relativement à un repère identifiable de la scène 3D.

L'exemple le plus évident de repère identifiable se résume à la désignation d'un second objet 3D en tant que référence de positionnement.

[Sélection (Opération de positionnement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (Valeur)]
& [Sélection (Unité)]
& [Sélection (Direction)]
& [Sélection (Objet Graphique 3D)]

Nous obtenons ainsi un positionnement par rapport à une entité sélectionnable de la scène.

Par extension, nous pouvons également étendre les « déplacement » en ajoutant la possibilité de sélectionner une entité de la scène 3D qui constituera une cible vers laquelle s'orientera la direction de translation.

[Sélection (Opération de déplacement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (Valeur)]
& [Sélection (Unité)]
& [Sélection (Direction)]
& [Sélection (Objet Graphique 3D)].

Cette nouvelle notion d'orientation de la direction de mouvement vers une cible ajoutera ici une nouvelle forme d'expression de direction différente de celles déjà mentionnées.

Le groupe des translations basé sur différentes façons de spécifier une direction se déclinera donc par la suite en quatre types de mouvements. Les « *positionnements* » se départagent d'une part en positionnements absolus par rapport à l'utilisateur, d'autre part en positionnements relatifs à la position d'une entité désignable de la scène. Les « *déplacements* » quant à eux se déclinent en déplacements simples par rapport à la position initiale de l'objet du mouvement, et en déplacements *ciblés* en direction d'une entité désignable de la scène.

D'autres types de translations sont envisageables, tels que le positionnement par rapport à deux entités de la scène, ou encore un déplacement orienté par rapport à une orientation prédéfinie de la scène ou de l'objet 3D lui-même. Nous n'aborderons pas ces autres types de mouvements dans cette étude car soit ils constituent des mouvements composites ne manipulant plus directement une direction (donc situés à un plus haut niveau d'interaction), soit ils font appel à une définition plus complexe de la notion d'objet 3D qui déborde de cette étude.

Examinons maintenant les « mouvements giratoires » qui recèlent deux catégories, Les « *rotations* » et les « *girations* ».

Le dictionnaire "Petit Larousse" définit la « *rotation* », (du latin *rotatio*, de *rotare*, tourner), comme le "*mouvement d'un corps autour d'un point ou d'un axe fixe*".

Cette définition du mouvement d'un corps autour d'un axe ne précisant pas toujours la distance entre ce corps et l'axe ou le point autour duquel s'effectue le mouvement⁸, une ambiguïté subsiste quant à la distinction entre un mouvement de rotation local (sur lui-même) et la description partielle d'une trajectoire orbitale autour d'un point distant.

Nous conserverons par conséquent pour la suite le terme de « *rotation* » pour désigner plus spécifiquement le mouvement d'un corps tournant sur lui-même.

En ce qui concerne le second type de mouvement circulaire, le terme candidat et polysémique de "*révolution*" présente un caractère périodique supposant d'une part un retour au point initial pour décrire une orbite complète, et d'autre part, un mouvement continu⁹. Les mouvements continus ne seront pas abordés dans cette première approche qui ne porte que sur les mouvements d'amplitude finie. Nous utiliserons donc le terme moins usité de « *giration* », *mouvement giratoire, ou circulaire, du latin girare*¹⁰, afin de distinguer des « *rotations* » les mouvements d'un corps autour d'un centre géométrique distant du corps lui-même.

Une commande de rotation d'un objet se manifestera alors par un ensemble de spécifications similaires à celles employées pour les translations, avec deux différences notables. Il nous faudra En effet, pour cette catégorie de mouvement, d'une part exprimer une opération pour laquelle la destination reste identique à la position d'origine de l'objet du mouvement, et d'autre part spécifier un axe et un sens de rotation en lieu et place d'une direction linéaire de translation.

[Sélection (Opération de rotation)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (Valeur)]
& [Sélection (Unité)]
& [Sélection (Axe)]
& [Sélection (Sens)]

Une rotation d'un objet sur lui-même telle que décrite ci-dessus de manière générale ne trouve plus de « destination ». Le résultat de ce mouvement laisse l'objet du mouvement « sur place » en n'adressant qu'une modification d'orientation de celui-ci sur un unique degré de liberté. Cette modification d'orientation implique la sélection d'un axe et d'un sens de rotation autour duquel s'effectue le mouvement, ainsi que celle d'une quantité exprimée à l'aide d'une valeur numérique et d'une unité de mesure angulaire.

Il en ira a priori de même pour les girations avec spécification de la position du foyer du mouvement.

⁸ Le Dictionnaire de la langue française, (Hachette, 1980), définit ainsi le terme « rotation » : « Mouvement d'un corps qui tourne autour d'un axe. » Il précise ensuite pour exemple : « Rotation d'un astre sur lui-même ».

⁹ "Atlas d'astronomie", Perrin, 1989.

¹⁰ Dictionnaire de la langue française, Hachette, 1980, ISBN 2.73.8200.05.02

[Sélection (Opération de giration)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (Valeur)]
 & [Sélection (Unité)]
 & [Sélection (Axe)]
 & [Sélection (Sens)]
 & [Sélection (Objet Graphique 3D)]

Ce type de mouvement, qui suppose un centre géométrique autour duquel s'effectue le mouvement, peut être effectué là encore de façon egocentrée (autour de l'utilisateur), ou exocentrée (autour par exemple d'un autre objet de la scène).

Par ailleurs, dans le cas particulier d'un plan de référence normal à l'axe vertical, la giration peut éventuellement se simplifier par la spécification d'une direction linéaire prenant le rôle de tangente à la courbe.

[Sélection (Opération de giration)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (Valeur)]
 & [Sélection (Unité)]
 & [Sélection (Direction)]

Pour résumer maintenant les mouvements qui ont été mentionnés jusqu'à présent par catégories, et que nous retiendrons pour la suite, nous allons les réunir au sein d'un tableau simplifié (cf. Tableau 2).

Translation	Positionnement	Egocentré	Par rapport à la Position de l'utilisateur.
		Exocentré	Par rapport à la position d'un second objet sélectionné.
	Déplacement	Exocentré	Dans la direction exocentrée Indiquée par l'utilisateur.
		Ciblé	En direction d'un objet cible sélectionné.
Mouvement	Rotation	Autocentrée	Toujours autour de l'objet sur lui-même.
Circulaire	Giration	Egocentrée	Par rapport à la Position de l'utilisateur.
		Exocentrée	Par rapport à la position D'un second objet sélectionné.

Tableau 2. Catégories de « mouvements » en fonction de la position du repère spatial de référence.

D'autres « mouvements » d'objets 3D au sein de la scène sont à prévoir, ne serait-ce que d'un point de vue géométrique, que nous approfondirons lorsque nous aurons abordé les modalités ainsi que les dispositifs en entrée.

5.2 Types de modalités existantes

De même qu'il dispose de plusieurs canaux perceptuels l'informant sur l'état de son environnement (vue, ouïe, toucher, etc.), l'être humain dispose de plusieurs moyens d'expression qui lui permettent de communiquer ses intentions (regard, voix, mouvements, etc.). Ces canaux de communication présentent chacun des compétences à l'expression d'une modalité (ex. un index tendu dénote un geste déictique difficilement réalisable avec la voix).

Une tâche donnée, pour être exécutée par le système, doit correspondre à une commande reçue de la part de l'utilisateur, celui-ci peut spécifier cette commande de diverse manières dont l'efficacité est en partie prévisible et en partie à vérifier par des mesures de performances.

Pour reprendre un exemple déjà évoqué, une "indication de direction" est une opération élémentaire qui peut être manifestée de diverses manières par l'utilisateur au travers de diverses modalités d'expression. La direction du regard et le geste déictique manuel (pointage avec index tendu dans la direction de l'objet d'intérêt), en sont deux exemples.

Nous allons, au cours de ce chapitre, d'abord considérer brièvement les modalités de sortie ou de présentation de l'information à l'utilisateur. Nous aborderons ensuite les modalités d'entrée observables chez l'être humain en situation de communication.

Evoquer l'affichage comme modalité principale de sortie fait suite d'une part au fait que la vision tient une place prépondérante dans les tâches de conception d'artefacts. Elle fait suite d'autre part à certaines considérations issues d'enquêtes de terrain dans le domaine du stylisme modélisme de l'industrie textile. En particulier au cours d'étapes d'évaluation de prototypes vestimentaires, révélant en l'occurrence, pour de tels objets présentant un caractère esthétique, une inspection de prototype essentiellement basée sur l'aspect visuel qu'ils présentent. L'affichage en sortie doit par conséquent être en mesure de présenter l'apparence du produit de conception d'une part de manière réaliste ou efficace à des fins d'évaluation, d'autre part de manière adaptée à une exploration des sous-ensembles qui le composent éventuellement, et enfin de façon à satisfaire les besoins de l'interaction mise en place.

D'autres catégories d'information participent cependant de façon périphérique à l'apparence du produit de conception, en particulier en ce qui concerne les retours d'information relevant de l'interaction elle-même. Ces informations pourront se voir présentées de manière classique (*i.e.* textuelle) ou éventuellement sonore (par exemple, à l'aide d'une synthèse vocale).

Pour aborder maintenant les modalités d'entrée, rappelons tout d'abord que la plupart des techniques mentionnées au cours de la revue de question semblent pouvoir être regroupées dans un contexte de manipulation directe d'objets

graphiques 3D. Un tel contexte conduit naturellement, tout en expliquant partiellement le phénomène, à tenter d'enrichir le retour sensoriel de l'utilisateur par des dispositifs "haptiques" encore relativement encombrants à long terme.

Parmi les solutions envisageables à plus court terme évitant les modalités liées au sens tactilo-kinesthésique, se trouve notamment la possibilité d'exploiter les "actes communicatifs" de l'utilisateur en vue de permettre au système de déterminer les opérations souhaitées par celui-ci.

Avec les principales manifestations d'expressions de communication observable chez l'être humain, nous allons évoquer, du plus général au particulier, les mouvements et les verbalisations. Les mouvements les plus généraux vont d'abord concerner les déplacements de l'utilisateur ainsi que ceux de l'ensemble du corps et des membres. Viennent ensuite les mouvements des extrémités telles que les mains, les pieds, la tête et les mouvements oculaires qui y prennent place.

5.2.1. Mouvements généraux.

Déplacement de l'utilisateur, (mouvement le plus général),
Mouvements de l'ensemble du corps et articulations des membres.



Cette catégorie de mouvements pouvant comporter plusieurs modalités d'expression utilisateur ne sera pas traitée dans le cadre des modalités d'entrée, nous considérerons ici l'utilisateur en situation fixe par rapport à l'environnement. On se souviendra cependant que de nombreuses informations fournies par ces mouvements peuvent s'avérer nécessaire au système selon la plateforme mise en oeuvre, et ce, pas nécessairement en tant que modalités caractérisées par une interprétation de ces flots d'informations.

Par ailleurs, ce type de mouvement peut enrichir ou présenter une alternative à d'autres techniques plus couramment utilisées pour informer d'autres modalités.

5.2.2. Mouvements des mains et des doigts.

L'utilisation de ce medium pour en extraire des informations utiles à une interaction avec contrôle graphique propre à la Manipulation Directe fut l'une des techniques expérimentée dès la fin des années 80 avec l'apparition des premiers "Gants de Données". Les premiers concepteurs destinaient ce medium à une interaction directe de type préhension de l'Objet comme cela se pratique dans la vie courante, ou encore, dans d'autres applications, à une commande par gestes iconiques utilisant les langages de sourd-muets. Dans le même ordre d'idée, la métaphore du "*Marionnettiste*" permet de piloter les mouvements d'un objet articulé tel qu'un personnage de synthèse dans le monde de la post-production audiovisuelle.

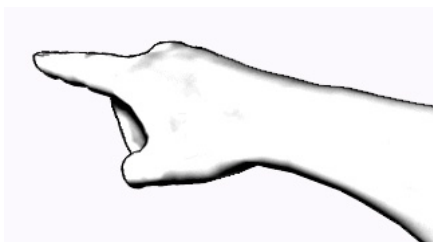


Certaines applications ont aussi cherché à le mettre en oeuvre pour en extraire une déictique permettant le pointage mais qui, à elle seule, montre rapidement des limites liées à l'imprécision, l'inconfort, et finalement une difficulté à concurrencer réellement les media déjà existants pour ce faire et accessibles pour un coût plus modeste.

Avec la fonction ergotique (Cadoz, 1994), notre main nous sert à interagir directement, physiquement, sur les objets usuels, et il est sans doute naturel que ce soit la première mise en oeuvre de ce medium à laquelle on songe avec des Objets virtuels. L'absence de retour haptique et proprioceptif reste cependant le principal écueil et met l'utilisateur dans la situation d'un accidenté conservant ses capacités motrices mais ayant perdu toute sensation provenant de la main. La fonction ergotique s'exerce alors en surchargeant le contrôle visuel nécessaire pour compenser l'absence de retour tactile mais obtenir le résultat d'une préhension, aussi bien en immersion qu'en pilotant une main virtuelle à distance.

Sous un aspect plus sémiologique, notre main nous sert aussi, dans la vie courante, à *signaler*, de façon approximative, une *Direction*, une *Distance*, à effectuer une *description spatiale* d'une forme, d'un volume en le dessinant littéralement dans l'espace (Bellik, 1995), suggérant une préférence de l'utilisateur quant à l'usage de la gestuelle manuelle pour convoier une information à caractère géométrique. Cette réalisation a été mise en oeuvre par exemple par Schkolne *et al.* sur le WorkBench, (Schkolne, Pruet & Schröder, 2001).

Plus rarement, la main sert à signaler quelques actes directifs (par exemple en environnement bruyé ou autre situation impliquant le silence). Des exemples de mise en oeuvre de la description spatiale à l'aide des mains révèlent l'aptitude particulière de cet unique medium pour le cas particulier de "création" d'une nouvelle forme, en particulier pour la description de formes curvilignes ou de concavités. L'homme ne dispose d'aucune autre modalité descriptive pour ce faire, sinon quelque dialecte verbal très spécialisé et relatif à la géométrie.



Deux cas particuliers de désignations monomodales effectuées à l'aide d'une main peuvent être pris en considération. D'abord celui qui indique une direction sans qu'une autre modalité ne vienne confirmer une sélection autrement que par l'énoncé vocal (par exemple lorsqu'on s'adresse à quelqu'un que l'on sommera de se rendre "*là-bas*" en le fixant du regard), et celui qui indiquera une direction pointant en dehors du champs de vision, comme on le pratique avec le pouce en évoquant un lieu défini par quelque identifiant lui aussi dans l'énoncé vocal.



Par ailleurs, la main utilisée seule peut également exprimer une distance d'un ordre de grandeur relatif à ses propres dimensions ou par rapport à un plan de référence.

Cette faculté est par exemple mise en oeuvre, comme déjà signalé au sujet de la technique "*Image Plane Interaction*", pour ajuster la valeur de l'angle solide d'un cône de sélection (cf. Figure 75 p. 92).

L'indication d'un écart mettant en œuvre deux doigts opposables et désignant sous forme de présélection maintenue une distance inférieure à une dizaine de centimètres, implique un acte supplémentaire de validation pour obtenir une sélection. Du fait qu'il est possible d'exprimer une telle distance à l'aide d'une seule autre modalité, une telle modalité alternative d'expression d'une distance ne semble se justifier que lorsqu'elle adresse directement une comparaison visuelle de l'écart qu'exprime la main avec une opération projetée sur l'objet de la tâche.

Pour ce qui concerne les actions utilisateur de plus bas niveau (présélectionner et valider) qui peuvent être effectuées à l'aide de ce médium d'information, l'usage d'une seule main permet donc d'effectuer au moins un geste déictique manifestant la présélection d'une direction. Elle permet éventuellement une validation moyennant quelques conventions de gestuelle manuelle lorsque n'apparaît pas d'ambiguïté d'indication de direction malgré l'imprécision du mouvement. Les cas d'exceptions restant cependant fréquents pour ce cas de figure, nous pourrions alors être conduit à considérer davantage le geste déictique manuel (avec par exemple l'index comme convention) non pas comme vecteur d'indication de direction, mais plutôt comme indicateur imprécis servant à valider un pointage plus précis exprimé à l'aide d'une autre modalité. Nous pourrions également devoir, à l'inverse, consigner ce geste déictique à la seule manifestation d'une présélection, en particulier celle d'une direction.



Figure 88 Indication d'un écart à l'aide des deux mains.

L'usage des deux mains de l'utilisateur ouvre un domaine entier mettant en scène deux exemplaires de ce même médium. Deux exemplaires d'un même médium de pointage permettent à priori de signaler par exemple deux points, (dont résulte un écart, une mesure de distance), deux directions, (dont résulte un angle), deux ensembles de points, donc éventuellement deux Objets Graphiques 3D. Il peut aussi être envisagé, après sélection des objets pointés par chaque main, d'ainsi maintenir une dépendance géométrique pour, par exemple, rapprocher ou séparer deux Objets ou groupe

d'Objets Graphiques 3D. Toutefois, la principale fonction permise par des pointages bi manuels et qui n'a pas d'équivalent reste la description de forme spatiale, le "geste du potier"(ex. Figure 89, Nishino et al., 1997, et Nishino et al., 1998.). Cette fonction tiendra une place particulière dans un environnement de conception où il pourra être question non pas de formes régulières mais relevant plutôt du modelage de formes irrégulières ("*free-form modeling*").

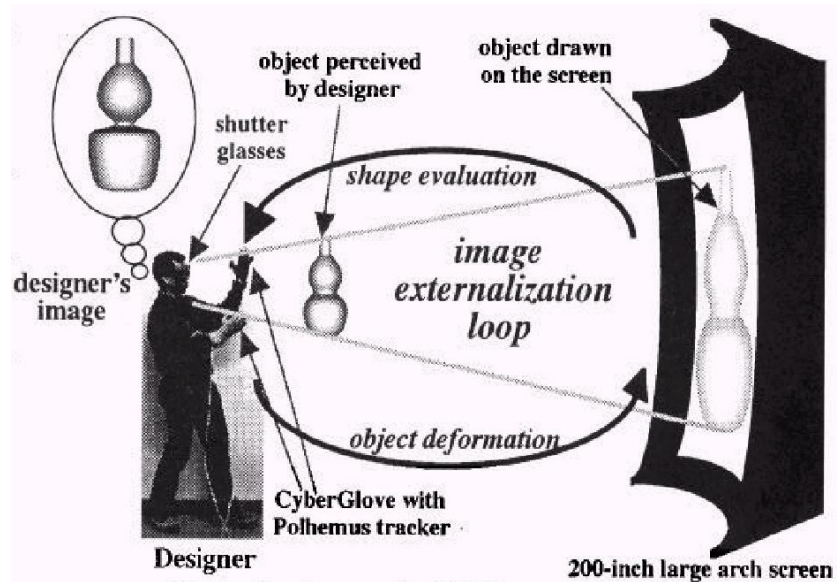


Figure 89 Exemple d'interaction adressant la modification bimanuelle d'une forme irrégulière. Illustration extraite de "3D Object Modeling using Spatial and Pictographic Gestures", H. Nishino, K. Utsumiya & K. Korida, VRST'98, pp. 51-58.

Une autre aptitude sans équivalent du geste bimanuel adresse les descriptions de formes spatiales au cours de situations d'interaction interpersonnelles (Figure 90).

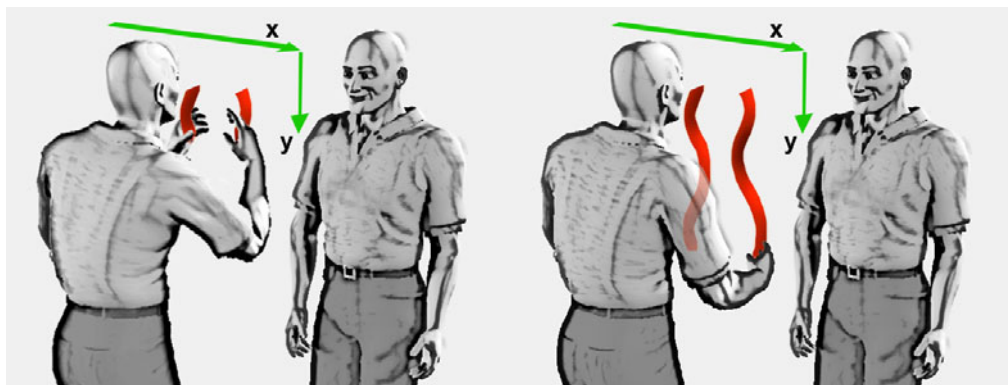


Figure 90 Illustration d'un gestuel bimanuel descriptif d'une forme "quelconque" prenant place dans un plan intermédiaire et normal aux deux interlocuteurs et indiquant des caractéristiques de description picturale 2D.

Ces mouvements mimétiques de description de formes spatiales peuvent s'avérer d'intérêt crucial pour des environnements de conception impliquant aussi bien les tâches de déformation que de création d'objets 3D. De telles descriptions adressent par ailleurs aussi bien les formes régulières qu'irrégulières. Ces catégories de tâches n'étant pas prises en compte dans le cadre de cette première approche, nous n'approfondirons pas davantage cette compétence des mains. Nous allons cependant considérer, dans le cadre d'indications liées à des mouvements dynamiques des

mains, les spécifications composites d'informations spatiales. En particulier, la description d'un mouvement circulaire de la main (Figure 91), traduit à la fois un axe

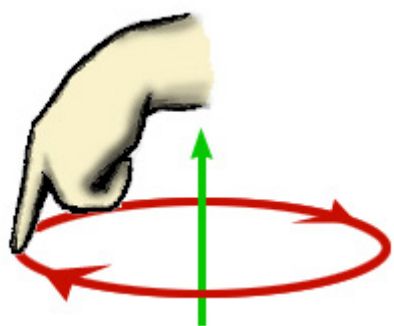


Figure 91. Mouvement circulaire d'indication de l'axe et du sens d'un mouvement circulaire.

autour duquel s'effectue ce mouvement ainsi qu'un sens de circulation. Ces deux informations constituent les paramètres tenant lieu de direction pour les tâches de rotation et de giration que nous avons définies. Par rapport à notre définition de la présélection, il s'agit bien d'une indication continue d'un état modifiable et mesurable, sous le contrôle moteur de l'utilisateur, ce qui en fait une présélection. D'un autre côté, la mise en œuvre d'une telle expression gestuelle manifeste également un acte volontaire de courte durée produisant une

interprétation immédiate qui, dans ce cas, en fait également une sélection implicite. Nous obtiendrons donc deux alternatives. Cette description spatiale sera d'abord une présélection d'axe et de sens de rotation à valider, par exemple à l'aide d'une seconde modalité. Ou bien elle constituera une sélection implicite qui désignera ces paramètres de mouvement circulaire.

Il nous faut enfin évoquer le cas du *geste instrumenté* (cf. Cadoz, 1994), qui peut augmenter les capacités de communication de la main. Une telle instrumentation de la main permet par exemple de spécialiser un outil préhensible pour effectuer une opération particulière (stylo laser pour un pointage précis sur une surface, couteau ou scalpel pour inciser ou trancher, "Gomme" 3D pour "trouer" dans une surface, etc.), ou encore pour signaler une conduite, une consigne (baguette du chef d'orchestre, drapeau, raquette de signalisation à motif explicite sur chaque face, etc.).

D'autres usages sont aussi envisageables si l'on retient la possibilité d'en extraire des actes directifs indiquant ou mimant un mouvement (translation ou rotation) à imprimer à un Objet sélectionné (gestes mimétiques), bien que ceux-ci, en première instance, relèvent *a priori* davantage de la manipulation directe.

Cet aperçu non exhaustif de la compétence des mains à exprimer des indications spatiales révèle ainsi une richesse d'expression de ce médium dans le cadre d'une interaction orientée communication, qui justifiera son emploi pour une interaction avec un environnement virtuel.

5.2.3. Mouvement des pieds (et des orteils).

L'utilisation éventuelle de ce medium pour en extraire des informations de modalité semble adresser d'éventuelles formes de manipulation directe à l'image de la façon dont on s'en sert pour conduire un véhicule, ou encore pour modifier le son d'un instrument musical. La littérature rapporte un exemple de contrôle locomoteur le mettant en œuvre pour une navigation au sein d'un environnement virtuel « main libre » (LaViola et al. 2001).

L'usage de cet appendice ne semble pourtant pas présenter d'indice exploitable dans le cadre de modalités de communication.

5.2.4. Mouvements de la tête.

Les mouvements de la tête, qui pourraient figurer précédemment dans les "Mouvements généraux de l'ensemble du corps et articulations des membres", peuvent tout d'abord être considérés comme une information nécessaire dans le cas d'utilisation d'un dispositif d'affichage portable, sans mériter alors l'appellation de "modalité".



Ces mouvements de la tête sont relevés dès lors que se voit mis en œuvre un dispositif d'affichage portable (HMD) ou un suivi du regard pour lequel il est nécessaire de rapporter l'orientation des yeux à celle de la tête. Le "Head-Tracking" (suivi de la position de la tête) est donc souvent utilisé sur de nombreuses plates-formes expérimentales sans que n'en soit pourtant, semble-t-il jusqu'à présent, extrait de modalité gestuelle exploitable autre que par exemple la "Navigation Orbitale" permettant, par les seuls mouvements de la tête de tourner autour d'un unique Objet 3D d'intérêt (Koller et al., 1996), ou encore l'indication d'une direction de déplacement virtuel de l'utilisateur, comme par exemple le "*Virtual Tour*" exposé au salon de Laval en 2002, (Conférence VRIC 2002). Cette application de "Voyage virtuel" présentée au salon de cette conférence permet d'établir *de facto* l'aptitude des mouvement de la tête à définir au moins une "*Direction*" égocentrée de locomotion.

En ce qui concerne la possibilité d'extraire de ce medium une modalité d'expression utilisateur, on pourra considérer d'une part la "gestuelle" quasi-déictique du menton parfois utilisée pour désigner la direction d'un objet visé du regard par un locuteur, et d'autre part l'*acquiescement* ou *désaccord* par un mouvement répété de lacet ou de tangage de la tête sur l'axe cervical.

Une telle information pourrait ainsi tenir lieu de modalité de validation d'une présélection effectuée par ailleurs (ou mieux encore, au cours d'un dialogue). On devra alors se souvenir que ces mouvements présentent éventuellement une composante culturelle importante. Ainsi, en Grèce, l'*acquiescement* ressemble à notre mouvement de négation, et le *désaccord* est manifesté par un unique mouvement emphatique soutenu de tangage vers le haut en abaissant les paupières.

Une telle modalité peut trouver usage lors, par exemple, de la mise en oeuvre par le système de l'équivalent actuel des boîtes de dialogues du GUI n'impliquant qu'un acquittement ou une annulation de la part de l'utilisateur.

En ce qui concerne l'indication de direction, certains mouvements de la tête peuvent être considérés comme des actes continus d'orientation globale en direction d'un objet d'intérêt. A ce titre, ces mouvements présentent les caractères d'une présélection de direction correspondant au centre du champ de vision. Certains chercheurs considèrent d'ailleurs le résultat d'une telle orientation de la tête comme une réduction de la « direction du regard » (Arns & Cruz-Neira, 2002). Une méthode de *présélection* d'objets basée sur une telle modalité mise en oeuvre seule est donc possible mais impliquerait sans doute un environnement distinguant les objets de la tâche sans ambiguïté du fait d'un manque de précision évident.

Nous avons évoqué l'éventualité d'indiquer une direction à l'aide du menton par un mouvement de tangage de la tête. Ce mouvement ne suffit cependant pas, seul, à indiquer effectivement une sélection (dont la présélection doit être maintenue, de préférence, simultanément).

5.2.5. Mouvements faciaux : Mouvements Oculaires.

La place que tiendra la direction du regard au sein des actions utilisateur de plus bas niveau (viser et valider), restera, comme le suggère la littérature, celle d'un *pointeur* très rapide (Ware & Mikaelian, 1987, Tanriverdi & Jacob, 2000, Sibert & Jacob, 2000), dont la précision à terme sera supérieure à celle obtenue par d'autres moyens directs. Toujours selon certains auteurs (ex. Zhai et al., 1999), la direction du regard est difficilement envisageable comme modalité de validation du fait que les mouvements oculaires perceptibles constituent davantage une conséquence involontaire de l'exercice de la vision qu'un mouvement volontairement issu d'un organe de commande.



La difficulté posée par la direction du regard consiste en ce qu'elle fonctionne en permanence sans nécessairement impliquer une désignation dont on pourrait extraire une sélection. L'usage de ce médium d'entrée implique alors soit d'interpréter tous les mouvements des yeux (exploration, intérêt soutenu pour un objet, introspection qui laisse le regard "dans le vague", etc.), soit de n'utiliser que ce qui est aisément interprétable. La mise en oeuvre de la seule composante déictique lorsque l'intérêt de l'utilisateur pour un objet se trouve confirmé par ailleurs poserait sans doute moins de difficulté que chercher à interpréter l'ensemble de l'évolution du regard.

D'autre part, le regard présente une activité continue (bien que saccadée) tant que les paupières restent ouvertes. Toutes ces caractéristiques en font par conséquent un candidat à la visée produisant une présélection.

5.2.6. Énonciation vocale.

Ce médium est l'un de nos principaux outils de communication. A lui seul, il peut être utilisé au cours de conversations téléphoniques, avec des applications concernant par exemple la recherche d'information et la réservation de titres de transport, par réduction à un sous-ensemble du langage (cas de la réservation SNCF).



D'autres terrains d'application utilisent ce médium à l'aide de langages d'échange d'information ou d'opération tels que protocoles d'énonciation radio dans le domaine des transmissions (radioamateurs, transmissions militaires), ou encore le commandement aveugle avec retour sonore de confirmation.

Le langage véhiculé par la parole prévoit la possibilité de désigner spécifiquement une entité d'intérêt dont la localisation est indiquée à l'aide d'une modalité gestuelle d'indication de direction. Il s'agira alors, par exemple, de l'énoncé comportant un adjectif démonstratif précédant l'évocation de la classe à laquelle appartient l'objet d'intérêt dans la direction indiquée par ailleurs. Si l'on admet pour acquis que le destinataire du message est effectivement à l'écoute de l'émetteur, cette façon de procéder revêt un caractère implicite de sélection de par sa seule énonciation.

La parole paraît dénuée d'aptitude à une indication conditionnée par une validation simultanée qui en ferait un outil de présélection. Elle semble au contraire, dans le cadre d'une communication interpersonnelle, constituer une modalité de choix pour effectuer la validation d'une information transmise par un canal gestuel.

Un mouvement de translation d'objet 3D est néanmoins concevable sur ce seul canal de communication par l'énonciation de spécifications de vitesse et de direction de progression. Une autre façon de se rendre en un endroit peut consister à "demander" à y être emmené (ex. Ballegooij & Eliéns, 2001). Dans ce cas qui rejoint le pointage (spécification "explicite" d'une destination), une modalité vocale en entrée pour la réalisation d'une tâche de "locomotion" pourrait apporter une amélioration mesurable de manipulation de l'environnement. Toutefois, un tel usage monomodal suppose que la destination dispose d'un identifiant énonçable.

En revanche, effectuer une description spatiale d'un objet complexe de forme irrégulière s'avérera malaisée par le seul usage de la parole, nous utilisons généralement pour ce faire soit des évocations d'autres objets de référence supposés faire partie d'un savoir commun, soit l'usage d'un autre médium complémentaire d'information plus approprié tel que des mouvements des mains.

Malgré l'obligation de structurer davantage le dialogue, la richesse d'expression de ce canal de communication autorise sans doute d'importants développements dans le domaine de la conception, même en tant que modalité unique d'entrée, par exemple en ce qui concerne la construction et l'aménagement d'une scène (Tamura, 2000).

On a précédemment signalé une approche particulière sous l'appellation de "*Declarative Modelling*". Si une telle approche s'avère possible avec un canal d'entrée dactylographique, elle présente des caractéristiques *a fortiori* accessibles au canal vocal d'expression pour la description d'une scène à construire par le système. (Coyne & Sproat, 2001).

D'autres exemples vont concerner l'aménagement de la scène par le système sous la conduite vocale de l'utilisateur, il s'agira donc là de "Manipulation" sans sélection explicite de l'objet autrement que par son appellation (Tamura, 2000). Dans ce type de scénario, la désignation de l'objet peut vraisemblablement se voir effectuée à l'aide d'un identifiant qui le distinguera sans ambiguïté (son nom, sa catégorie s'il en est le seul représentant dans la scène, ou une caractéristique dont il sera le seul objet de la scène à en disposer).

Si la définition d'un objet 3D présent sur la scène comporte un identifiant énonçable qui le désigne ou qui en distingue la catégorie (ex. pour la conception de vêtement, "col", "manche", "bouton", etc.), une présélection n'est pas mise en œuvre. Une telle situation rappelle davantage la recherche d'un document dans une base de données à partir d'un terme identifiant l'objet ou sa catégorie, constituant ainsi une sélection implicite de l'objet de la commande.

L'énonciation ne présente donc pas de caractère continu modifiable et maintenu en attente de validation, mais produit en effet des successions de mots dont certains présentent les caractéristiques de sélections (actes volontaires de courte durée produisant une interprétation immédiate convoyant une information de désignation non ambiguë), cependant que les autres structurent l'énoncé, agissent comme modificateurs de l'information exprimée par un autre terme, ou tiennent un autre rôle.

En particulier, l'énonciation d'un verbe impératif traduisant une opération à exécuter constitue à elle seule une sélection.

D'autres termes ou groupe de termes d'un énoncé présentent de telles caractéristiques que nous approfondirons plus loin lorsqu'il sera question de commande multimodale.

5.2.7. Combinaisons de modalités.

Afin d'aborder maintenant diverses combinaisons des modalités que nous avons évoquées, nous allons d'abord les réunir dans un tableau de distribution sous forme de diagramme de Venn. La Figure 92 montre une telle distribution. Encadrée de bleu s'y trouve la direction du regard, le cadre vert délimite la parole. La colonne encadrée d'orange concerne ensuite les mouvements de la tête, alors que la colonne noire encadre la mise en œuvre de la main. Les mouvements généraux du corps tels que déplacements de l'utilisateur, non traités, sont cependant figurés en dehors de cette distribution pour rappeler qu'ils peuvent éventuellement jouer un rôle vis-à-vis d'un autre groupe de modalités.

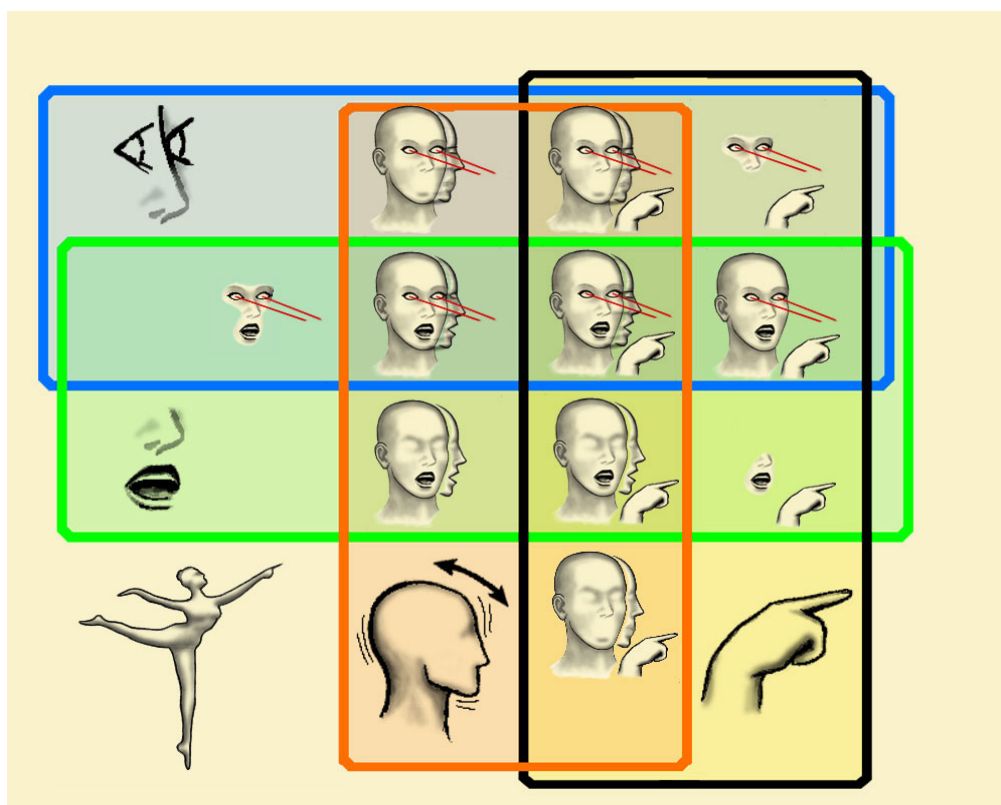


Figure 92 Distribution combinatoire des organes supportant diverses modalités observables de communication.

Une telle représentation permet aisément d'identifier de façon visuelle les diverses combinaisons deux à deux, ou combinaisons bimodales au nombre de six. Elle permet également de distinguer ces dernières des quatre triplets de modalité, qui eux-mêmes encadrent l'unique quadruplet.

Combinaisons de deux modalités

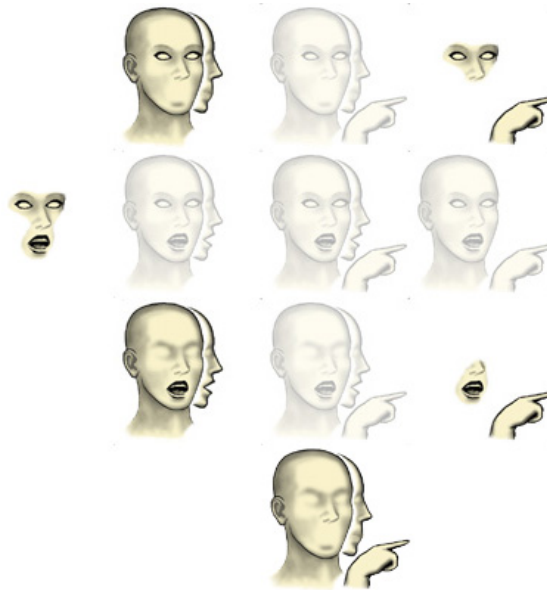
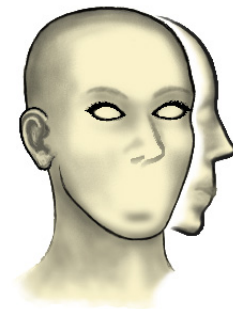


Figure 93 Combinaisons bimodales de la distribution.

Combinaison de la direction du regard avec les mouvements de la tête.

On dispose, avec une telle combinaison, de la possibilité d'extraire deux indications de direction différentes (direction de la tête et celle de visée oculaire) ainsi qu'une validation si l'on accepte d'extraire quelques gestes iconiques de façon restreinte avec une gestuelle de la tête.

La direction de regard reste cantonnée à ne fournir qu'un pointage rapide d'une direction, éventuellement d'un point à courte distance si l'on dispose de la direction de chacun des deux yeux (on dispose alors de trois indications de direction).



Dans un tel cas de figure, l'utilisateur dispose de la possibilité, en combinant les deux modalités, de *Sélectionner* un *Point*, *Ensemble de Points* ou un *Objet Graphique*, et une *Direction*. On réservera sans doute ces deux media d'information à ce seul usage, car cette configuration se révélera sans doute cependant inappropriée à la *Sélection* d'un *Point* à la surface d'un *Objet Graphique 3D* ou à celle d'une *Propriété* d'un *Objet Graphique 3D*, (encore que l'on pourrait convenir d'un geste iconique supplémentaire signifiant l'incompréhension et traduisant une interrogation concernant l'Objet en cours de pointage oculaire, le cas existe mais il n'est pas forcément présent dans toutes les cultures).

Par rapport aux tâches utilisateur de bas niveau (présélection et validation), on a indiqué l'éventualité assez rare d'indiquer une direction à l'aide du menton (pointage, donc) par un mouvement de la tête en tangage. Ce mouvement ne suffit cependant

pas, seul, à indiquer effectivement une désignation (dont le pointage se trouve alors validé). Lorsque ce mouvement apparaît effectivement au cours d'une désignation, une autre modalité d'expression le confirme en tant que pointage, il s'agira généralement d'une direction de regard soutenu sur la cible, laquelle modalité se trouve aussi être une présélection. Nous conviendrions donc, dans la vie courante, qu'une double présélection à l'aide de deux modalités différentes peut, dans certains cas, constituer une désignation à part entière, l'une des deux modalités d'expression validant le pointage effectué par l'autre. Une décomposition de ce mouvement de désignation traduit en fait l'inverse, c'est la direction de regard qui présélectionne l'objet, et le mouvement de la tête, acte volontaire de courte durée produisant une interprétation immédiate, qui valide en produisant une sélection, cependant que le regard reste fixé sur l'Objet au cours du mouvement.

Il n'est pas certain que ce mouvement particulier de la tête appartienne aux universaux du comportement humain, ce qui en fait éventuellement une modalité d'origine culturelle dont l'usage se verrait simplement restreint s'il n'existe pas dans d'autres cultures. Le danger d'un tel usage réside cependant dans la possibilité qu'un tel mouvement existe et soit porteur d'une autre signification dans une autre culture, comme c'est le cas concernant l'acquiescement ou la négation pour les seuls mouvements de la tête.

Combinaison de la direction du regard avec la gestuelle manuelle.

Nous retrouvons là encore une situation mettant en jeu deux media pouvant convoier un pointage déictique dont l'un, plus précis que l'autre, nécessite via une autre modalité la confirmation de son caractère déictique ainsi qu'une validation.



Le rôle prédominant de pointage tenu par la direction du regard n'est pas exclusif, la désignation à l'aide de la main peut aussi, bien que moins précise, tenir un rôle d'indicateur de direction, ce sera notamment le cas lorsque l'utilisateur indique de l'index une autre direction que celle dans laquelle il regarde, le cas est cependant particulier en ce que la direction mentionnée du doigt devient relative, et de façon très approximative, à un lieu le plus souvent défini dans le contexte par un énoncé vocal éventuellement antérieur au geste déictique.

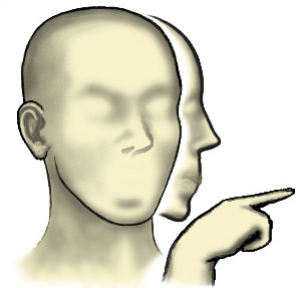
Nous considérerons donc en premier lieu le geste déictique manuel comme une éventuelle validation de présélection obtenue par le regard, en ne conservant provisoirement que le principe selon lequel ce geste manuel consiste effectivement en un geste volontaire dénotant une désignation, ne requérant pas de précision, mais permettant la confirmation d'une intention utilisateur de désignation. Ce type de rôle revient à encoder les mouvements de la main pour produire un acte volontaire de courte durée produisant une interprétation immédiate.

Le second rôle que peut tenir la main concernera les compétences particulières qui distinguent ce médium des autres, à savoir, comme nous l'avons évoqué plus haut, au sujet du rôle de description spatiale de formes à l'aide d'une gestuelle dynamique des deux mains. Il s'agira alors, par exemple, d'un pointage continuellement validé pour peu qu'une telle description ait été préalablement déclarée, (ce qui en fait encore un usage encodé). Ces deux médias d'entrée permettent donc, en première considération, le pointage, la validation, la description spatiale de formes, ainsi qu'une double présélection si les mouvements des deux mains sont renseignés.

Cette double indication de direction peut ainsi tenir le rôle de « pointeur » indiquant une position dans l'espace proche de l'utilisateur à l'intersection des deux directions manifestées. Ce qui rappellera le même éventuel rôle d'une double direction évoqué pour la seule direction du regard.

Combinaison des mouvements de la tête avec la gestuelle manuelle.

Nous retrouvons, avec cette combinaison de modalités, approximativement la même situation que précédemment avec une moindre précision. Nous disposons ainsi de deux indicateurs de direction. Nous disposons encore de deux modalités pouvant chacune signifier une validation de ce que présélectionne l'autre sous les mêmes conditions d'usage déjà évoquées.



Combinaison des mouvements de la tête avec la parole.

Les mouvements de la tête peuvent, on l'a déjà signalé, fournir l'indication d'une direction. Ils sont d'ailleurs utilisés dans certains cas comme "direction de regard" sur certaines plateformes pour lesquelles on va considérer que l'utilisateur tourne généralement la tête de façon à faire face aux objets d'intérêt. Bien qu'une telle "direction de regard" reste approximative, elle permet néanmoins, comme c'est le cas dans l'application "Virtual Tour", d'indiquer une direction efficace de déplacement virtuel dont l'utilisateur peut corriger la trajectoire en cours de déplacement. L'énonciation vocale apparaissant comme second médium disponible dans le présent cas de figure, peut tenir le rôle de validation d'une direction et, par extension, une sélection d'objet de la scène, ainsi bien sûr que celui d'entrée de commandes à exécuter. Le médium de pointage restant approximatif, une sélection efficace indiquerait un faible peuplement de l'environnement en Objets 3D ainsi que la nécessité de s'en rapprocher pour sélectionner sans ambiguïté.



Ce qu'apporte cette combinaison de modalités, et ceci tiendra pour chaque association de la parole avec une modalité permettant la présélection d'une direction égo-centrée, est l'indication possible, en tenant compte de la sélection d'un objet

quelconque de la scène, d'une direction exclusivement vocale. La sélection d'un objet de la scène fournit en effet au système deux positions virtuelles privilégiées dont l'une est occupée par l'utilisateur, l'autre par l'objet de la sélection. Du fait que l'axe vertical est généralement commun à l'utilisateur et au système, ces deux positions délimitent un segment au bout duquel peut être manifestée vocalement une direction strictement orthogonale (ex. « *vers la droite* »). La parole permet difficilement, en effet, l'indication d'une direction quelconque, mais se prête bien, et de façon exclusive, à l'indication de directions orthogonales en s'appuyant sur un repère désigné au sein de l'environnement. Nous formulons habituellement ce type de direction par des énonciations telles que « *sur la droite* » s'agissant de quelque objet préalablement désigné, ou encore « *à droite de* » précédant un objet désigné. Pour la suite, nous distinguerons ces directions particulières sous le terme de « *directions orthogonales* ».

En dehors de ces quelques considérations, les mouvements de la tête conservent, pour les dispositifs d'affichage portables, le rôle qui permet la cohérence géométrique de présentation de la scène à l'utilisateur lors d'une orientation de la tête.

Combinaison de la parole et du mouvement des mains.

Dans ce cas de figure, l'indication de direction est encore effectuée à l'aide de la direction indiquée par la main, on retrouvera les mêmes observations que précédemment concernant une imprécision relative de ce médium qu'on utiliserait alors de préférence pour une interaction de proximité. Le médium vocal tenant le même rôle que précédemment pour l'énoncé de commandes à exécuter.



On verra ainsi par exemple se distribuer, au sein d'une proposition comportant une composante spatiale, une simplification de la composante spatiale sur la modalité déclarative (l'énoncé vocal, ex. "*Allez voir là-bas si j'y suis!*") ne permettant pas, à lui seul, de savoir où se trouve "là-bas"), complétée par la mise en oeuvre d'un autre canal d'expression en entrée apportant davantage de possibilités d'expression pour indiquer une direction en permettant notamment d'écourter la durée de l'énoncé (le geste déictique).

Une indication de direction egocentrée et/ou exocentrée ainsi qu'une sélection de celles-ci à l'aide du médium vocal deviennent possibles. L'usage de la main seule à des fins d'indication de direction soit egocentrée, soit exocentrée, impliquera quelque spécification supplémentaire de désambiguation du même geste déictique mis en oeuvre, par exemple de manière vocale. Des tâches plus complexes deviennent toutefois ainsi accessibles. En dehors de ce que chacune de ces modalités apporte utilisée seule et de ce qui vient d'être mentionné, il semble que cette association ne permette guère plus que la validation vocale de quelques présélections manuelles.

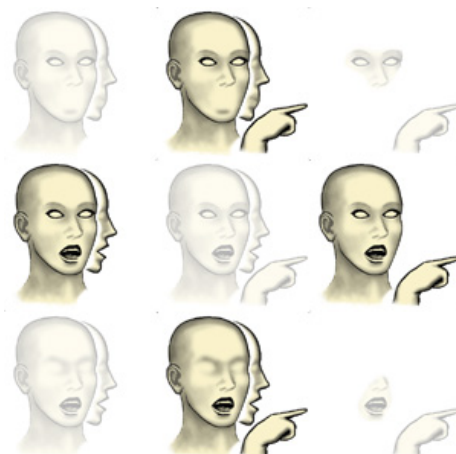
Combinaison de la direction du regard avec la parole.

Cette combinaison permet de tirer profit des présélections obtenues par visée au sein de l'environnement. Cette visée apporte davantage de précision que les gestes déictiques manuels ainsi sans doute qu'un gain de temps prévisible. L'association conserve donc cette capacité de sélection d'objets et de directions au moins égocentrées et orthogonales, ouvrant l'accès aux tâches de translation d'objets. En revanche, les tâches de mouvements circulaires d'objets deviennent problématiques mais restent possibles à l'aide d'indications vocales plus complexes.



Combinaisons de trois modalités

Nous allons maintenant examiner les quatre combinaisons trimodales en les comparant entre elles à partir de ce qui a été mentionné en terme d'apports pour les combinaisons bimodales. Les combinaisons de trois groupes de modalités apportent une complémentarité de tâches accessibles qui augmentent *a priori* l'efficacité d'expression d'une commande.



Sans manifester de réelles redondances, certaines modalités présentent en effet certaines équivalences pour certaines tâches, ce qui permettra de leur attribuer des affectations préférentielles selon le meilleur compromis.

Combinaison de la parole, du mouvement des mains et des mouvements de la tête.

Cette combinaison de trois modalités permet de cumuler les avantages de chacune des associations bimodales incluant la parole qui la composent.

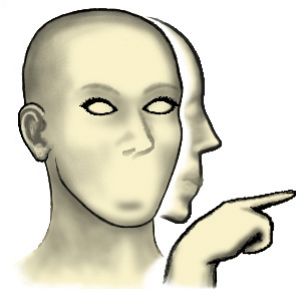
Disposer de la gestuelle manuelle permet de plus d'affecter plus particulièrement celle-ci, en terme d'indicateur de direction, aux indications de directions exocentrées (cependant que la direction de la tête conserve son seul rôle possible d'indicateur de direction égocentrée). De cette manière se voit supprimé l'usage alternatif de la main pour indiquer deux catégories



différentes de directions (egocentrée vs exocentrée). Cette affectation qui rend ainsi complémentaires les modalités vis-à-vis des tâches impliquant des directions, permet à cette combinaison de manifester sans ambiguïté tous les différents types de direction qui ont été évoqués (ainsi que le sens giratoire d'un mouvement circulaire). On obtient alors une multimodalité en entrée qui couvre l'ensemble du spectre des tâches envisagées, celles de mouvement d'objets.

Combinaison de la direction du regard, des mouvements de la tête et du mouvement des mains.

Cette combinaison présente deux modalités équivalentes d'indication de directions egocentrées qui en réduisent la complémentarité. Par ailleurs, en l'absence de la parole, les sélections redeviennent implicitement gestuelles.



Comparée à la combinaison précédente, celle-ci ne permet plus les sélections implicites liées à l'usage de la parole ni, par conséquent, l'accès aux directions orthogonales.

Combinaison de la direction du regard, de la parole et des mouvements de la tête.

Cette combinaison présente les deux mêmes modalités équivalentes d'indication de directions egocentrées qui en réduisent également la complémentarité. D'autre part, en l'absence de la main, elle ne permet plus la spécification d'une direction exocentrée ni d'un axe et d'un sens de rotation en un seul mouvement de celle-ci. Cette combinaison ne fournit plus, par conséquent, un accès aisé aux tâches de mouvements circulaires.



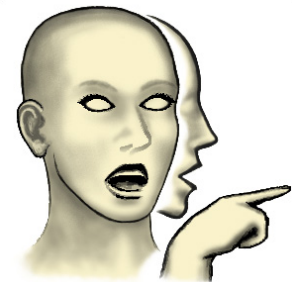
Combinaison de la direction du regard, de la parole et du mouvement des mains.

Cette association de modalités présente les mêmes caractéristiques que la combinaison de la parole, du mouvement des mains et des mouvements de la tête. Elle apporte de plus les avantages propres à la direction du regard, un gain de temps prévisible, ainsi qu'une meilleure précision de la visée.



Combinaison de la direction du regard, de la parole, du mouvement des mains et des mouvements de la tête.

La dernière combinaison réunissant les quatre modalités qui ont jusqu'ici été distribuée de façon combinatoire est équivalente à l'association de la direction du regard, de la parole et du mouvement des mains. Cette ressemblance est due à l'équivalence des directions egocentrées exprimées par l'orientation de la tête et la direction du regard. Cette association n'apporte donc rien de plus que la combinaison trimodale équivalente, en supposant un bénéfice plus faible du spectre de tâches accessibles sur le nombre de modalités mises en oeuvre.



Nous avons examiné l'ensemble des combinaisons de modalités distribuées de manière combinatoire. Nous constatons comme prévisible que sur l'ensemble de ces associations, seules deux combinaisons de trois modalités complémentaires couvrent l'ensemble du spectre des tâches utilisateur évoquées, permettant les mouvements d'objets. Ces deux combinaisons présentent les caractères suivants : l'indication d'une direction egocentrée est affectée à un medium spécifique, l'indication d'une direction exocentrée est attribuée à un medium distinct. L'une des modalités gestuelles permet d'exprimer un axe et un sens de rotation nécessaires à l'expression d'une commande de mouvement circulaire, et une modalité particulière permet la sélection des directions, des objets de la scène, ainsi que d'autres sélections implicites elles aussi nécessaires.

5.3 Types de dispositifs correspondants.

La réalisation d'une tâche par le système peut être commandée par l'usage d'une ou plusieurs modalités d'expression en entrée. Chacune des modalités envisagées, lorsqu'elle est mise en oeuvre par l'utilisateur, consiste en une information qui doit être convoyée au système. Cette transcription repose sur l'usage de dispositifs techniques qui permettront de mesurer les variations d'expression significatives que manifesterà l'utilisateur pour chaque modalité envisagée.

Un examen des dispositifs au cours de la revue de questions a permis ensuite de classer ceux-ci en terme de compétence vis-à-vis des modalités auxquelles on souhaite les attribuer.

La modalité déictique manuelle peut ainsi être transmise au système à l'aide de plusieurs techniques. Une méthode exocentrée utilisera par exemple, en guise de dispositifs, des caméras d'observation qui transmettront au système les images reflétant les mouvements de l'utilisateur. Un traitement d'image, dont une reconnaissance de forme, permettra alors au système de déterminer si l'utilisateur pointe un objet graphique à l'aide d'un doigt tendu. Une autre méthode plus centrée sur l'utilisateur (parfois baptisée "egocentrée"), utilisera par exemple un tracking magnétique de la main ainsi qu'un gant de données fournissant les informations de

position, d'orientation et de flexion des phalanges des doigts de l'utilisateur. Un traitement de données supplémentaire devra ensuite permettre de détecter quand l'utilisateur effectue un pointage déictique manuel. Plusieurs techniques peuvent ainsi parfois permettre de détecter l'expression d'une même modalité en entrée, ce qui permet de classer les modalités en termes de compétence et de capacité à transmettre une expression modale.

On considère également que les dispositifs techniques permettant de transmettre les expressions utilisateur ne sont pas forcément disponibles. Tout au plus pouvons-nous, dans le cadre d'une approche dite « Top-Down », définir un portrait-robot, un cahier de charges auquel devrait satisfaire le dispositif idéal propre à transmettre au mieux l'expression de telle ou telle modalité. En l'absence, le cas échéant, du dispositif idéal pour ce faire, une classification des dispositifs existants permet de s'en approcher en tenant compte de la maturité respective des techniques.

Ces trois notions de tâches utilisateur, de modalités d'expression utilisateur en entrée, et de dispositifs techniques, permettent d'établir des correspondances qui semblent nécessaires en vue de l'élaboration d'une grammaire multimodale en entrée du système. Certaines modalités paraissent en effet spécifiques à un type d'opération de façon quasiment exclusive (exemple, la spécification d'une information à caractère spatial est effectuée de façon prédominante à l'aide des mains). A l'opposé, certaines opérations peuvent requérir différentes combinaisons de modalités éventuellement redondantes (exemple, la spécification d'une direction soit avec le regard et l'index tendu sans énonciation vocale, soit à l'aide du regard et d'une énonciation vocale sans mouvement des mains, etc.). L'examen d'une telle correspondance permet de déterminer, par type de tâche, les combinaisons modalités-dispositifs dont l'emploi couvre l'essentiel du spectre d'expression utilisateur en regard des tâches à effectuer. Cet examen permet également, sous forme de scénarios, de mettre en évidence les conflits qui peuvent résulter de certaines affectations de modalités à certains dispositifs pour des tâches différentes. D'autre part, le principe d'économie mettra en avant l'intérêt à user d'une modalité mesurable à l'aide d'un dispositif initialement prévu pour un autre usage quand c'est possible. Ainsi, les mouvements des mains seront-ils sans doute mis en oeuvre à des fins de déictique manuelle, on peut tabler sur le fait que d'autres modalités pourront être exprimées à l'aide du même dispositif que celui qui permettra la détection d'une telle modalité déictique, qu'il s'agisse de caméras ou de gants de données. Le dispositif qui permettra de déterminer ce que montrent éventuellement les mains de l'utilisateur couvre ainsi un spectre de modalités d'expression qui en mesure l'importance. En ce qui concerne le principe d'économie, le hochement de tête peut par exemple parfois signifier une direction (on dit aussi "pointer du menton"), une telle modalité d'expression d'une direction n'est certainement pas celle qui prédomine pour ce faire, au point qu'on pourrait tout simplement la négliger. Cependant, si les mouvements de la tête de l'utilisateur sont mesurés pour satisfaire par ailleurs, par exemple, le calcul de l'affichage graphique,

on aurait sans doute tort de ne pas mettre ce geste déictique en oeuvre dont l'usage pourra occasionnellement s'avérer d'intérêt, aussi rare fut-il.

Nous avons, au chapitre précédent, examiné les compétences de diverses modalités dénotant un rôle communicatif dans le cadre de tâches de mouvements d'objets à réaliser. Nous avons également envisagé diverses combinaisons de ces modalités pour évaluer le rapport du spectre d'expressions accessibles au nombre de modalités d'entrée à renseigner.

Nous avons convenu que deux combinaisons de trois modalités particulières semblent couvrir les besoins de spécification nécessaires aux tâches utilisateur définies pour les mouvements d'objets.

Nous allons maintenant évoquer à un niveau plus technique le profil des dispositifs nécessaires à la prise en compte des informations produites par les sources de modalités retenues en entrée. Ce faisant, nous tiendrons compte des éventuelles contraintes apportées par les techniques correspondant à la modalité principale en sortie, c'est-à-dire par les dispositifs permettant la présentation de l'objet de la tâche.

Les deux combinaisons de modalités candidates sont :

Une combinaison de la parole, du mouvement des mains et des mouvements de la tête.

Une combinaison de la parole, du mouvement des mains et de la direction du regard.

Pour commencer par ce qui distingue ces deux combinaisons en participant aux tâches de plus bas niveau, examinons d'abord les profils des dispositifs de suivi des mouvements de la tête, puis ceux permettant d'obtenir la direction du regard.

5.3.1. Mouvements de la tête

Relever les mouvements de la tête, qui dispose d'une « face », ou en d'autres termes d'une orientation, revient à déterminer sa position dans l'espace ainsi que cette orientation de la « face ». Selon l'affichage utilisé en sortie, cette position sera nécessaire pour modifier la représentation de la scène en fonction de la position de la tête de l'observateur (en fait, de la position de l'œil dominant), notamment pour les affichages dit « immersifs ». Renseigner la position de la tête de l'utilisateur est en particulier nécessaire pour un environnement virtuel dont l'affichage est fixe et pour lequel on admet une déambulation de l'utilisateur par rapport à cet affichage. C'est le cas des affichages immersifs de type « Cave », mais également celui d'un moniteur de type courant quand le système gère la déformation de perspective en fonction de la position de la tête de l'utilisateur. Une réduction de la présentation en sortie est cependant possible en supposant une position fixe de l'observateur par rapport à l'affichage lui aussi fixe. Ce sera alors le cas de dispositifs du type « Reality Center » qui supposent un utilisateur assis en face de l'affichage.

Réduire la mobilité de l'observateur à une position déterminée apporte plusieurs avantages. En particulier, la gestion de la présentation en sortie peut s'affranchir de la prise en compte de la position de l'observateur alors supposée être constante et connue. La seule information nécessaire dans ce cas quant à la tête de l'utilisateur réside alors dans son orientation par rapport au dispositif d'affichage. Laquelle détermine une direction egocentrée de visée à partir de la position de l'utilisateur. Pour ce cas que nous allons retenir, nous devons par conséquent disposer d'une méthode permettant de déterminer l'orientation de la tête de l'utilisateur sur trois degrés de liberté.

Plusieurs catégories de solutions sont envisageables. Une reconnaissance active de l'environnement (*outward looking*), à partir de la position de l'utilisateur l'encombrerait d'un dispositif de capture d'informations présentes de manière stable au sein de l'environnement et d'un système de traitement de ces informations. L'observation de la tête de l'utilisateur présente l'intérêt de ne pas l'encombrer, tout en restant compatible avec une position fixe de l'affichage. A ce titre, nous ne retiendrons pas les méthodes mécaniques du type pantographe qui impliquent une liaison physique de la tête avec un dispositif fixe de l'environnement. Les méthodes basées sur un traitement d'images font appel à la reconnaissance de formes et peuvent être simplifiées par une recherche de marqueurs disposés sur la tête de l'utilisateur de manière aisément identifiable. Le traitement doit alors être suffisamment rapide pour produire un échantillonnage de fréquence suffisamment élevée pour satisfaire les besoins d'une visée. De tels principes sont disponibles mais impliquent un système de traitement à eux seuls. Une autre catégorie de méthodes consiste à mesurer l'émission d'un phénomène physique à partir d'une source de position connue, à l'aide d'un ensemble de capteurs dont l'encombrement est réduit en sorte d'être portable par l'utilisateur. Deux principales méthodes existent avec l'émission d'une onde ultrasonore et celle d'un champ magnétique. La première méthode autorise de plus grandes distances de mesures en souffrant d'un éventuel effet d'obstruction, la seconde est plus précise mais de portée plus réduite et nécessite quelques précautions préalables relatives à une éventuelle perturbation de champ magnétique. Ces deux méthodes constituent par conséquent des candidats convenables aux mesures d'orientations de la tête de l'utilisateur en l'encombrant de manière minimale.

5.3.2. Mouvements des yeux

Pour aborder maintenant la direction du regard, nous convenons que cet organe particulier ne saurait accepter une mesure mécanique impliquant un contact physique. L'essentiel des techniques possibles relèveront par conséquent de l'observation de l'œil. Une première technique qui rejoint celles basées sur celles de traitement d'images consiste à observer, depuis une position fixe de l'environnement, la position d'une tache cornéenne par rapport à la surface apparente de la pupille observée à distance. Une telle tache cornéenne est obtenue par l'émission ponctuelle,

au niveau de la caméra d'observation, d'un faisceau d'infrarouge réfléchi par la surface de l'œil. Ce type de technique souffre de l'obligation de disposer le dispositif de mesure dans un angle solide de faible ouverture en face de l'œil de l'observateur, et mesure par ailleurs des amplitudes de mouvements oculaires modestes. Son usage en ce qui nous concerne le cantonne à des affichages de dimensions réduites à côté desquels il peut prendre place.

Une autre méthode possible, et elle se révèle être en fait la seule méthode non intrusive disponible, consiste à effectuer le même type de mesure non plus à partir d'une position fixe de l'environnement, mais par rapport à la position de la tête de l'observateur. Le dispositif devient alors portable et est installé sur la tête de l'utilisateur. Cette méthode présente l'avantage d'augmenter sensiblement l'amplitude des mouvements oculaires mesurables. Mais elle implique, utilisée seule, une orientation fixe et connue de la tête de l'observateur. Cette fixation de la tête réduit alors son usage à des affichages dont la surface apparente ne déborde pas du champ de vision de l'utilisateur. Une autre méthode plus ancienne et plus intrusive fournissant un résultat approximativement similaire consiste à relever, au moyen d'électrodes, les variations de courant produites par les muscles responsables de l'orientation du globe oculaire. Cette méthode n'étant plus utilisée, nous ne ferons que la mentionner.

Nous connaissons donc une méthode qui fournit une information d'orientation du globe oculaire par rapport à la tête de l'utilisateur. Pour contourner l'inconvénient (ainsi que l'inconfort) d'une situation emprisonnant la tête de l'utilisateur dans une orientation « conforme » aux besoins de l'interaction, l'addition d'une méthode de référencement de la position de la tête de l'utilisateur par rapport à l'affichage permet d'obtenir indirectement une direction du regard relativement au dispositif fixe d'affichage.

La méthode précédemment retenue pour les mouvements de la tête convient en réduisant le choix de l'une des deux techniques retenues à celle qui offre la plus grande précision de mesure.

Nous retrouvons alors avec cette addition de deux méthodes pouvant chacune fournir une direction egocentrée, l'unique combinaison de quatre modalités que nous avons évoquée au chapitre précédent. A ceci près que cette fonction communicative ne sera plus mise en œuvre avec la direction de la tête, mais à l'aide de la direction résultante du regard, plus précise et plus rapide. Cette association de méthodes obéit par conséquent à une contrainte technique, et aboutit à un compromis acceptable respectant le confort de l'utilisateur et apportant une mesure des mouvements oculaires de grande amplitude. La méthode de suivi de la direction de la tête n'apparaît plus, dans ce cas, en tant que modalité d'entrée à part entière, mais davantage comme l'un des composants techniques du dispositif convoyant l'information de direction du regard. Cette intégration de la technique de suivi de la

direction de la tête à celle du regard réduit donc l'ensemble à une combinaison de trois modalités.

5.3.3. Mouvements des mains

Pour aborder maintenant les mouvements des mains, nous leur avons attribué au moins deux rôles nécessaires à l'expression d'une direction exocentrée, et à celles, simultanées, d'un axe et d'un sens de mouvement circulaire.

L'expression d'une direction à l'aide de la main n'implique qu'une connaissance de son orientation ainsi que celle d'un modificateur signalant un acte déictique en cours. Le geste déictique le plus couramment effectué consiste en une extension de l'index avec les autres doigts refermés sur la paume pour indiquer de manière fixe une direction approximative dans un angle solide approximatif de 180° vers l'avant du torse. Un autre geste déictique d'usage moins fréquent et plus approximatif encore consiste en une extension du pouce avec les autres doigts refermés sur la paume pour indiquer une direction plus vague dans un angle solide approximatif de 180° vers l'arrière du torse.

L'indication simultanée d'un axe et d'un sens de mouvement circulaire s'effectue généralement avec la même pose de la main que pour un geste déictique mais sans position fixe de la main qui effectue alors un mouvement tournant. La mesure de cette indication nécessite donc de plus une information de position de la main qui effectue ce mouvement.

Du mouvement observable le plus général au plus particulier, il nous faudrait donc renseigner le système sur la position du torse de l'utilisateur, puis d'au moins deux degrés de liberté de l'épaule en rotation, d'un degré de liberté supplémentaire du coude en flexion, et de trois degrés de liberté en rotation au niveau du poignet pour connaître la seule position de la paume de l'utilisateur avant de nous intéresser aux mouvements des doigts. De tels instruments de mesure existent sous la forme d'exosquelettes pour le moins encombrants. Nous avons toutefois déjà retenu, pour les mouvements de la tête, une technique alternative fournissant une mesure similaire de la production d'un champ magnétique, imposant le seul port d'un capteur de taille modeste. Cette méthode permet simultanément plusieurs mesures d'une même production de champ magnétique à l'aide de plusieurs capteurs similaires. On tirera donc profit de cette méthode en la mettant également en œuvre pour renseigner les positions et orientations de la main seule, non plus par rapport à l'utilisateur, mais par rapport à une position fixe connue.

Maintenant que nous connaissons la position et l'orientation de la paume de l'utilisateur dans l'espace, il nous faut renseigner les mouvements des doigts de la main ou définir une méthode distinguant le geste déictique principalement utilisé.

Les mouvements des doigts ont motivé de nombreux développements avec différentes méthodes de mesure de flexion basées sur différentes techniques au cours des deux dernières décennies. Zimmerman *et al.* furent les premiers à développer en

1987 un "Gant de Données" dont l'une des vocations premières était de permettre à l'utilisateur de manipuler des objets graphiques virtuels. Force est de constater, 17 ans plus tard, que l'usage du « gant de données » est resté marginal. Les raisons en sont sans doute multiples, mais nous pouvons cependant avancer certaines d'entre elles par expérience. D'une part, un très faible nombre d'applications impliquent de renseigner les flexions des doigts au point de justifier un investissement en des techniques qui relèvent encore souvent du laboratoire. Ce dispositif ne trouve guère d'usage au sein d'une plateforme courante de type GUI pour laquelle les dispositifs courants d'interaction avec un plan reste plus efficaces. L'usage du gant de données est, dans la plupart des cas, démontré au cours d'interactions pour lesquelles il joue un rôle de manipulation directe déroutant par l'absence de retour tactilo-kinesthésique, et au cours desquelles il semble se résumer à la mise en œuvre de gestes globaux tels que fermer la main, ouvrir la main, etc. pour mimer grossièrement une sélection à l'aide d'un geste de préhension. D'autre part, le port d'un tel gant ne permet guère d'autres activités à son porteur qui doit alors généralement le retirer. Enfin, en rapport avec l'utilité relative d'un tel dispositif, son coût constitue sans doute un critère limitant sa diffusion, qui doit de plus faire face aux mêmes impératifs que la diffusion d'un vrai gant, c'est à dire se décliner en différentes tailles. (Il n'y a guère que les gants de jardinage, de chantier ou de boxe qui n'existent qu'en trois tailles, lesquels dédient tous la main de l'utilisateur à la seule activité correspondante).

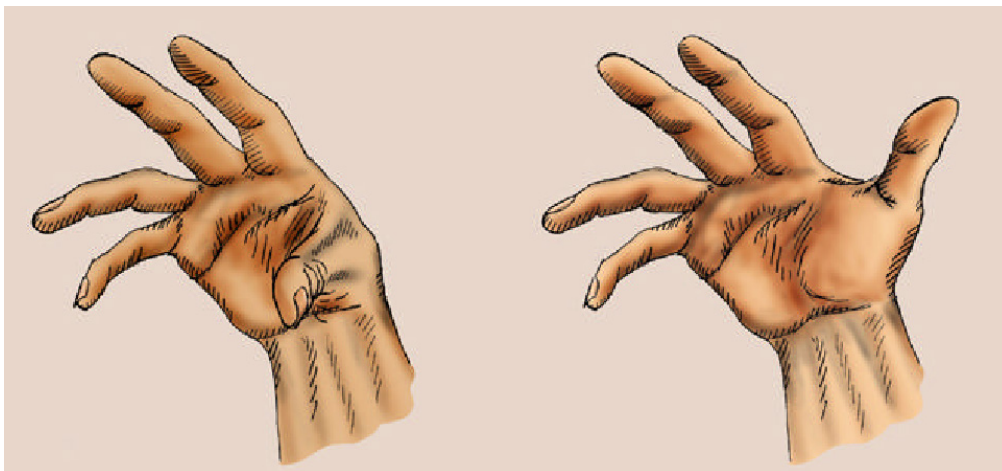


Figure 94. Mouvement du pouce qui n'est pas renseigné par un gant de données d'entrée de gamme.

Dans le cadre de recherche qui nous concerne, nous pouvons admettre des approximations en comptant sur de futures améliorations de ce dispositif. Dans le détail, les gants de données disponibles présentent de nombreuses variations liées aux techniques mises en œuvre. Ainsi, le gant financièrement le plus accessible¹¹ fournira-t-il une valeur de flexion globale de l'ensemble de chaque doigt, sans

¹¹ Un tel gant de données fut mis en œuvre au cours d'un projet de master of computer science in HCI. C. Tison, « Immersive Modelling Aids for Building Designers in Java3D and VRML », Herriot-Watt University, Edinburgh, 1999.

renseigner les mouvements d'abduction des doigts ou d'opposition du pouce (Figure 94). D'autres gants de données apportent davantage de mesures des mouvements des doigts, mais deviennent rapidement d'un coût prohibitif.

Considérant alors le modeste besoin, pour notre modalité d'entrée, qui ne consiste qu'à renseigner une unique position de la main dénotant un geste déictique, l'éventualité d'un dispositif artisanal fut envisagée.

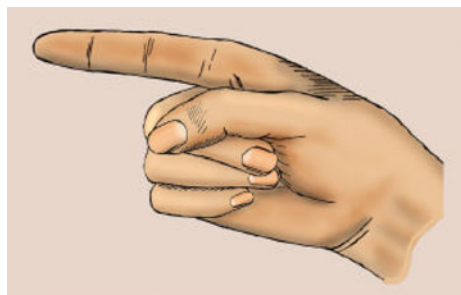


Figure 95 Représentation de la « pose » adoptée lors d'un geste déictique.

Une mesure de cette « pose » particulière de la main peut résider en une mesure directe ou indirecte de la flexion des doigts, elle peut également résider en des mesures indirectes des contacts entre diverses parties des mains au cours d'un tel geste. Nous pouvons également, en ajoutant une tâche motrice supplémentaire, mettre en place un dispositif électromécanique signalant cette « pose » à l'endroit le plus judicieux, en alourdissant toutefois la tâche de l'utilisateur.

L'observation du fait que le pouce entre généralement en contact avec la deuxième phalange du majeur au cours d'un geste déictique (Figure 96), a conduit à opter pour le compromis d'un capteur tactile disposé à cet endroit.

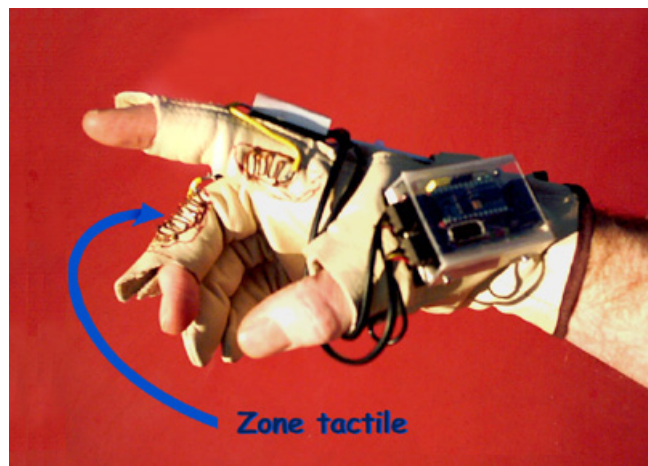


Figure 96. Disposition des zones sensibles tactiles.

La Figure 96 montre la disposition de deux zones sensibles tactiles dont l'une est installée sur le gant à l'endroit mentionné. De telles zones de capture sont réalisées par un « peigne » conducteur cousu à la surface du cuir.

Davantage de détails de fonctionnement ainsi que la description de l'application de démonstration vidéo ([gant01.wmv](#)), figurent dans l'annexe.

5.3.4. Enonciations

Pour aborder maintenant la modalité vocale commune aux deux combinaisons de modalités retenues, voyons quels dispositifs sont nécessaires.

La modalité vocale nécessite une entrée de la voix de l'utilisateur sur le système. La plupart des systèmes actuels présentent une orientation « multimédia » qui les fait

généralement disposer d'un équipement sonore aussi bien en sortie qu'en entrée. L'entrée physique « microphone » est ainsi disponible. Le seul élément de capture nécessaire reste finalement le microphone lui-même.

Plusieurs modèles de microphones sont disponibles.

Les microphones de table tels qu'en utilisent les radioamateurs et autres cibistes présentent l'intérêt d'un poussoir qui permet de manifester une prise de ligne lorsqu'on s'adresse au destinataire. Cependant, ce dispositif poussoir supplémentaire mobilise une main de l'utilisateur. Un modèle similaire dispose d'un mécanisme de détection de la voix qui produit la prise de ligne, mais qui réalise trop souvent cette prise de ligne avec un retard dont résulte une amputation du début de l'énoncé à la réception du message.

Les microphones dynamiques courants de type dit « cravate » sont particulièrement discrets, mais présentent la particularité de capter également tous les bruits ambiants, ce qui représente alors un inconvénient notable en environnement bruyant, car la reconnaissance vocale cherche également à y reconnaître des termes du langage.

S'affranchir des bruits ambiants afin d'éviter ce problème ne connaît qu'un nombre limité de solutions. La première consiste à utiliser un microphone directionnel qu'il faut disposer face à la source sonore, ce qui présente alors un problème de capture du bruit de la respiration. Une seconde solution consiste à utiliser un ensemble microphonique à soustraction du bruit ambiant. Ce type de microphone utilise deux capteurs directionnels dont l'un est disposé en direction de la source sonore d'intérêt, l'autre dans la direction opposée pour capter le bruit ambiant. Ce type de microphone s'est répandu chez les détaillants d'équipement pour motards depuis la libération de bandes de fréquences radio, (qui a permis la diffusion grand public d'équipements de communication de type « talky-walky » de qualité semi professionnelle).

Une autre solution garantissant une immunité au bruit ambiant est l'osophone. Ce type de dispositif se fixe sur la gorge de l'utilisateur et capte les vibrations vocales transmises par les cartilages de la trachée respiratoire. (Un tel dispositif semble cependant peu répandu, et n'a pas été expérimenté).

Une autre solution garantissant une immunité au bruit ambiant est l'osophone. Ce type de dispositif se fixe sur la gorge de l'utilisateur et capte les vibrations vocales transmises par les cartilages de la trachée respiratoire. (Un tel dispositif semble cependant peu répandu, et n'a pas été expérimenté).

Les dispositifs que nous avons mentionnés permettent maintenant d'esquisser un ensemble technique correspondant à une multimodalité de communication en entrée.

Nous constatons notamment que l'une des deux combinaisons initialement retenues impliquant la direction du regard est en réalité une extension de l'autre par la seule adjonction d'un dispositif portable de capture des mouvements oculaires. Cette constatation présente l'intérêt de pouvoir segmenter la mise en œuvre de ces

multimodalités en entrée de manière techniquement additive en une solution intermédiaire basée sur la seule direction de la tête pour manifester une direction égocentrée. Cet intérêt se double par ailleurs du fait que la direction de la tête est parfois considérée comme une réduction de la direction du regard.

Pour conclure, nous allons donc maintenant rappeler les dispositifs d'entrée retenus.

La tâche d'indication d'une direction égocentrée transite via une capture magnétique des position et orientation de la tête pour la première solution de trois modalités retenue. La seconde solution utilise cette information non plus à titre de modalité d'entrée, mais comme composante de la direction indiquée par une capture des mouvements oculaires réalisée à l'aide d'un dispositif portable. Cette direction de visée oculaire constitue alors la modalité d'indication d'une direction égocentrée pour la seconde solution multimodale.

Pour les deux solutions multimodales, l'indication d'une direction exocentrée et l'indication d'un sens et axe circulaire de mouvement sont transmises là encore à partir d'une information de position et d'orientation produite par un second dispositif de capture magnétique assujetti à la main dominante. Pour constituer la détection d'un geste déictique entrant dans la composition de ces deux indications d'intérêt, un dispositif artisanal de capture produisant un état binaire manifestant un geste déictique est ajouté.

Enfin, un microphone de tête transmet la voix de l'utilisateur au système.

5.4 Rôle des modalités d'entrée au sein de l'interaction.

De nombreuses contraintes pèsent encore sur l'efficacité d'une interaction de type manipulation directe, notamment manuelle. Les dispositifs haptiques qui seraient nécessaires pour ce faire relèvent encore largement de l'expérimentation en laboratoire. Par ailleurs, la préhension manuelle d'un objet 3D suppose en sortie un affichage graphique de proximité respectant l'indice visuel d'accommodation de l'œil de l'utilisateur, ce qu'aucun dispositif d'affichage n'est encore capable de réaliser à notre connaissance. Une interaction ne reposant pas sur une préhension directe de l'objet de la tâche peut alors adopter une approche relevant de la communication homme-machine, elle-même faisant appel aux modalités d'expression affichées par l'utilisateur. Ce type d'interaction s'inspire des "actes de langage" étendus à l'ensemble des indices communicatifs observables sur l'utilisateur qui s'exprime oralement, fixe du regard en orientant sa tête, effectue divers gestes avec ses mains et d'autres parties du corps. Tous les indices observables n'étant pas forcément pertinents en entrée système au regard des tâches à effectuer (ex. les mouvements des sourcils n'apportent sans doute guère d'information dans une situation de conception), une évaluation des compétences de chaque modalité à exprimer une information pertinente relative à l'activité de l'utilisateur doit être réalisée. Les modalités les plus significatives sont alors retenues afin de couvrir, à l'aide d'un

équipement technique le plus réduit possible, un spectre d'expression multimodale le plus étendu possible.

Avec la *visée* que nous avons considérée comme atomique et universelle (au sens énoncé par Bowman, 2002), nous nous sommes démarqués de la notion de pointeur 3D ponctuel difficile à manipuler au sein d'un environnement virtuel. Nous avons ensuite décliné cette *visée* avec la notion plus générale d'*indications de directions* au plus bas niveau de l'interaction. Avec la qualité supplémentaire *implicite vs explicite*, qui adresse en particulier la parole, nous avons ensuite obtenu la possibilité de *présélectionner* et de *sélectionner* de manière *implicite* ou *explicite*, permettant alors l'exploitation de mouvements continus involontaires chez l'utilisateur. Nous obtenons ainsi un spectre plus important de constituants d'interaction exploitables. L'assemblage de ces « composants » peut poser des problèmes de cohérence voire de conflits eu égard à diverses tâches.

5.4.1. L'interaction au niveau le plus bas.

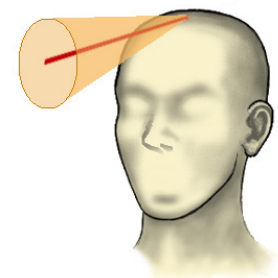
Il va donc maintenant nous falloir établir un ensemble cohérent de ces combinaisons à mettre en œuvre du point de vue de l'interaction afin de permettre à l'utilisateur de constituer des expressions multimodales de commande.

Présélection de Direction égocentrée

L'indication de direction *égocentrée* par visée à l'aide de la direction de la tête présente l'inconvénient d'être approximative. Notre sens proprioceptif n'est pas, en effet, accoutumé à produire une information nous permettant de connaître précisément une direction d'orientation de notre tête. Afin d'augmenter la précision de cette visée, il doit être possible de mettre en œuvre une réponse permanente du système informant l'utilisateur sur le résultat de cette visée.



Figure 97. Exemple de représentation d'une visée.



La visée concernant plus particulièrement une perception visuelle, on conçoit un retour visuel clairement identifiable qui ne doit cependant produire qu'une obstruction minimale de la scène présentée avec une métaphore la plus évidente possible. L'ensemble de ces conditions désigne la métaphore du " viseur ", ou du " collimateur de visée ", comme candidat de choix pour informer l'utilisateur de la direction dans laquelle il " vise ".

Un viseur présente généralement une forme caractéristique réticulée dans les instruments d'optique. Le collimateur (\oplus) adopte une forme similaire mais ajoute parfois un cercle réunissant les éléments orthogonaux du réticule autour du centre apparent.

La mise en œuvre d'un tel collimateur s'effectue en le disposant entre l'œil de l'utilisateur et l'objet de la visée, généralement à une faible distance, constante, de l'utilisateur.

Afin de présenter un minimum d'obstruction sur l'objet de la visée, ce viseur peut adopter des dimensions apparentes plus importantes qu'un viseur réel. Enfin, il doit présenter un aspect qui le distingue clairement de la scène et des objets de la tâche.

Présélection d'un objet

Une telle mise en œuvre d'un retour visuel d'une direction égocentrée constitue la représentation d'une présélection constante de cette direction. Comme nous avons également attribué cette visée à la présélection d'objets de la scène, cette représentation devra sans doute afficher un retour visuel de présélection d'objet situé dans la ligne de visée. Cet affichage peut devoir être effectué au niveau de la représentation du viseur. Toutefois, un objet de la scène devenant dans ce cas impliqué, sa représentation peut également être affectée de manière à distinguer un état de présélection de celui-ci.

L'état présélectionné d'un objet de la scène peut également impliquer la production d'informations complémentaires relatives à l'objet d'une éventuelle sélection. Il ne s'agira pas ici d'informations relatives à l'exploration, informant quant à l'objet lui-même, mais plutôt d'informations de *situation* de celui-ci par rapport à l'utilisateur au sein d'une représentation souffrant d'un défaut d'affichage permettant une évaluation correcte de la profondeur. Nous n'évoquerons ici, en particulier, que la distance virtuelle de l'objet d'une présélection, le séparant de l'utilisateur, qui pourra être représentée en réponse à une présélection d'objet de la tâche.

Présélection de Direction exocentrée

L'indication d'une direction exocentrée s'effectue avec un geste déictique de la main. Ce geste est généralement imprécis et mériterait lui aussi un retour visuel permettant un ajustement de cette indication. Toutefois, cette direction ne pointera généralement pas en direction la scène et sa représentation ne saurait trouver son origine à la position virtuellement occupée par la main de l'utilisateur. De ce fait, un retour visuel est difficilement représentable sur la scène, sauf sans doute à un niveau de dialogue plus élevé lorsque est sélectionné un objet de la tâche auquel s'applique cette indication de direction dont il occupe l'origine.

L'indication d'un axe et d'un sens de mouvement circulaire à l'aide de la main présente lui aussi une difficulté de représentation immédiate d'un retour visuel. Comme déjà mentionné, cette présélection peut en effet constituer une sélection implicite. Par ailleurs, un tel mouvement intermittent présente peu d'aptitude à un "ajustement" maintenu en attente de validation. Un retour visuel est cependant concevable à un niveau de dialogue élevé, en restant éventuellement d'intérêt marginal.

Sélections

Les sélections ont été définies en deux catégories principales, implicites et explicites.

Nous avons déterminé une sélection explicite comme une association d'un état de présélection maintenu et d'un acte volontaire de courte durée convoyant une information de confirmation non ambiguë.

En distribuant les indications de direction sur les organes expressifs les plus compétents, sous forme motrice et continue, nous en réduisons la capacité à effectuer une autre tâche.

Par ailleurs, nous avons également considéré la parole comme recèlant des actes volontaires ponctuels convoyant une information de désignation non ambiguë. Un rôle simple de confirmation préétabli pourrait suffire pour obtenir cette caractéristique "validante" d'une présélection. Toutefois, ce serait négliger la capacité de la parole à produire des énoncés à la fois porteurs d'informations de désignation non ambiguë et présentant simultanément les caractéristiques d'actes volontaires ponctuels convoyant une information de confirmation elle aussi non ambiguë.

Aussi pouvons-nous, afin de tirer parti de cette capacité de sélection *implicite*, spécialiser la fonction de validation par une double confirmation simultanée. Une telle confirmation spécialisée permet d'envisager ou de réserver d'une part plusieurs usages possibles des indications de direction, et d'autre part de convoier une information de désambiguation le cas échéant. Par exemple, la sélection d'une direction visée dans laquelle se trouve un objet graphique produit une ambiguïté quant à l'objet d'intérêt candidat d'une sélection (la direction égocentrée de visée ou l'objet 3D situé dans la visée). La parole permet alors d'une part de préciser la nature de l'objet d'intérêt à l'aide d'un énoncé désignant la classe à laquelle il appartient (« *cette direction* » vs « *cet objet* »), et d'autre part de présenter les conditions suffisantes pour produire une sélection.

Cette validation est donc en particulier envisagée sous la forme d'une référence vocale à la classe à laquelle appartient l'objet d'intérêt pour la tâche.

Sélection explicite d'un objet 3D de la scène

Une sélection explicite se présente alors sous forme d'une présélection ajustée et maintenue accompagnée de l'énonciation d'un groupe nominal comportant un adjectif démonstratif.

Ainsi, pour citer un exemple, la présélection par visée d'un objet de la scène peut-elle se trouver validée à l'aide de l'énoncé " *cet objet* " traduisant la classe générique à laquelle appartiennent tous les objets de la scène.

Sélection explicite d'une direction

L'usage du même modèle de double confirmation spécialisée nous mène alors dans ce cas à la production d'un énoncé similaire (" *cette direction* "), comportant cette fois le type de l'information indiquée (désambiguation).

L'adoption de ce modèle de validation à l'aide d'un énoncé spécialisé trouve sa justification en permettant notamment la désambiguation d'une situation pouvant survenir quand, lors d'une visée en direction d'un objet de la scène, celui-ci, en réponse à la visée, devient présélectionné, alors que c'est la direction *égocentrée* qui est l'objet d'intérêt.

Ce style de validation autorise ainsi, en permettant dans ce cas de sélectionner une direction dans laquelle se trouve un objet éventuel de la tâche, plusieurs usages de la même indication d'une direction *égocentrée*.

Autre conflit potentiel, la spécification vocale de la classe générique trouvant simultanément plusieurs présélections candidates à une sélection (ex. « *cette direction* ») qui peut apporter une situation ambiguë lors de l'indication d'une direction *exocentrée* à l'aide d'un geste déictique. L'indication de direction *égocentrée* restant par nature continue (celle de la tête ou du regard), la production du seul énoncé "*direction*" trouve alors à valider, en concurrence, deux informations appartenant à la même classe, à savoir une direction *égocentrée* et une direction *exocentrée*.

Une solution de spécialisation complémentaire de l'énoncé peut alors consister à spécifier également la nature complétant le type d'information indiquée, (ex. « *cette direction depuis moi* » vs « *cette direction depuis cet objet* », voire « *cette direction égocentrée* » vs « *cette direction exocentrée* »), au prix d'un apprentissage pesant sur le langage qui la rend peu envisageable.

L'autre solution consiste à répartir la validation non plus seulement sur le seul énoncé, mais également sur celle des deux indications de direction qui comporte un acte volontaire ponctuel (qui entre dans notre définition de la validation). L'adoption par la main de la "pose" caractéristique du geste déictique constituant effectivement un acte moteur qui répond à cette condition, cela revient à instaurer une priorité de l'indication d'une direction *exocentrée* sur celle d'une direction *égocentrée* continuellement visée.

Ainsi, l'énoncé peut-il conserver le même modèle en ne spécifiant que la classe de l'information désignant l'objet d'intérêt.

La production de l'énoncé "*cette direction*" en présence d'un geste déictique manuel validera donc une direction *exocentrée* de façon préférentielle sur la direction *égocentrée* exprimée en permanence.

Sélection implicite de valeurs et d'unités

Le caractère implicite d'un énoncé semble dans certains cas reposer sur une combinaison des informations convoyées par plusieurs termes successifs. Ainsi, l'énonciation d'une valeur numérique ne désigne-t-elle rien d'autre qu'un type numérique d'ordre primitif (qu'on limitera volontairement aux entiers de façon provisoire). Une telle désignation ne présente pas, à elle seule, la sélection de l'objet d'une tâche particulière à exécuter. De même, la sélection d'une unité de mesure, par exemple métrique, constitue bien une désignation mais ne constitue pas l'objet d'une tâche.

Cependant, l'association de ces deux indications devient potentiellement porteuse d'une information pertinente d'un niveau supérieur en constituant une unité interprétable et pouvant satisfaire les conditions d'une sélection implicite. Ainsi, l'énoncé d'une valeur numérique entière associée à celle d'une unité métrique désigne-t-elle implicitement un type "distance" non primitif de valeur finie et transposable par le système à l'environnement. Néanmoins, une distance, là encore, ne constitue pas, en soi l'objet d'une tâche.

Sélection explicite d'une position de destination

En considérant maintenant l'utilisateur comme le centre de référence de l'environnement, muni d'autre part de la capacité à exprimer une direction égocentrée ainsi maintenant qu'une distance, nous pouvons évoquer une nouvelle association dont résulte comme précédemment une nouvelle interprétation pertinente. Ainsi, associer l'indication d'une distance à la sélection d'une direction égocentrée désigne-t-elle une position dans un système de coordonnées polaires (pour simplifier, nous nous cantonnerons volontairement au plan par la suite).

Une telle désignation d'une position contient une information qui mène a priori à la sélection de cet objet "position" complètement identifiable au sein de l'environnement.

Cependant, l'indication d'un lieu présente à son tour une carence d'information qui n'en fait pas un objet sélectionné d'une tâche quelconque. Nous allons donc délibérément, dans ce cas précis, considérer une telle "position" sous la forme d'une indication de "*destination*" ponctuelle pour laquelle cette appellation nous semble plus appropriée. Une telle indication de "*destination*" nous permet en effet comme précédemment d'admettre l'éventualité d'une association supplémentaire avec un terme énonçable dont le résultat conduira à la désignation d'un éventuel objet d'une tâche.

Sélection d'opération

Nous allons maintenant, en nous limitant provisoirement de façon délibérée aux catégories de tâches utilisateur que nous avons défini, évoquer les désignations d'opérations s'appliquant à un objet de la tâche.

L'énoncé d'une opération correspond à l'indication d'une tâche à exécuter. Nous allons admettre d'abord qu'une désignation d'opération va consister ici en une sélection implicite qui supposera au moins un objet de l'opération.

La catégorie de tâches que nous adressons dans le cadre que nous avons fixé est constituée par les mouvements d'objets au sein de l'environnement.

Cette catégorie d'opérations suppose effectivement un objet de la tâche sur lequel s'appliquera l'opération désignée.

Nous avons globalement distingué deux types de translations.

Le "*positionnement*", ou placement, qui présente un caractère absolu du fait qu'une telle opération ne tient pas compte de la position initiale de l'objet du mouvement et implique en particulier une position de destination clairement décrite pour être identifiable par le système à partir de la position virtuellement occupée par l'utilisateur.

Le "*déplacement*" qui, étymologiquement, et à la différence du "*placement*", devrait en toute rigueur rester relatif à la position initiale de l'objet du mouvement. Car en effet, on "*déplacera*" un objet de l'endroit où il se trouve, par exemple d'une certaine quantité dans une certaine direction.

Toutefois, on notera d'une part l'usage admis de la formule qui consiste à "*déplacer un objet d'un endroit à un autre*", qui peut lui aussi constituer une destination clairement identifiable. Cette formulation présente ainsi les caractéristiques d'un "*déplacement*", mais contient également celles d'un positionnement, ce qu'il est en réalité.

Par ailleurs, et d'autre part, on "*déplace*" souvent, par abus de langage, un objet à une destination clairement identifiée, signifiant ainsi un positionnement à l'aide d'un verbe inapproprié.

On remarque alors que si les termes "*placer*" ou "*positionner*" désignent clairement une opération de positionnement absolu, il n'en va pas de même pour le terme "*déplacer*" qui restera ambigu en exprimant aussi bien une opération de "*déplacement*" qu'un "*positionnement*" par abus de langage.

La question de savoir s'il faut ou non contraindre l'utilisateur à s'exprimer avec rigueur de façon académique dans une langue dite "*vivante*" n'est pas du ressort de cette étude. Pour notre part, dans le cadre d'une thèse dans le domaine de l'Interaction Homme-Machine, nous conserverons ici le principe selon lequel c'est le système qui doit au mieux être adapté à l'utilisateur en admettant les abus de langage dès lors que leur signification peut être identifiée à l'aide d'autres indices.

Ceci étant posé, il reste à déterminer les indices permettant de distinguer les "*déplacements*" des "*placements*".

Si l'on utilise le terme plus général de "*mouvement fini*" à appliquer à un objet, on définit alors deux catégories par lesquelles un mouvement s'effectuera d'une part "*d'une*" certaine quantité à partir de la position initialement occupée par l'objet, (ex. "...*de 10m...*", "...*de 20°...*", "...*d'un demi-tour...*", etc.), ou qui s'effectuera, d'autre part, "*à*" une position de destination clairement indiquée, (ex. "...*à 10 m dans cette direction...*").

L'usage plus systématique de la préposition correcte "*de*" versus "*à*" correspondant à la catégorie de mouvement signifiée constitue donc un indice permettant de lever l'ambiguïté lors de l'usage du verbe "*déplacer*". On « *déplacera* » ainsi par exemple un objet « *de 10 m* » dans une certaine direction exocentrée, ou encore éventuellement par abus de langage, « *à 10m* » dans la direction égocentrée simultanément visée.

Nous allons donc considérer que dans certains cas, l'énoncé d'un unique terme d'opération ne suffit pas à désigner l'opération à exécuter, que cette unique énonciation désigne alors plutôt dans ce cas un groupe d'opérations, et que la désignation identifiant effectivement l'opération à exécuter est alors distribuée sur plusieurs termes de l'énoncé dont seule l'association constitue une désignation déterministe de ce qui doit être exécuté.

Pour adresser maintenant plus particulièrement les mouvements circulaires, nous avons distingué deux sous-ensembles que nous allons illustrer avec un exemple. Nous avons restreint la *rotation* à un mouvement tournant d'un corps autour d'un axe passant par le centre géométrique de l'objet d'un tel mouvement. Cette restriction nous a alors permis de distinguer un autre mouvement tournant d'un corps autour d'un axe ne passant plus par le centre géométrique de l'objet d'un tel mouvement. Afin de distinguer ce dernier mouvement d'une rotation, nous lui avons cherché une désignation. Le terme de « révolution » ayant paru désigner un mouvement périodique, nous avons opté pour le terme de « *giration* » de façon quelque peu arbitraire.

Géométriquement, la définition des deux mouvements ne se distingue que par la position de l'axe autour duquel ils s'effectuent. Ce que nous avons défini comme une « *giration* » est donc une « *rotation* » autour d'un axe distant, cependant que la *rotation* dans son sens restreint présente seulement la particularité de situer l'axe du mouvement en un lieu qui n'a de particulier que le fait de passer par le centre géométrique de l'objet.

Cependant, lorsqu'on souhaitera de la part d'un tiers qu'il effectue une rotation d'un objet sur lui-même, on lui demandera de « *tourner cet objet d'une certaine quantité dans une certaine direction* » en spécifiant l'axe et le sens du mouvement de façon gestuelle. De même, pour obtenir une *giration*, lui demandera-t-on cette fois de « *tourner cet objet d'une certaine quantité dans une certaine direction autour d'un lieu désigné* » en spécifiant là encore l'axe et le sens du mouvement de façon gestuelle.

Nous en concluons d'abord que l'expression d'une commande de *rotation* (d'un objet sur lui-même) est incomplète, et qu'en l'absence de spécification de la localisation de l'axe du mouvement, on admet comme sous-entendu que cet axe passe par le centre de l'objet.

En d'autres termes, nous constatons ici que le *cas particulier* que constitue la *rotation*, se désigne par une expression *plus générale* convoyant *moins* d'information que la désignation du cas plus général de *giration*.

Nous constatons ensuite que le même terme d'opération (« *tourner* ») est mis en œuvre pour désigner les deux mouvements circulaires que nous avons distingué. Le seul énoncé du verbe d'opération ne détermine donc pas lequel des deux mouvements est indiqué. Nous retrouvons encore, à l'instar de l'énoncé d'une commande de translation, le fait que le seul énoncé d'un unique terme d'opération ne suffit pas à désigner l'opération à exécuter, que cette unique énonciation désigne alors plutôt dans ce cas un groupe d'opérations, et que la désignation identifiant effectivement l'opération à exécuter est alors distribuée sur plusieurs termes de l'énoncé dont seule l'association constitue une désignation déterministe de ce qui doit être exécuté.

Pour résumer, nous admettrons que l'énonciation d'un unique terme d'opération de translation peut constituer une sélection qui ne détermine de façon préférentielle que la sélection d'un groupe d'opérations dans lequel figure l'opération à exécuter, et qu'en conséquence, une sélection supplémentaire d'un modificateur (préposition), détermine ensuite, au sein du groupe, l'opération à exécuter.

Différemment, l'énonciation d'un unique terme d'opération de mouvement circulaire peut constituer une sélection qui ne détermine de façon préférentielle qu'une opération de *rotation* par défaut. La détermination d'une autre opération de mouvement circulaire impliquant une sélection supplémentaire d'un modificateur (dit « locution prépositive », ex. « *autour* »), détermine ensuite, au sein du groupe, l'opération à exécuter en excluant alors la sélection d'une *rotation*.

Nous retiendrons pour les deux sous ensembles de mouvements, l'existence d'une désignation d'opération distribuée sur une somme de sélections successives de termes relatifs aux opérations. Avec un cas particulier, la *rotation*, désignée par la seule sélection d'un groupe d'opérations, celui des mouvements circulaires réunissant les mouvements à la fois de *rotation* et de *giration*.

La carence que nous constatons en termes uniques énonçables désignant clairement une opération géométrique nous amène maintenant à évoquer la pluralité possible de termes désignant un même groupe de mouvements.

En effet, un autre aspect intéressant particulièrement la sélection d'opérations est la synonymie.

Une opération de mouvement géométrique trouve plusieurs termes pouvant la signifier dans plusieurs cas. Le groupe des positionnements s'exprime à l'aide de termes verbaux tels que « *placer* », « *positionner* », « *mettre* », « *disposer* », etc. qui

les distinguent clairement du groupe des déplacements auxquels correspond un ensemble, semble-t-il moins nombreux, de termes tels que « *déplacer* ».

Nous avons évoqué, lors de la distinction entre les opérations de positionnements et celles de déplacements, une confusion possible sur l'appellation de cette dernière catégorie qui présente un caractère générique de désignation (que nous avons délibérément restreint pour des besoins de clarté au profit du terme générique de « *mouvement* » englobant à la fois les « *positionnements* » et les « *déplacements* »). Nous avons également admis précédemment qu'une certaine liberté avec l'emploi du langage peut conduire à désigner une opération de positionnement à l'aide du terme « *déplacer* ». Cette restriction étant ici de notre fait, on peut donc s'attendre à ce que l'usage du terme « *déplacer* » soit également mis en œuvre pour signifier un positionnement.

Par ailleurs, toujours au titre de la synonymie, des termes génériques de mouvement qui ne dénotent aucune distinction entre positionnement et déplacement semblent également pouvoir être mis en œuvre (ex. « *bouger* »), qui justifient alors, pour déterminer une opération de mouvement, la prise en compte d'un autre indice utilisé de façon plus tardive dans l'expression, la préposition.

Sélection implicite d'une direction

Pour revenir brièvement sur l'expression de certaines directions particulières, nous allons maintenant évoquer celles qu'autorise l'usage de la parole.

Le caractère implicite d'une énonciation permet en effet de désigner une direction de manière implicite dans au moins trois cas.

Nous avons précédemment évoqué la désignation d'un objet de la scène situant une cible vers laquelle orienter une direction, ainsi qu'une désignation de direction orthogonale conventionnelle par rapport à une direction sous-entendue de référence et de deux points.

Dans le premier cas, la mention d'un déplacement « *en direction de* » suivi d'une désignation d'objet détermine en effet une orientation de la direction du mouvement ciblant un second objet désigné.

Dans le second cas, la sélection d'un déplacement à appliquer sur un objet sélectionné « *vers la droite* », à partir de l'endroit où il se trouve, constitue une désignation exocentrée particulière ne mettant pas en œuvre une désignation gestuelle de direction. Ces directions constituent un cas particulier car elles n'adressent que les directions orthogonales, ce qui en limite l'usage. Cependant, leur particularité réside également dans le fait qu'elles peuvent être sélectionnées à l'aide d'une unique modalité, ce qui nous laissera supposer une plus grande facilité de mise en œuvre.

Une troisième sélection implicite d'une direction constitue un autre cas particulier cependant important à signaler. Elle réside en la désignation d'une cible constituée

par l'utilisateur lui-même. L'usage de l'adjectif démonstratif au sein du groupe nominal utilisé jusqu'à présent pour valider une présélection à l'aide d'un énoncé se décline en effet avec celui d'un adjectif possessif. Ainsi, la désignation du déplacement d'un objet sélectionné « *dans ma direction* » cible-t-il clairement un déplacement exocentré vers la position virtuellement occupée par l'utilisateur, et correspond pour ainsi dire à une direction « egociblée ».

5.4.2. Expressions de commande

Nous avons été conduit à l'élaboration d'une grammaire multimodale approximative permettant d'une part de caractériser de manière systématique l'ensemble des configurations [tâche x modalités x dispositifs], et d'autre part de définir des scénarios de commandes sous-jacentes aux tests utilisateurs qui permettent d'en ajuster la cohérence. Ainsi, par exemple, l'usage d'un geste déictique manuel pour la sélection d'un objet de la scène entraine-t-il en concurrence avec l'usage du même geste pour la désignation de la destination d'un déplacement d'objet en présence d'un autre objet de la scène situé au-delà de la destination souhaitée dans la direction indiquée (occasionnant logiquement alors la sélection du second objet). La mise en évidence de tels conflits d'usage des modalités à l'aide de scénarios conduit à préciser davantage le rôle d'une modalité. Dans l'exemple cité, il apparaît que le geste déictique manuel prend généralement place pour n'indiquer qu'une direction qui ne saurait à elle seule permettre une sélection.

Le tout constituant davantage que la somme des parties, nous allons maintenant réexaminer plus particulièrement les tâches utilisateur de mouvements d'objets plus en détail pour estimer la capacité d'une expression de combinaisons de modalités à exprimer une commande de manière non ambiguë.

La syntaxe générale d'une translation d'objet est définie par un énoncé de la forme :

[Désignation (Opération de translation)]
& [désignation (*Objet Graphique 3D*)]
& [désignation (destination)]

Syntaxe pour laquelle une destination se décline sous plusieurs combinaisons possibles, par exemple:

[Désignation (direction)]
& [désignation (quantité linéaire ou angulaire)]
& [désignation (Origine géométrique)]

La syntaxe générale d'une orientation d'objet obéit à une forme similaire :

[Désignation (Opération de mouvement circulaire)]
& [désignation (*Objet Graphique 3D*)]
& [désignation (Valeur angulaire)]
& [désignation (axe et sens du mouvement)]

En introduisant, dans l'expression, de telles variables relatives à l'origine ou la destination d'une direction exprimée, ainsi que l'origine à partir de laquelle doit s'appliquer le mouvement, nous avons d'abord établi une distribution combinatoire des cas possibles (non reproduite pour des raisons de reproductibilité et de lisibilité). L'élimination des combinaisons peu envisageables voire aberrantes a ensuite permis d'obtenir un nombre plus réduit de combinaisons. Nous avons conservé ensuite les combinaisons les plus pertinentes exprimant différents mouvements d'objets.

Les tâches utilisateur de mouvements d'objets ont été regroupées dans un tableau d'abord à partir des deux façons de caractériser une direction en fonction de l'endroit d'où elle pointe et de l'endroit vers où elle pointe, puis à partir de la position autour de laquelle elle indique (ou n'indique pas) l'axe et le sens d'un mouvement circulaire. Ce tableau est reproduit ci-après (Tableau 3).

Nous allons examiner au moins un cas pour chaque catégorie de mouvement en restreignant l'amplitude de ces mouvements au plan horizontal par rapport à l'utilisateur (il s'agira généralement d'un plancher de référence). Nous décrirons dans chaque cas les successions de combinaisons de modalités mises en œuvre au sein de l'expression, ce qui permettra notamment, le cas échéant, la mise en évidence de conflits, ainsi qu'une estimation de la conformité de l'énoncé avec les expressions courantes du langage oral.

Translation	Positionnement	Egocentré	Par rapport à la Position de l'utilisateur.
		Exocentré	Par rapport à la position d'un second objet sélectionné.
	Déplacement	Exocentré	Dans la direction exocentrée Indiquée par l'utilisateur.
		Ciblé	En direction d'un objet cible sélectionné.
Mouvement	Rotation	Autocentrée	Toujours autour de l'objet sur lui-même.
Circulaire	Giration	Egocentrée	Par rapport à la Position de l'utilisateur.
		Exocentrée	Par rapport à la position D'un second objet sélectionné.

Tableau 3. Reproduction distribuée des mouvements décrits.

5.4.3. Les positionnements

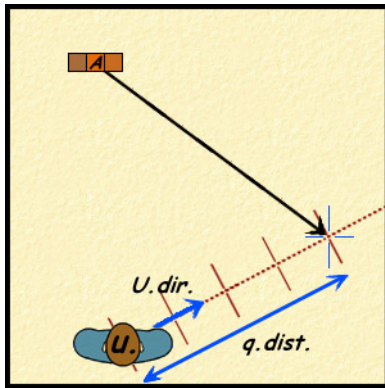


Figure 98. Positionnement égocentré.

La Figure 98 montre une vue de dessus figurant l'utilisateur (« *U.* ») et l'objet d'un mouvement (« *A* ») ainsi que les caractéristiques géométriques du mouvement. En bleu se trouvent figurées une direction égocentrée (« *U.dir.* ») ainsi qu'une quantité métrique de distance (« *q.dist.* »). Le résultat apparent du mouvement est manifesté par une flèche en noir.

Nous avons défini un positionnement égocentré (Figure 98), sous la forme

[Sélection (Opération de positionnement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (destination)]

dont la destination apparaît sous la forme

[Sélection (*distance*)]
& [Sélection (*direction*)]

Typiquement, une commande à un tiers s'énoncera par exemple de la façon suivante :

"**Place** [objet (A)] **à** [quantité distance] **dans** [direction égocentrée]."

Pour le mouvement illustré Figure 98, La sélection du groupe d'opérations de positionnement se réalise à l'aide de l'énoncé correspondant, celle de l'objet de l'opération s'effectue par visée et validation par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt (« *cet objet* »). L'énoncé de la préposition spécifie ensuite le positionnement de manière redondante. La détermination de la destination du mouvement implique ensuite d'abord la sélection d'une quantité métrique (valeur, unité), puis celle d'une direction, qui sera effectuée ici par l'indication d'une direction égocentrée, validée par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt (« *cette direction* »).

La bimodalité mise en œuvre ici permet *a priori* d'exprimer une commande de positionnement égocentré défini sans ambiguïté.

Examinons maintenant un cas général de positionnement exocentré (Figure 99 qui représente également un second objet « *B.* » vu de dessus).

Le positionnement exocentré a été défini sous la forme :

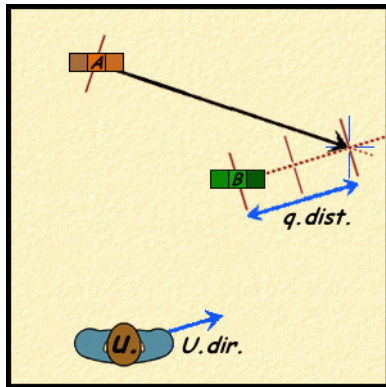


Figure 99. Cas général de positionnement exocentré.

[Sélection (Opération de positionnement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (destination)]

avec, pour la destination :

[Sélection (*distance*)]
 & [Sélection (*direction*)]
 & [Sélection (*Objet Graphique 3D*)]

Une expression de commande à un tiers s'énoncera par exemple de la façon suivante :

"**Place** [objet (A)] **à** [quantité distance] **de** [objet (B)] **dans** [direction exocentrée]."

Là encore, la sélection du groupe d'opérations de positionnement se réalise à l'aide de l'énoncé correspondant, celle de l'objet de l'opération s'effectue par visée et validation par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt. (L'énoncé de la préposition spécifie ensuite le positionnement de manière redondante).

La détermination de la destination du mouvement implique ensuite trois informations : une position origine à partir de laquelle appliquer une distance dans une direction exocentrée souhaitée.

Une distance est désignée par la sélection d'une quantité métrique (valeur, unité), la désignation de la position d'origine se traduit ensuite par la sélection d'un second objet référent, laquelle se réalise par visée et validation produite par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt. Enfin, l'indication d'une direction exocentrée, validée par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt constitue la dernière information nécessaire.

L'aspect séquentiel d'une expression cadencée par l'énoncé permet ici de distinguer des successions de combinaisons de deux modalités qui ne produisent pas de conflits. Une telle séquence permet donc *a priori* d'exprimer une commande de positionnement exocentré sans ambiguïté pour un cas général de positionnement d'un objet à une certaine distance d'un second objet dans une direction s'appliquant à partir du second objet de façon exocentrée.

Pour évoquer maintenant une variante du positionnement exocentré, introduisons une sélection de direction relative (Figure 100). Il s'agit, rappelons-le, d'une direction limitée aux directions orthogonales dans un plan de référence, qui présente l'avantage de s'énoncer oralement.

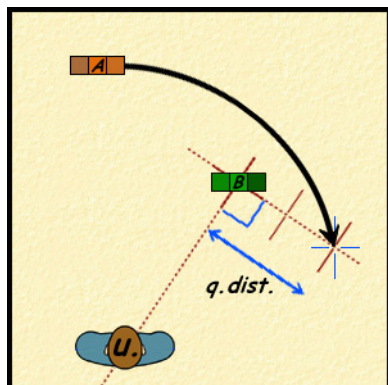


Figure 100. Exemple de positionnement exocentré orthogonal.

La forme générale de l'expression reste
 [Sélection (Opération de positionnement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (destination)]

avec encore, pour la destination :

[Sélection (*distance*)]
 & [Sélection (*direction*)]
 & [Sélection (*Objet Graphique 3D*)]

Une expression de commande correspondante peut s'énoncer de la façon suivante:

" **Place** [objet (A)] **à** [quantité distance] **à** [direction relative] **de** [objet (B)]."

Exemple, "**Place cet objet à 2m à droite de cet objet**".

ou encore:

" **Place** [objet (A)] **à** [quantité distance] **de** [objet (B)] **vers** [direction relative]."

Exemple, "**Place cet objet à 2m de cet objet vers la droite**".

La sélection de l'opération est similaire au cas précédent, ainsi que pour l'objet de l'opération.

La détermination de la destination du mouvement implique encore trois informations: une position origine à partir de laquelle appliquer une distance dans une direction souhaitée.

Une distance est désignée par la sélection d'une quantité métrique (valeur, unité), la désignation de la position d'origine se traduit ensuite par la sélection d'un second objet référent du mouvement (objet « B » sur la Figure 100), laquelle sélection se réalise de la même façon que pour l'objet du mouvement. Enfin, à la suite de l'adverbe, la sélection implicite d'une direction exocentrée orthogonale (« *vers la droite* ») complète le nombre d'informations nécessaire.

Aucun conflit n'apparaît au cours de l'expression.

Examinons maintenant deux cas particuliers de positionnements ciblant l'utilisateur.

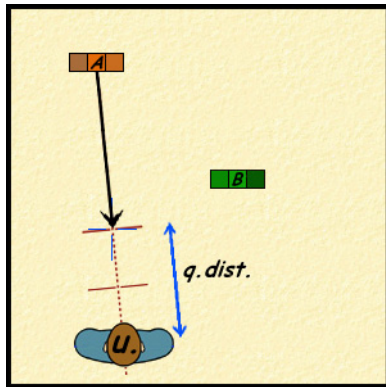


Figure 101. Cas particulier de positionnement ciblé.

Le premier cas est celui pour lequel la désignation de l'opération s'effectue à l'aide d'un groupe verbal spécialisé qui désigne également la direction. Cette sélection d'opération ne semble pas trouver, dans le langage habituellement parlé, d'équivalent à l'aide d'un terme plus général de positionnement, voire de déplacement (Figure 101).

En effet, l'application de la forme générale de positionnement implique :

[Sélection (Opération de positionnement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (destination)]

avec, pour la destination :

[Sélection (*distance*)]
& [Sélection (*direction*)]
& [Sélection (*Objet Graphique 3D*)]

La syntaxe équivalente pourrait apparaître ainsi :

"**Place** [objet (A)] **à** [quantité distance] **de** [Utilisateur]
dans [direction objet (A)]."

Exemple : « *place cet objet à 2m de moi dans ma direction !* »

Un tel énoncé est toutefois difficilement concevable, alors que l'usage d'un verbe spécialisé du langage courant manifeste ce positionnement sans présenter d'éventuel conflit.

"**Approche** [objet (A)] **à** [quantité distance]."

Exemple : « *Approche cet objet à 2m !* »

Le second cas particulier de positionnement implique une destination exprimée par une direction trouvant sa source à la position d'un second objet référent, et orientée vers l'utilisateur (Figure 102).

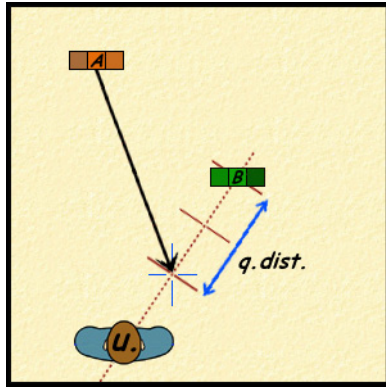


Figure 102. Cas particulier de positionnement ciblé à partir d'un objet référent.

La forme de l'expression est celle d'un positionnement.

[Sélection (Opération de positionnement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (distance)]
 & [Sélection (direction)]
 & [Sélection (Objet Graphique 3D)]

Une expression de commande correspondante pourra afficher la syntaxe suivante :

" **Place** [objet (A)] **à** [quantité distance] **de** [objet (B)] **dans** [direction Utilisateur]."

Exemple : « *Place cet objet à 2m de cet objet dans ma direction !* »

C'est la direction sélectionnée qui présente une particularité. Elle désigne en effet implicitement l'utilisateur en tant que cible par l'usage d'un adjectif possessif.

Là encore, aucun conflit ne semble apparaître dans cette syntaxe qui paraît néanmoins peu naturelle malgré l'existence concevable d'un tel mouvement géométrique.

5.4.4. Les déplacements

Pour aborder maintenant les « *déplacements* », nous en avons défini deux catégories principales, les déplacements exocentrés avec indication de direction exocentrée, et les déplacements ciblés qui feront parcourir une distance indiquée en direction d'une cible désignée. Les cas particuliers peuvent être appliqués comme précédemment avec la mise en œuvre d'une direction relative ou de la désignation d'une cible en direction de laquelle se trouvera la destination (Figure 103).

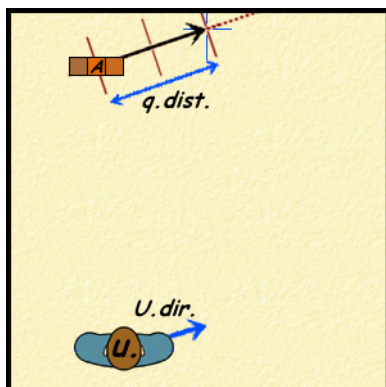


Figure 103. Exemple de déplacement exocentré.

Nous avons défini la forme générale d'un déplacement comme suit :

[Sélection (Opération de déplacement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (distance)]
 & [Sélection (Direction)]

Un exemple de syntaxe correspondante sera :

" **Déplace** [objet (A)] **de** [quantité distance] **dans** [direction exocentrée]."

La sélection de l'opération et de l'objet de la tâche est maintenant classique.

La détermination de la destination du mouvement implique une sélection d'une quantité métrique (valeur, unité), ainsi que celle d'une direction, qui sera effectuée ici par l'indication d'une direction *exocentrée*, validée par l'énoncé démonstratif désignant la classe générique de l'objet d'intérêt, c'est-à-dire la direction. On notera, pour la détermination d'une destination, une permutation possible de la spécification de quantité avec celle de direction.

Exemple : "*Déplace cet objet de 2m dans cette direction !*"

ou "*Déplace cet objet dans cette direction de 2m !*"

Les bimodalités successivement mises en œuvre permettent *a priori* d'exprimer une commande de déplacement exocentré défini sans ambiguïté, et sans conflit apparent.

Apparenté au déplacement avec indication exocentrée de direction, le cas

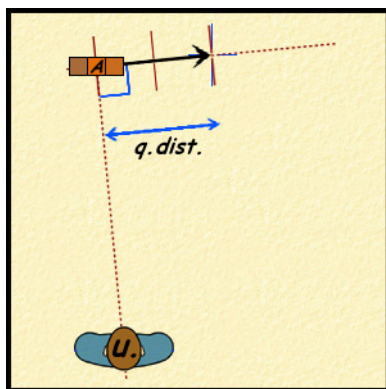


Figure 104. Exemple de déplacement dans une direction orthogonale.

particulier d'indication d'une direction orthogonale se distingue par l'économie de mise en œuvre d'une modalité gestuelle pour la désignation de direction relative (Figure 104).

En conservant la même forme générale, une syntaxe de l'expression pourra être la suivante :

"**Déplace** [objet (A)] **de** [quantité distance] **vers** [direction relative]."

Un déplacement ciblé implique la sélection d'un second objet désigné comme cible vers laquelle s'oriente la direction indiquée (Figure 105).

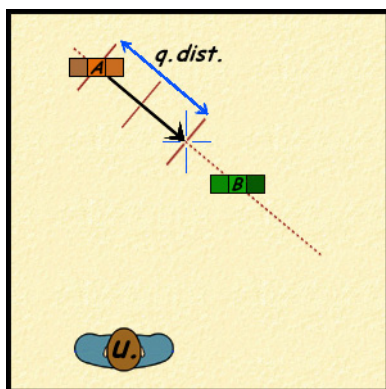


Figure 105. Exemple de déplacement ciblé.

La forme de l'expression est similaire à précédemment avec désignation d'une cible:

[Sélection (Opération de déplacement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (distance)]
& [Sélection (Direction)]
& [Sélection (Objet Graphique 3D)]

avec pour syntaxe possible :

"**Déplace** [objet (A)] **de** [quantité distance] **en direction de/vers** [objet (B)]."

Une variante particulière adresse un déplacement d'une distance indiquée dans la direction de l'utilisateur (Figure 106).

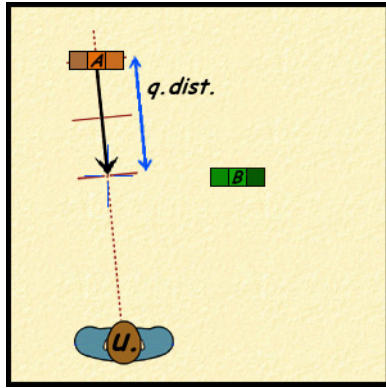


Figure 106. Cas particulier d'un déplacement ciblant l'utilisateur.

Ce déplacement correspond encore au modèle :

[Sélection (Opération de déplacement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (distance)]
 & [Sélection (Direction)]
 & [Sélection (Objet Graphique 3D)]

Avec par exemple pour syntaxe :

"**Déplace** [objet (A)] **de** [quantité distance] **en direction de/vers** [Utilisateur]."

Avec toutefois une forme spécialisée simplifiant l'expression :

"**Rapproche** [objet (A)] **de** [quantité distance]."

5.4.5. Les mouvements circulaires

En ce qui concerne maintenant les mouvements circulaires, nous avons défini deux catégories, les rotations et les girations.

La forme générale que nous avons retenue pour les rotations est la suivante :

[Sélection (Opération de rotation)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (Valeur)]
 & [Sélection (Unité)]
 & [Sélection (Axe)]
 & [Sélection (Sens)]

Nous aurons, pour les girations, une forme similaire à laquelle s'ajoute une définition de position de l'axe autour duquel s'effectue le mouvement :

[Sélection (Opération de rotation)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (Valeur)]
 & [Sélection (Unité)]
 & [Sélection (Axe)]
 & [Sélection (Sens)]
 & [Sélection (Objet Graphique 3D)]

Avec, pour ces girations, une éventuelle simplification à l'aide d'une spécification de direction orthogonale prenant le rôle de tangente.

[Sélection (Opération de giration)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (Valeur)]
 & [Sélection (Unité)]
 & [Sélection (Direction)].

La rotation d'un objet s'exprime de manière spécifique à l'aide d'une sélection d'axe et sens de rotation spécifique, deux informations auxquelles nous avons attribué un geste mimétique manuel les indiquant, voire les sélectionnant.

Dans le premier cas pour lequel l'acte mimétique n'apporte qu'une *présélection*, une validation vocale complémentaire peut s'énoncer sous une forme démonstrative telle que « *comme ceci!* », ou encore « *de cette façon!* », etc., cependant qu'on effectue une désignation manuelle d'axe et de sens de rotation à l'aide d'un geste mimétique.

Une syntaxe correspondante se décrira sous la forme :

" **Tourne** [objet (A)] **de** [quantité angulaire] **comme ça !** "

Dans le second cas pour lequel on considérera que le même geste mimétique produit une *sélection* d'axe et sens de rotation, du fait qu'il constitue un acte volontaire de courte durée, la syntaxe peut être simplifiée de la manière suivante :

" **Tourne** [objet (A)] **de** [quantité angulaire] + [axe & sens] "

Pour ce qui concerne maintenant les girations (Figure 107), l'application du même modèle s'opère. Avec une syntaxe résultante peu usitée, du fait de l'absence de terme d'opération spécifique déterminant ce type de mouvement.

" **Tourne** [objet (A)] **de** [quantité angulaire] **autour de** [Utilisateur] **comme ça !** "

Exemple : « *Tourne cet objet de 15° autour de moi comme ça !* ».

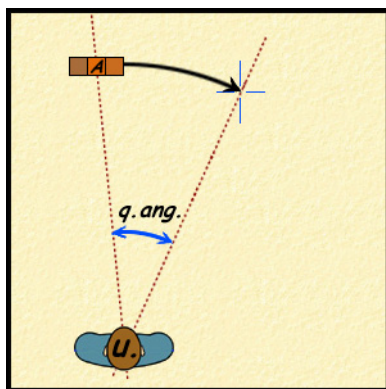


Figure 107. Exemple de mouvement de giration avec spécification de quantité angulaire.

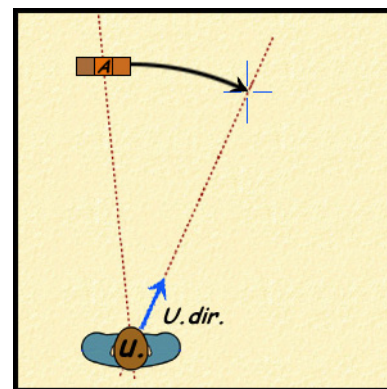


Figure 108. Exemple de mouvement de giration avec spécification de direction égocentrée.

Il semble toutefois qu'un tel mouvement puisse s'exprimer plus couramment, à l'aide des modalités mises en œuvre, sur la base de la forme générale d'expression d'un *positionnement égocentré* dont la distance à l'utilisateur serait omise de l'expression (Figure 108).

[Sélection (Opération de positionnement)]
& [Sélection (Objet Graphique 3D)]
& [Sélection (direction)].

" **Place** [objet (A)] **dans** [direction égocentrée]. "

Forme qui ne correspond plus à la tâche du fait de la difficulté, pour l'utilisateur, d'évaluer l'angle, mais qui peut cependant figurer au sein d'une tâche de giration n'impliquant pas de précision particulière.

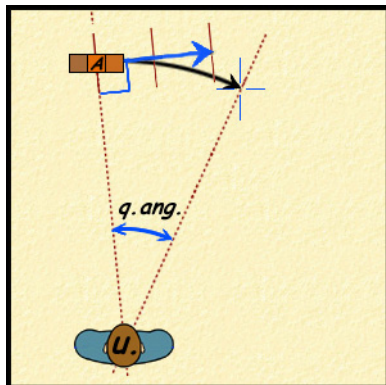


Figure 109. Giration exprimée sous la forme d'un déplacement.

Une mise en oeuvre différente des modalités disponibles trouve une autre solution sur la base du modèle général des *déplacements* avec désignation d'une direction orthogonale (Figure 109).

[Sélection (Opération de déplacement)]
 & [Sélection (Objet Graphique 3D)]
 & [Sélection (quantité angulaire)]
 & [Sélection (Direction)]

De syntaxe :

"**Déplace** [objet (A)] **de** [quantité angulaire] **vers** [direction relative]."

Outre le fait que cette syntaxe ne constitue pas une expression étrangère au langage courant, Nous ne constatons pas de conflit de mise en oeuvre des modalités associées.

Cette dernière application d'un modèle relevant des translations laisse envisager celle du modèle d'un positionnement relatif à un second objet référent (Figure 110).

[Sélection (Opération de positionnement)]
 & [Sélection (Objet Graphique 3D)]
 [Sélection (quantité angulaire)]
 & [Sélection (direction)]
 & [Sélection (Objet Graphique 3D)]

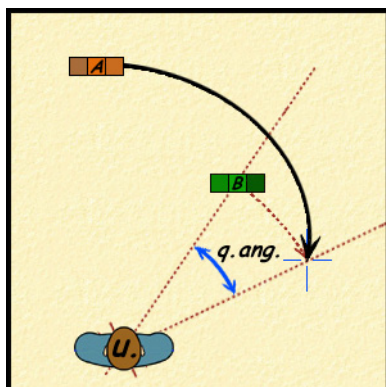


Figure 110. Exemple de positionnement utilisant une quantité angulaire.

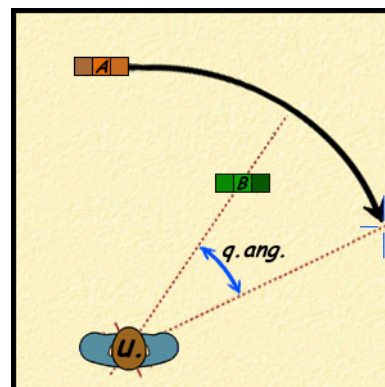


Figure 111. Second exemple de positionnement utilisant une quantité angulaire.

Une syntaxe correspondante serait alors :

"**Positionne** [objet (A)] **à** [quantité angulaire] **à** [direction relative] **de** [objet (B)]."

Exemple : « *Place cet objet à 15° à droite de cet objet!* ».

Cette forme de positionnement est géométriquement concevable par association de la sélection d'une opération de positionnement à celle d'une quantité angulaire dans une direction exocentrée. Toutefois, la forme syntaxique correspond à deux opérations possibles à exécuter (Figure 110 et Figure 111). Par conséquent, outre le fait qu'une telle syntaxe semble peu usitée, nous obtenons par ailleurs un conflit pour une tâche utilisateur non prévue initialement dans les catégories retenues.

Pour résumer ce dont il a été question dans ce chapitre, en premier examen de définitions de syntaxes correspondant aux modèles de tâches définis, nous constatons une conformité à une mise en œuvre séquentielle d'actes monomodaux ou bimodaux non ambigus.

L'absence de conflits d'expression pour un exemplaire de chaque catégorie de tâches retenue, nous permet de conclure maintenant que pour les mouvements d'objets 3D d'une scène virtuelle, le modèle défini d'expression multimodal en entrée du système, basé sur trois modalités d'entrée successivement mises en œuvre seules ou de façon généralement combinées par paires, permet *a priori* à l'utilisateur d'exprimer chaque catégorie retenue de mouvements géométriques qui constituent les objectifs des tâches de mouvement de ces objets 3D.

Chapitre 6. Expérimentation

Cette partie traite de l'objectif expérimental que nous cherchons à atteindre, puis des contraintes qui président à la conception d'une expérience avec des utilisateurs qui permettra de qualifier le style d'interaction proposé. Elle présente ensuite la manière suivie pour répondre aux principales questions qui se posent, décrit l'expérience mise en place et les traitements effectués sur les mesures réalisées.

6.1 Objectif

Nous avons considéré pour ensemble de tâches utilisateur l'équivalent de la manipulation d'objet 3D dans l'espace. Nous avons ensuite défini une association de modalités d'entrée propre à traduire une communication d'intention utilisateur correspondant à l'ensemble de ces tâches. Ces choix de modalités et dispositifs sont déterminés par l'hypothèse préalable selon laquelle l'intervention de la direction de la tête ou du regard pour effectuer une visée améliore sensiblement l'exécution de certaines tâches (Ware & Mikaelian, 1987, Tanriverdi & Jacob, 2000, voir aussi page 71 et suiv.), alors que l'intervention du geste manuel sera plus appropriée pour d'autres tâches comportant une information à caractère spatial (le geste déictique ou mimétique spatial. Bellik, 1995, voir aussi § 5.2.2 page 115), et que l'intervention de l'énonciation vocale contribue davantage à la commande d'opérations à exécuter pour peu que l'on réduise la tâche de reconnaissance vocale à un sous-ensemble du langage naturel ne couvrant que le domaine de l'application envisagée (Rosenfeld & al, 2001).

Puis, nous avons déterminé les profils de techniques propres à convoier les informations de ces modalités en leur cherchant un dispositif adéquat. Nous avons ensuite décrit des scénarios de mise en œuvre de ces modalités au sein d'un exemplaire de chaque tâche comme préalable à la définition complète d'une grammaire. Ces scénarios sont établis pour former un ensemble cohérent et homogène de mises en œuvre des modalités, qui fait apparaître d'éventuels conflits et soulève de nouvelles questions. Par exemple, l'acte mimétique manuel à manifester pour signifier l'axe et le sens d'une rotation est-il à considérer comme une sélection ? Une tâche particulière, pour constituer sans ambiguïté une expression de commande, trouve-t-elle dans la langue un énoncé correspondant ? Quelles sont les différentes formes d'énoncés possibles d'une tâche compte tenu de la rigidité relative du rôle assigné aux actes gestuels ? La différence entre la langue académique et le langage parlé apporte-t-il une ambiguïté, ce qui semble prévisible, dont il faudra nécessairement tenir compte dans la description plus complète d'une grammaire « utilisable » ? Etc.

Il nous faut donc également, avant d'aborder l'étude de niveaux d'interaction plus élevés, voire d'aborder d'autres ensembles de tâches utilisateur, procéder par étapes successives et notamment ici répondre à une question de fond concernant la pertinence de ce style d'interaction vis-à-vis des utilisateurs. A ce stade nous ne savons pas, en effet, si cette manière d'interagir avec un système informatique en vue

d'obtenir une action sur un objet graphique 3D constitue réellement une alternative aux techniques connues.

Nous recherchons également à obtenir des informations sur la variabilité d'usage de la parole et son impact sur l'ensemble de l'expression de commande en vue d'établir une définition d'une grammaire plus détaillée.

Ceci influencera directement la conception de l'expérience en termes de présentation des objectifs à atteindre aux utilisateurs.

6.2 Problématique et hypothèses

D'un point de vue global, nous avons conçu un système d'interaction autour d'une métaphore de commande que l'utilisateur met en œuvre à l'aide d'expressions vocales et gestuelles similaires à celles que l'on observe au cours d'un dialogue interpersonnel de commande. Bien que ce style d'interaction soit observable au cours d'activités réelles, nous ne disposons pas d'informations concernant l'efficacité d'une interaction multimodale orientée communication de commandes avec un environnement 3D d'un point de vue informatique. Une telle carence nous conduit par conséquent à envisager une expérimentation à caractère exploratoire pour déterminer la pertinence d'une telle métaphore. Ce caractère exploratoire nous réduit, au cours d'une première approche, à quelques hypothèses générales ainsi qu'à la mise en place de mesures qui confirmeront ou infirmeront l'intérêt de cette voie de recherche.

Dans le cas qui nous occupe, ne sachant qu'imparfaitement comment un utilisateur potentiel mettra la modalité vocale en œuvre, nous n'avons fait que réduire cet usage à une syntaxe approximative relevant des « *langages de commande* » en évitant une définition trop précise qui répondrait éventuellement à un choix délibéré du concepteur de l'interaction. Il existe des études sur la parole en situation d'IHM dans un environnement multimodal d'interaction avec du graphique (Carbonell & Dauchy, 1999. Robbe et al. 2000. Carbonell, 2003). Mais ce graphique est du 2D et non du 3D, (D'où l'intérêt de faire une étude dans le cas du 3D). Nous avons en effet rapidement constaté que la variabilité d'énonciation d'une commande, vocale par conséquent, était la contrepartie de la richesse offerte par cette modalité d'expression. Ce qui revient à établir que nous ne disposons pas d'une connaissance précise de la façon de s'exprimer des utilisateurs. Ce qui signifie également que nous ne pouvons provisoirement tester, en l'attente de pouvoir le faire sur la syntaxe de commandes complètes, que les sous tâches atomiques d'interaction (présélections et sélections implicites), ainsi que celles situées au niveau immédiatement supérieur (sélections explicites), sans pouvoir présager de leur agencement au sein de l'expression de commandes.

Nous allons néanmoins chercher à déterminer, au cours d'une expérience impliquant des utilisateurs :

-
1. d'une part la pertinence d'une interaction orientée communication incluant les actes non verbaux pour effectuer des tâches de mouvements d'objets 3D au sein d'un environnement virtuel,
 2. tester d'autre part à cet effet la syntaxe que nous avons définie pour l'usage et l'affectation des modalités gestuelles au niveau des tâches de bas niveau (présélections et sélections) ainsi que l'orientation « *langages de commande* » pour la mise en œuvre combinée des trois modalités retenues,
 3. et enfin obtenir davantage d'informations sur l'usage de la modalité vocale et son impact d'ensemble sur la formulation d'expressions de commande.

Pour ce faire, nous allons pratiquer un test utilisateur. Ce type de test permet d'obtenir des performances et des préférences à l'usage d'un produit en vue par exemple d'en déceler les défauts et de l'améliorer.

Dans notre cas, particulier de par l'usage d'une modalité vocale dont on cherche à déterminer l'usage correspondant à un ensemble de tâches afin de promouvoir une meilleure adaptation de l'interaction à l'utilisateur, nous ne disposons que d'une multimodalité d'entrée manifestant les actes communicatifs de l'utilisateur munie d'une grammaire de commande incertaine et approximative à préciser. Ne connaissant pas la variabilité d'expression qu'apporte l'usage de la parole, nous ne pouvons pas par exemple, et en particulier, présenter les objectifs des tâches à atteindre aux utilisateurs de manière traditionnellement textuelle dont ils n'auraient qu'à reproduire les termes.

Pour aborder maintenant la question des profils d'utilisateurs auxquels soumettre notre application à tester, nous avons précédemment indiqué que la défection du partenaire industriel du projet initial nous a conduit à définir un ensemble de tâches utilisateur présentant un caractère plus général que des tâches spécifiques au domaine de la conception vestimentaire. Ne disposant plus, en conséquence, d'un accès à une population cible d'utilisateurs parmi les concepteurs de ce domaine, nous ferons appel, pour cette première approche qui n'est plus spécifique au domaine initial, à des utilisateurs proches du site d'expérimentation. Le laboratoire dans lequel l'expérience a été mise en place est situé dans l'enceinte du centre de recherche INRIA de Rocquencourt en région Parisienne. Ce centre de recherche rassemble plusieurs centaines d'utilisateurs potentiels parmi lesquels on distingue les chercheurs et étudiants en informatique du personnel administratif.

D'autres arguments militent par ailleurs en faveur du choix d'une population d'informaticiens pour une telle expérimentation. Au cours d'une expérience mesurant le rappel des termes d'un langage de commande généré par les utilisateurs, D. Scapin (1982) soulève le problème que pose l'incertitude de l'expérimentateur concernant la compréhension effective par l'utilisateur de la tâche qui lui est assignée au cours d'un test de performance, en particulier lorsque des fonctions complexes sont susceptibles de lui paraître "peu naturelles" du fait qu'elles ne présentent pas de similitude avec

une tâche connue. L'auteur évoque par ailleurs le "*problème d'induction*" qui risque d'influencer les réponses de l'utilisateur au cours d'une telle expérience dans laquelle la présentation de la tâche assignée par l'expérimentateur contient les éléments de la commande. Une conclusion de cette expérience consistait en ce que l'activité de génération de noms de commandes par les utilisateurs facilite leur rappel ultérieur. Ceci nous conduit à rechercher une solution expérimentale qui évite les deux problèmes soulevés tout en permettant à l'utilisateur de mettre ses propres expressions en oeuvre.

Au cours d'une autre expérience mesurant, chez des utilisateurs, le rappel des termes d'un langage de commande à partir de définitions "*opérationnelles*" versus "*fonctionnelles*", (respectivement, opération réalisée du point de vue de l'opérateur versus fonction exécutée par le système), D. Scapin (1981) montre que les utilisateurs expérimentés en informatique semblent disposer d'un atout qui les aide à reconnaître des commandes et des définitions, même lorsqu'elles sont décrites soit de manière orientée opérateur, soit de manière fonctionnelle. Les utilisateurs expérimentés rappelant d'autre part davantage de synonymes que les utilisateurs non expérimentés, il suppose par ailleurs, à l'époque, que l'absence de standardisation des commandes d'un système à l'autre laisse à l'utilisateur expérimenté plusieurs termes possibles en concurrence pour une même définition d'une commande à rappeler, là où les utilisateurs novices ne disposeraient que du terme enseigné à reproduire pour la commande demandée (Scapin 1981).

Ces résultats nous conduisent à suspecter une différence de performances des utilisateurs du point de vue de leur expérience avec l'outil informatique. Les différences d'expérience en informatique à notre époque se situent en effet en grande partie dans la diversité des styles d'interaction auxquels ont été confrontés les utilisateurs. L'utilisateur d'informatique non expérimenté n'aura généralement eu affaire qu'à des applications informatiques standardisées autour d'un unique style d'interaction WIMP sur des plateformes classiques "écran-clavier-souris". Les utilisateurs expérimentés en informatique auront, pour leur part, également eu affaire à des langages de commande constituant un modèle alternatif d'interaction rappelant davantage la syntaxe générale que nous proposons pour la forme d'interaction proposée dans cette étude. Ces utilisateurs expérimentés auront par ailleurs éventuellement été sensibilisés à d'autres dispositifs d'entrée non classiques, tels que ceux permettant le suivi des mouvements ou encore la reconnaissance vocale. Nous allons par conséquent supposer que l'expérience en informatique constitue une variable expérimentale pertinente pour tester la forme d'interaction proposée au cours d'un test de performance avec des utilisateurs.

Ces travaux de D. Scapin et le fait que nous cherchons également à déterminer comment s'expriment les utilisateurs pour manifester une commande de repositionnement géométrique particulière, confortent d'autre part une présentation des tâches à accomplir non textuelle évitant d'influencer les utilisateurs et

minimisant de plus, si possible, l'incertitude de l'expérimentateur concernant la compréhension de la tâche à accomplir par l'utilisateur.

Nous devront enfin procéder en sorte de n'imposer aux utilisateurs que la syntaxe générale informant une commande complète, et procéder à une estimation qualitative des expressions de commande produites.

Une éventuelle convergence vers des résultats similaires pour chacun des deux groupes envisagés nous laissera soupçonner un défaut de conception de l'environnement d'interaction ou éventuellement de mise en œuvre de l'expérience, voire de facilité « universelle » d'usage sans permettre de conclure. De faibles résultats similaires pourront ainsi indiquer une difficulté à exploiter un style d'interaction mal conçu. Une différence de résultats entre les deux groupes, selon son importance, nous permettra d'infirmer une mise en cause de la conception du modèle interactif proposé ou de l'expérience, et confirmera accessoirement la pertinence de la variable choisie dénotant l'existence d'un facteur constituant soit un avantage permettant à l'un des deux groupes d'obtenir de meilleurs résultats, soit inversement un handicap octroyant de moindres résultats à l'autre groupe, (soit encore les deux).

6.3 Méthode

Un échantillon d'utilisateurs sujets se verra proposer la mise en œuvre de séquences d'interaction avec des objets graphiques 3D disposés sur une scène 3D. Ces utilisateurs sont munis des dispositifs d'entrée leur permettant de transmettre leurs expressions multimodales de commande.

"L'utilisabilité" d'un système se juge traditionnellement sur l'examen de facteurs tels que l'efficacité ("*effectiveness*"), la facilité à en apprendre l'usage ("*learnability*"), sa flexibilité ("*flexibility*"), qui permet de modifier l'état du système sans en affecter la facilité d'usage, et l'attitude ("*attitude*") qu'il génère chez l'utilisateur du point de vue satisfaction (Shneiderman, B., 1998).

Ce que l'on mesure concerne typiquement les durées de réalisation des tâches assignées, les erreurs effectuées, le rapport entre succès et insuccès, la fréquence avec laquelle les diverses commandes sont utilisées, etc.

L'utilisateur se voit par ailleurs soumettre un questionnaire en début de séance afin d'y porter diverses informations le concernant, permettant de situer son expérience tant en informatique graphique que multimodale, son âge, sexe, etc.

Ce questionnaire se voit complété par ses soins en fin de séance afin d'y reporter des informations subjectives et qualitatives telles que préférences, critiques, observations, etc.

Un ensemble de tâches utilisateur ayant été défini, il nous faut proposer aux utilisateurs d'exprimer une commande pour les plus représentatives d'entre elles à l'aide de la multimodalité d'entrée. Les expressions produites permettront d'évaluer le caractère d'universalité et/ou d'atomicité de chaque sous tâche, de caractériser

l'association simultanée de sous tâches complémentaires pour réaliser les tâches séquentielles de présélection et de sélection, et d'évaluer la variabilité d'énonciation des opérations d'une part, de l'ensemble des expressions d'autre part.

La question suivante consiste à résoudre la manière de présenter l'expérience à l'utilisateur. Nous pouvons en effet « enseigner » à l'utilisateur une mise en œuvre préétablie des modalités d'entrée, éventuellement arbitraire, afin de déterminer s'il lui est ensuite possible d'utiliser le système en assimilant des expressions préétablies. S'agissant d'un contexte de communication en situation de commande, nous pouvons également, à l'inverse, ne renseigner l'utilisateur qu'au minimum requis pour lui laisser exprimer librement sa propre façon de commander le système pour chaque tâche afin de déterminer la variabilité d'expression de l'échantillon pour chaque tâche proposée, ou, *a contrario*, de constater une homogénéité des commandes et des termes employés pour chaque tâche, qui réduiront les sous-ensembles d'expression ainsi que la grammaire à finaliser pour cet ensemble de tâches de mouvement d'objets 3D.

La seconde méthode présente l'avantage de fournir une réaction observable de l'utilisateur au premier contact avec un nouvel environnement d'interaction inconnu. Cette réaction renseigne sur la capacité du système à être appréhendé rapidement par l'utilisateur novice. Par ailleurs, constatant la difficulté à établir clairement les profils génériques ou synonymiques des énoncés vocaux d'une commande, un tel échantillonnage permet de déterminer comment s'expriment les utilisateurs pour exécuter une tâche de mouvement d'objet.

Huit séquences définies comme les plus représentatives concernent les "translations" et "mouvements circulaires" d'objets graphiques 3D présents sur la scène. Plusieurs d'entre elles concernent des translations d'objet exprimées de façon absolue relativement à la position virtuelle de l'utilisateur, relativement à la position occupée par l'objet, ou relativement à la position occupée par un second objet désigné, et incluant des expressions spécifiant une quantité de déplacement soit linéaire, soit angulaire. Une seule séquence d'orientation d'objet est retenue.

Le tableau ci-après résume les huit cases expérimentales retenues pour un déplacement ou orientation d'objet graphique 3D.

Chacune des cases expérimentales se voit attribuer vingt utilisateurs répartis en deux groupes de dix utilisateurs informaticiens et dix utilisateurs non informaticiens.

Chaque utilisateur se voit proposer l'ensemble des séquences dans un ordre différent pour contrebalancer l'effet d'apprentissage.

Combinaisons de modalités. Exemples d'énoncés.	I. Bimodalité : Dir.Tête + Vocal	II. Trimodalité : Dir.Tête + Mvt Main + Vocal
F1 Positionnement égo-centré. Ex. " <i>Place cet objet à 2 mètres dans cette direction</i> "		Opération: Vocal Objet: Dir.Regard + Vocal Quantité: Vocal Direction: Déict.Main + Vocal ou Dir.Tête + Vocal
F2 Déplacement exocentré. " <i>Déplace cet objet de 2 mètres dans cette direction</i> "		Opération: Vocal Objet: Dir. Regard + Vocal Quantité: Vocal Direction: Déict.Main + Vocal
F3 Déplacement relatif " <i>Déplace cet objet de 2 mètres vers la droite</i> "	Opération: Vocal Objet: Dir. Regard + Vocal Quantité: Vocal Direction: Vocal	
F4 Repositionnement égo-centré approximatif " <i>Met cet objet par là</i> "		Opération: Vocal Objet: Dir. Regard + Vocal Quantité: pas de Qté Direction: Déict.Main + Vocal ou Dir.Tête + Vocal
F5 Giration relative " <i>Déplace cet objet de 10 degrés vers la droite</i> "	Opération: Vocal Objet: Dir. Regard + Vocal Quantité: Vocal Direction: Vocal	
F6 Déplacement ciblé " <i>Déplace cet objet de 2 mètres en direction de cet objet</i> "	Opération: Vocal Objet: Dir. Regard + Vocal Quantité: Vocal Direction: Vocal + [Dir.Tête + Vocal]	
F7 Positionnement relatif référencé à un second objet. " <i>Déplace cet objet à 2 mètres à droite de cet objet</i> "	Opération: Vocal Objet: Dir. Regard + Vocal Distance: Vocal Direction: Vocal + [Dir.Tête + Vocal]	
F9 Rotation " <i>Tourne cet objet de 30° comme ça</i> "		Opération: Vocal Objet: Dir. Regard + Vocal Quantité: Vocal Direction: Vocal + Mimétique Main

Tableau 4. Huit séquences de mouvement d'objet 3D retenues pour l'expérience

6.3.1. Pré-expérience

Deux types de tests de l'environnement ont été réalisés.

Une inspection ergonomique réalisée par des utilisateurs du laboratoire, experts proches du domaine de l'ergonomie des environnements virtuels, a permis, le cas échéant, de situer d'éventuels problèmes ergonomiques connus.

Quelques utilisateurs témoins se sont vus ensuite proposer la mise en œuvre des séquences d'interaction définies afin de tester l'efficacité de l'ensemble du dispositif.

6.3.2. Le « Magicien d'Oz »

Le partenaire industriel impliqué dans le projet originel ("COMEDIA") faisant défaut pour la réalisation technique, un moyen terme est adopté en juillet 2003 afin de permettre néanmoins à l'étude d'aboutir à des résultats utiles.

La méthode adoptée est alors celle dite du "*Magicien d'Oz*", et consiste en une simulation réaliste des réponses du système aux commandes de l'utilisateur. L'utilisateur fait face à la présentation réaliste de l'application avec laquelle il interagit, mais c'est un comparse invisible de l'utilisateur qui gère la réponse du système (le « *Magicien* »). L'expérimentateur renseigne d'abord le participant quant aux modalités d'entrée à utiliser ainsi que les contraintes adoptées (syntaxes multimodales pour les sous-tâches en entrée et complétude nécessaire de l'expression de commande, notamment). Il le renseigne par ailleurs en ce qui concerne l'objectif à atteindre pour chaque séquence d'interaction qui lui sera présentée (des commandes de relocalisation d'objets). Chaque séquence de déplacement d'objet se voit donc attribuer un déroulement prédéfini dont le résultat est soumis à l'utilisateur en terme d'objectif à obtenir de la part du système. Le rôle du « *Magicien* » consiste alors soit à valider la réalisation de la séquence prédéfinie correspondante si l'expression de l'utilisateur est complète, soit à transmettre une demande de reformulation de l'expression.

Le plan adopté prévoit l'enregistrement analogique et digital des commandes multimodales des utilisateurs afin de permettre l'étude, par la suite, des occurrences des événements dans chaque modalité qui permettront ultérieurement une meilleure définition de la "*fusion*" des modalités d'entrée.

6.3.3. Les utilisateurs sujets

Un échantillon de population composé de deux groupes d'utilisateurs, un groupe homogène composé d'informaticiens, un autre, hétérogène, constitué de non informaticiens, a effectué les tâches assignées et a été recruté sur le site de l'INRIA de Rocquencourt. Nous utilisons deux groupes afin de tester de prime abord, une éventuelle convergence pouvant orienter une expérimentation ultérieure.

Il s'agit de préparer une situation en laboratoire dans laquelle l'utilisateur, faisant face à une représentation réaliste et cohérente de l'application envisagée, pourra

mettre en oeuvre la syntaxe proposée par l'expérimentateur pour obtenir les divers déplacements prédéfinis d'objets de la scène de la part du système.

Cette mise en situation de l'utilisateur implique un scénario établissant d'une part les rôles de l'expérimentateur et du magicien, d'autre part les rôles proposés aux utilisateurs lors de l'expérience. Pour ce faire, un matériel a été mis en oeuvre.

Vingt utilisateurs sont recrutés sur le site de l'INRIA Rocquencourt afin de constituer deux groupes de dix utilisateurs pour cette expérience.

Le premier groupe (*cf.* Tableau 5 ci-dessous) est constitué d'informaticiens (cinq hommes et cinq femmes). Parmi ces informaticiens, aucun n'est accoutumé à quelque modalité d'entrée inhabituelle aux plateformes classiques *écran-clavier-souris*, mais deux d'entre eux proviennent d'un projet de recherche sensibilisé à l'informatique graphique.

Tous déclarent utiliser régulièrement les systèmes d'exploitation « Windows » ainsi que « Unix ». L'un d'entre eux utilise également « MacOS-X » de façon régulière. Un autre « beOS » lui aussi de manière régulière.

Utilisateurs	Sexe	Age	Nature
GP	Masculin	39	Informatique Graphique
RF	Masculin	33	Informatique Analyse de Données
LV	Masculin	29	Informatique architecture
RD	Masculin	31	Informatique Réseau
RR	Masculin	27	Informatique Graphique
VA	Féminin	58	Informatique Chercheur
CF	Féminin	46	Informatique Admin Syst
BE	Féminin	22	Informatique Appl web
MC	Féminin	49	Informatique SGBD
CI	Féminin	27	Informatique Réseau
Moyenne		36,1	
ET		11,56	

Tableau 5. Utilisateurs informaticiens participant à l'expérience.

Le second groupe (*cf.* Tableau 6), est constitué de non informaticiens (cinq hommes et cinq femmes). Tous sont utilisateurs à plein temps d'applications informatiques de type GUI¹² avec interaction WIMP¹³. Deux d'entre eux utilisent exclusivement des applications sous le système d'exploitation « MacOS-X », six autres n'utilisent que des applications sous « Windows ». Parmi les deux derniers, l'un d'entre eux utilise régulièrement « Windows » et « Unix », l'autre utilise régulièrement « Windows » et « Linux » de manière plus discontinue.

Les applications employées relèvent généralement de la bureautique et des applications grand public du réseau Internet. Deux d'entre eux sont réalisateurs multimedia et utilisent à ce titre des applications informatiques vidéo et audio

¹² « GUI » : Graphical User Interface, ou Interface Graphique Utilisateur.

¹³ « WIMP » : Windows-Icon-Menu-Pointing, ou pointage sur fenêtres présentant icônes et menus.

professionnelles. Un utilisateur est sensibilisé aux applications CAD dans le domaine de la mécanique.

Utilisateurs	Sexe	Age	Nature
MM	Masculin	32	Mathématiques Appliquées
CV	Masculin	29	Journaliste océanographe
BC	Masculin	39	Réalisateur Multimédia
BJ	Masculin	25	Juriste
LS	Masculin	31	Gestion
GE	Féminin	27	Assistance de projet
RC	Féminin	23	Administration Logistique
DC	Féminin	24	Technologie Mécanique (IUFM)
AC	Féminin	25	Administration RH
SI	Féminin	54	Réalisateur Multimédia
Moyenne		30,9	
ET		9,4	

Tableau 6. Utilisateurs Non informaticiens participant à l'expérience.

6.3.4. Construction du montage expérimental

Présentation des tâches au participant à l'expérience

L'expérimentateur doit disposer, pour chaque type de déplacement (huit retenus en tout), d'une représentation du résultat du comportement du système à présenter à l'utilisateur afin de lui permettre de déterminer la formulation de la commande multimodale à fournir au système.

Cette représentation se compose *a priori* d'une situation de départ, (l'état courant de la scène 3D), et d'une situation résultante dans laquelle l'un des objets graphiques 3D de la scène aura subi une relocalisation.

La syntaxe à mettre en oeuvre par l'utilisateur étant multimodale et comportant notamment une partie verbalisation orale, il convient de définir une présentation de la tâche qui n'influence pas directement l'énoncé de l'utilisateur, ce qui serait par exemple le cas avec la présentation d'une consigne textuelle telle que "Déplacer l'objet X de 10m vers la droite ", dont l'utilisateur risquerait de reprendre les termes.

Le type candidat de présentation d'objectifs à atteindre envisagé comme le plus approprié reste une présentation graphique de la scène en perspective plongeante augmentée de figures géométriques renseignées telles que flèches, dimensions, etc.

Une vue plongeante telle que figurée ci-après permet une meilleure compréhension d'un mouvement de l'un des objets à obtenir du système. Cette figure montre l'environnement sur lequel est figuré l'utilisateur ainsi que le champ de vision dont il dispose sur la scène qui lui est présentée au cours de l'expérience. C'est sur ce type de figure que sera représenté le mouvement à obtenir pour chaque séquence d'interaction.

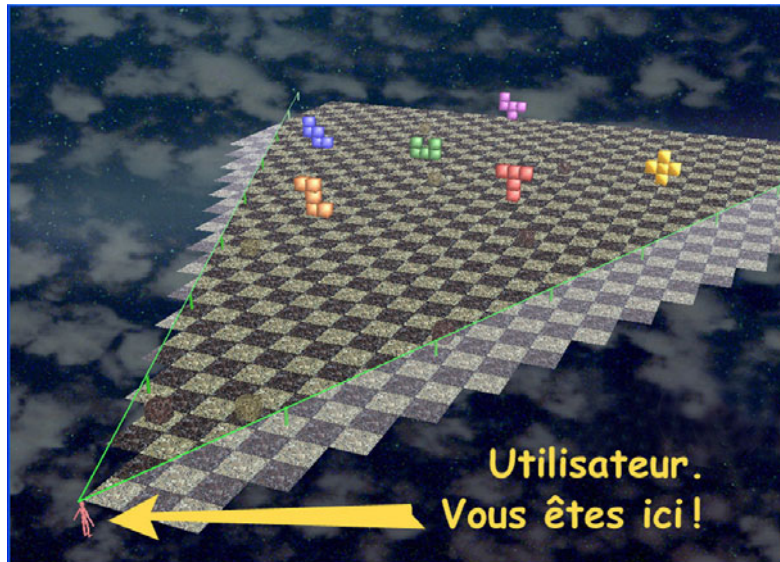


Figure 112. Présentation « *Bird's Eye View* » de la position occupée par l'utilisateur (figuré en rouge en bas à gauche) par rapport à la scène.

Les copies d'écran Figure 113 à Figure 120 sur les pages qui suivent montrent ce qui fut présenté aux participants dans l'ordre suivant :

- F1 : Tâche de positionnement égocentré.
- F2 : Tâche de déplacement exocentré.
- F3 : Tâche de déplacement relatif orthogonal.
- F4 : Tâche de repositionnement égocentré approximatif.
- F5 : Tâche de giration relative orthogonale.
- F6 : Tâche de déplacement ciblé.
- F7 : Tâche de positionnement relatif référencé à un autre objet.
- F9 : Tâche de rotation.

On notera ici l'absence d'une tâche « F8 ». Nous verrons plus loin qu'une huitième tâche de positionnement angulaire orthogonal avait, pour l'expérimentation, initialement été envisagée telle que décrite au chapitre précédent en commentaire des Figure 110 et Figure 111 page 160. L'expérimentation de cette forme peu usitée d'association d'une quantité angulaire à une opération de positionnement orthogonal exocentré ne fut finalement pas mise en œuvre au profit de huit tâches de mouvement munies d'une syntaxe plus courante.

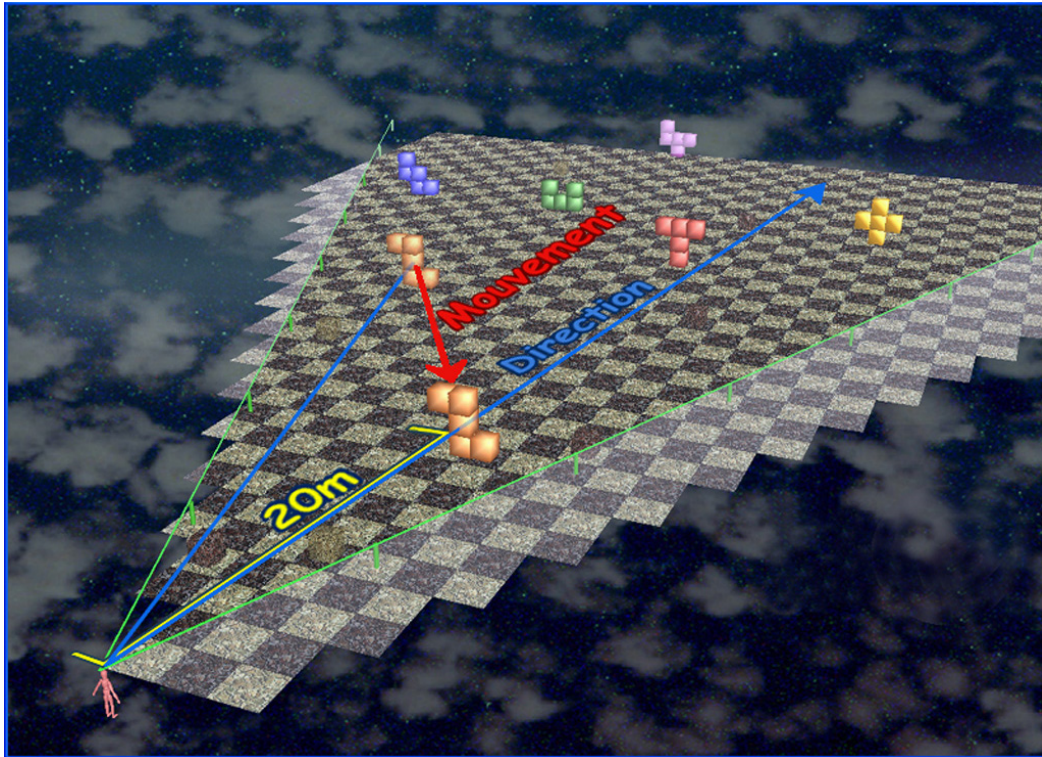


Figure 113. Présentation de la tâche F1 de positionnement égocentré.

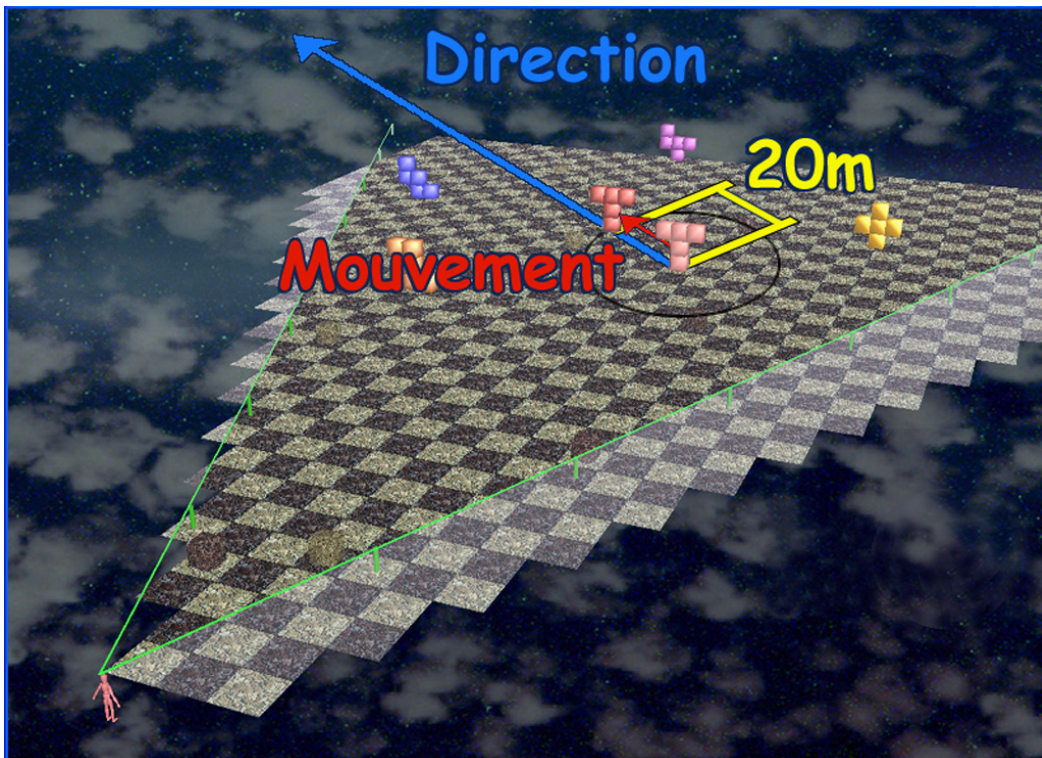


Figure 114. Présentation de la tâche F2 de déplacement exocentré.

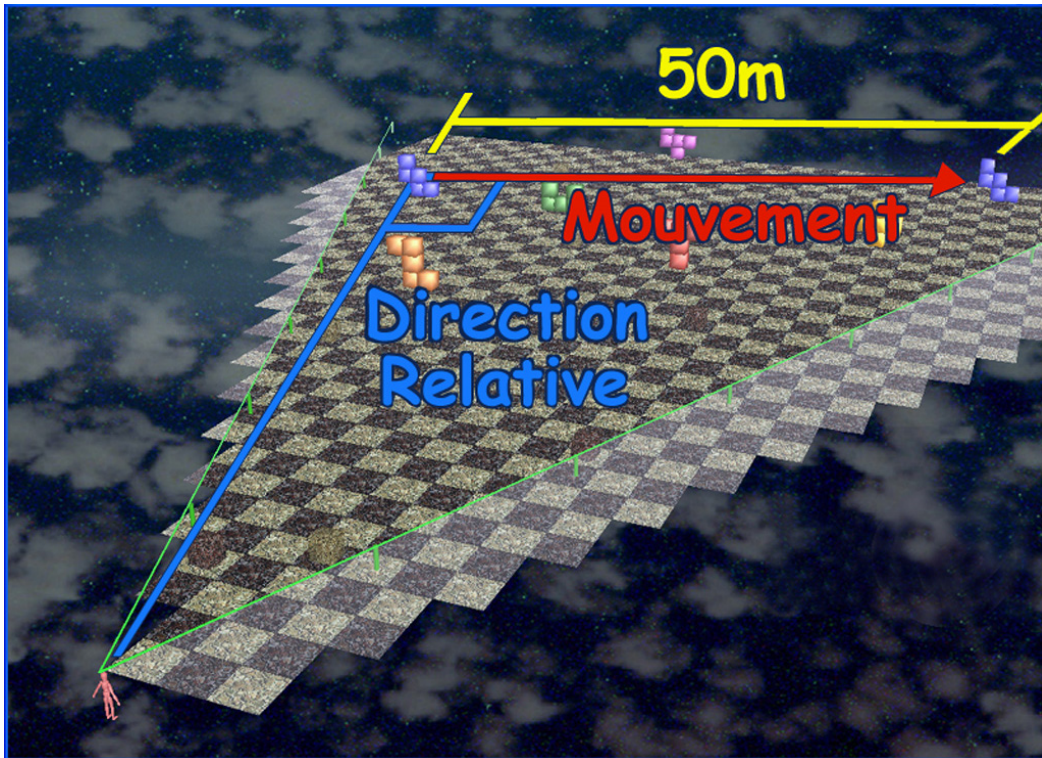


Figure 115. Présentation de la tâche F3 de déplacement relatif orthogonal.

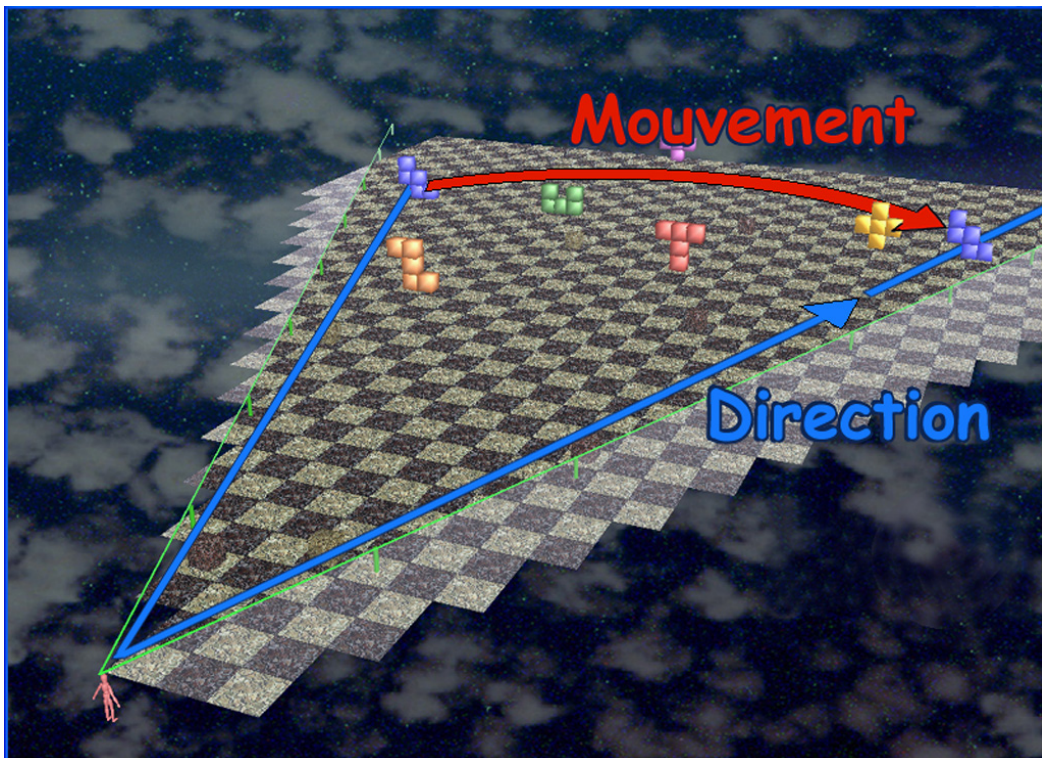


Figure 116. Présentation de la tâche F4 de repositionnement égocentré approximatif.

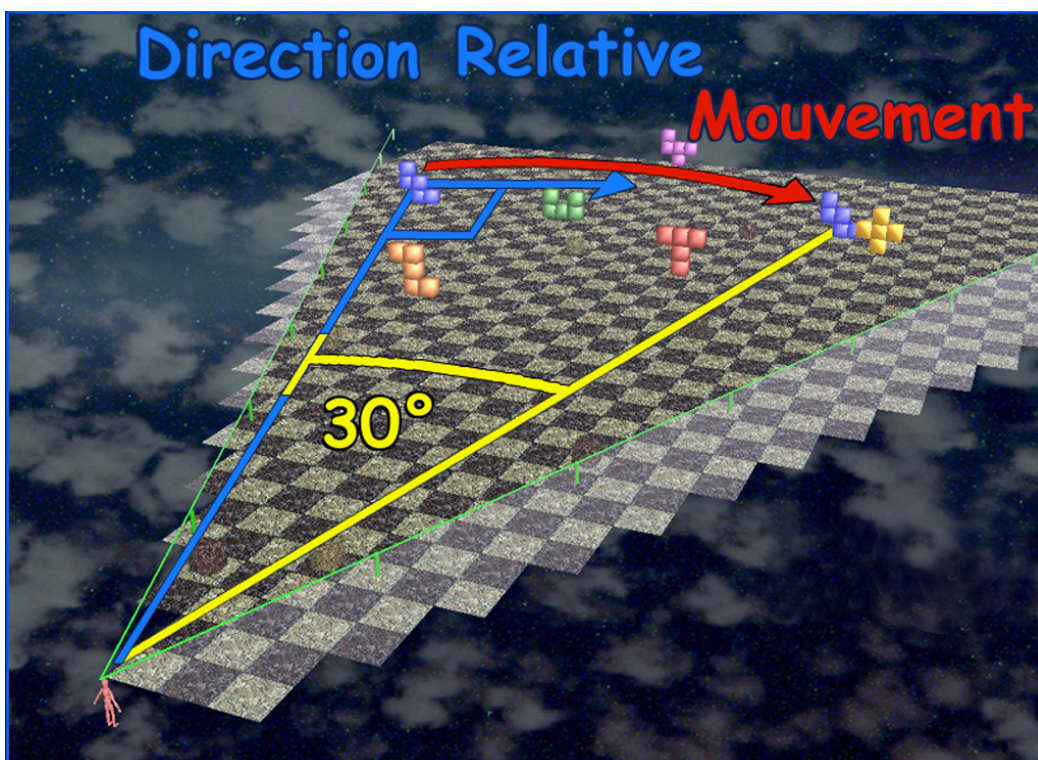


Figure 117. Présentation de la tâche F5 de giration relative.

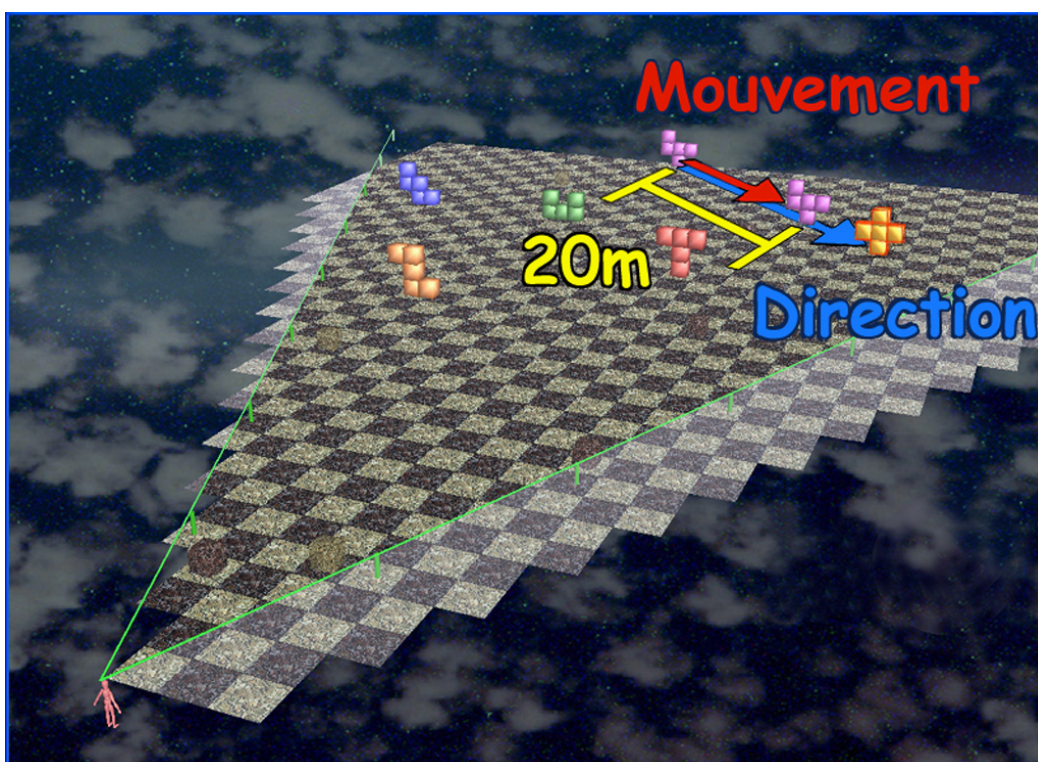


Figure 118. Présentation de la tâche F6 de déplacement ciblé.

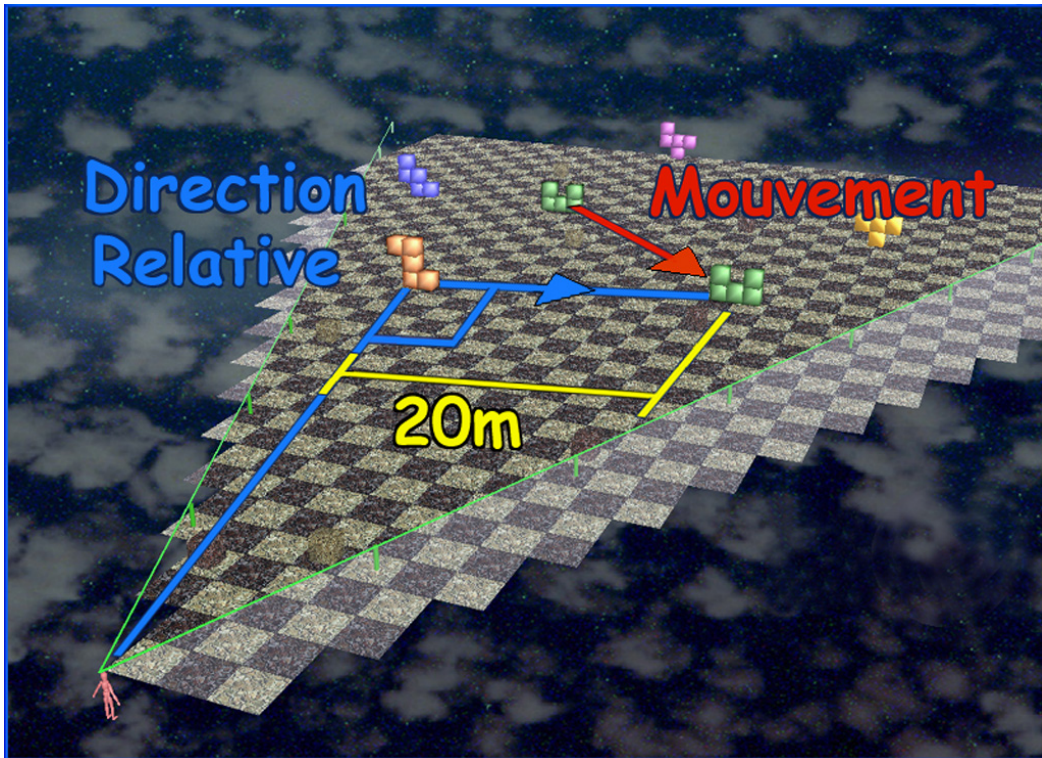


Figure 119. Présentation de la tâche F7 de positionnement orthogonal référencé à un autre objet.

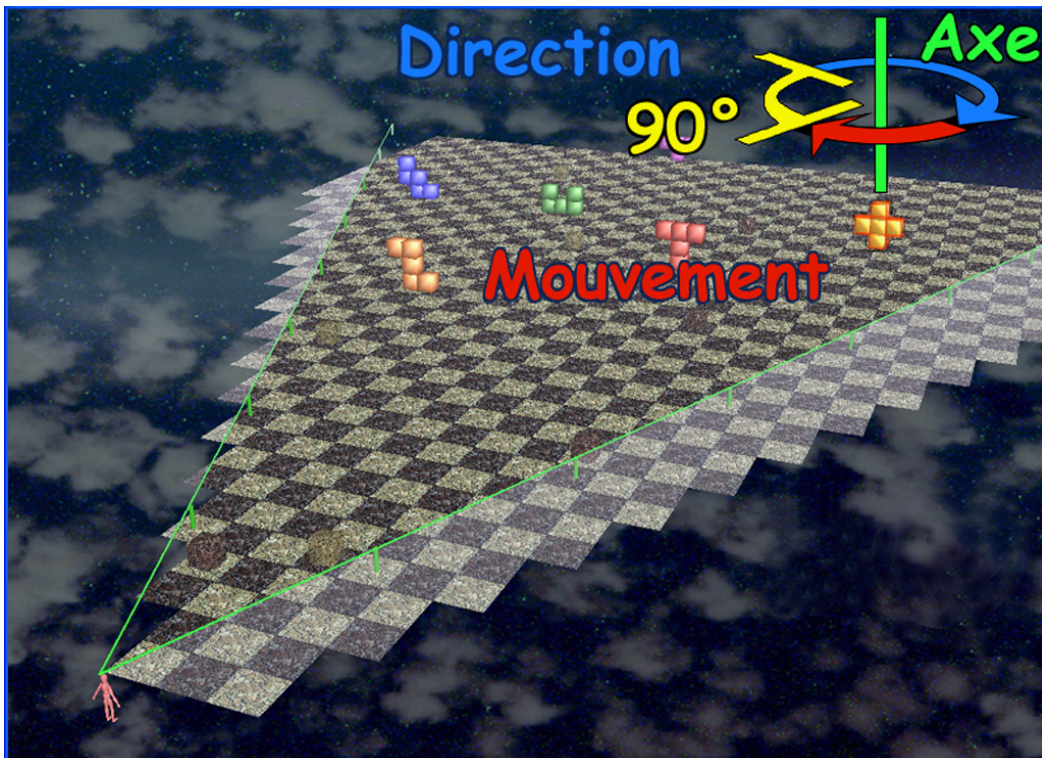


Figure 120. Présentation de la tâche F9 de rotation.

Dispositif

Initialement, la plateforme technique consiste en une application répartie sur trois machines échangeant des données, dont la première récupère les événements et informations des modalités d'entrée provenant de l'utilisateur à des fins d'enregistrement digital daté des séquences d'expression. Une seconde machine gère l'affichage de la scène 3D. Une troisième machine, distante de l'utilisateur et à disposition du magicien, dispose d'un service distant qui lui permet d'initier les séquences d'affichage à l'aide d'un serveur installé sur la machine graphique.

Par ailleurs, afin de compléter les enregistrements digitaux des informations des modalités d'entrée, l'utilisateur est filmé au cours de chaque séquence d'interaction.

La structure initialement envisagée est représentée Figure 121. Sur cette représentation, Les données du tracking magnétique de la tête (HMT) et de la main (GloveMT) de l'utilisateur, sont récupérées par la machine graphique. L'option d'enregistrement de ces informations s'effectuant parallèlement sur une seconde machine récupérant elle aussi les informations vocales et gestuelles (en bas à droite sur la Figure 121).

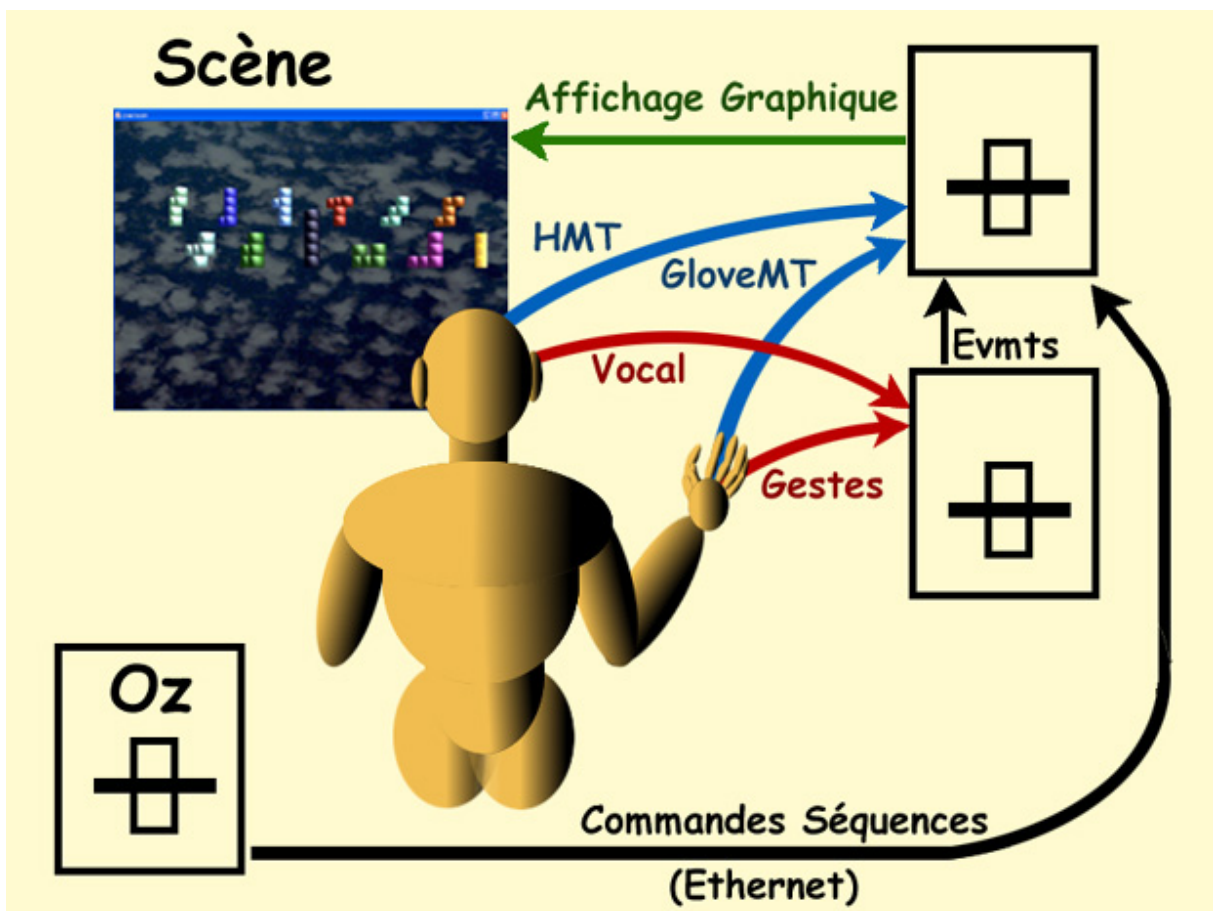


Figure 121 . Structure de la plateforme initialement envisagée.

En pratique, il s'avéra plus judicieux de réunir les informations de modalités d'entrée sur une unique machine de traitement à l'instar des spécialisations de processeurs de sortie dédiés aux restitutions multimedia. Cette machine d'entrée d'information, baptisée « *input* », voit ainsi transiter toutes les informations à enregistrer, trie les informations à transmettre à la machine d'affichage de la scène (baptisée « *machine 3D* »), et présente les objectifs de tâches à l'utilisateur sur un second affichage sous le contrôle radio (RF) de la machine « *ZardOz* » du magicien. Cette répartition s'imposait également, après quelques essais, pour soulager la machine 3D de présentation de la scène. Accessoirement, cela permet de distinguer clairement la machine « 3D » en tant que dispositif dédié aux modalités graphiques de présentation en sortie. Enfin, ce partitionnement indépendant de l'application d'affichage de la scène 3D permet de la développer de façon rapidement réutilisable pour l'application ultérieure de restitution 3D des enregistrements digitaux des gestuelles utilisateur. On peut observer l'organisation de cette répartition sur la Figure 135 page 191, et le synoptique d'ensemble du montage expérimental Figure 122.

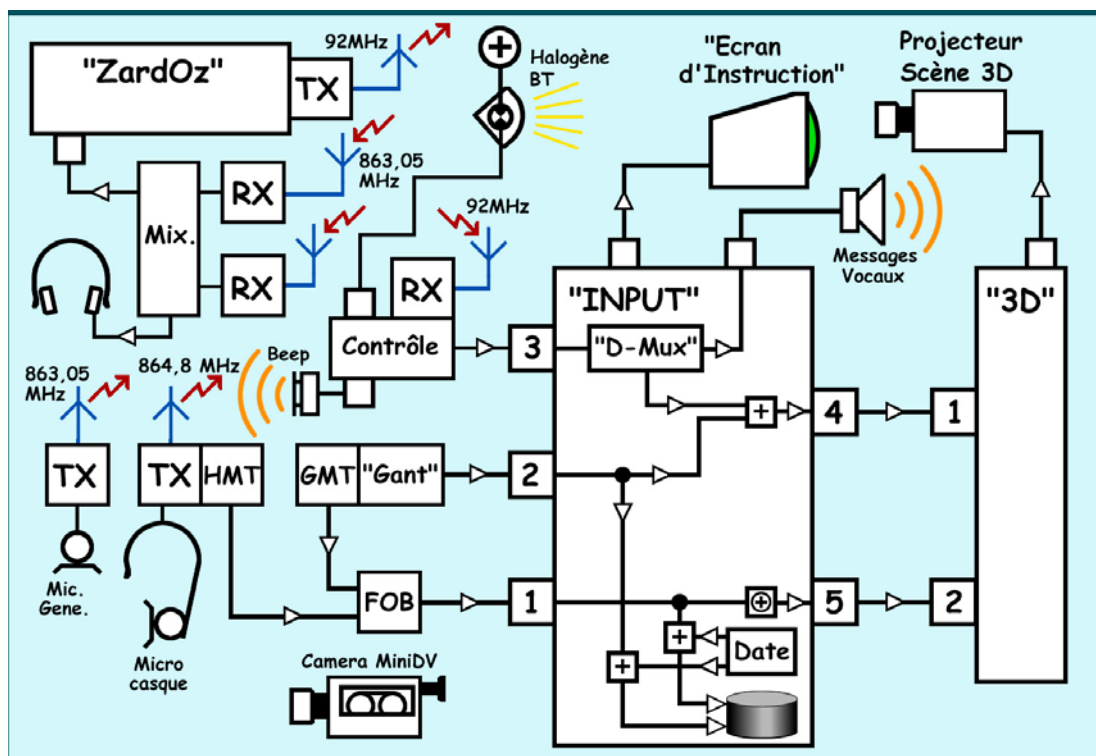


Figure 122. Synoptique d'ensemble du système d'interaction utilisateur avec des objets 3D.

L'acquisition de dispositifs d'entrée relatifs aux gestuelles utilisateurs (tracking magnétique de la tête et d'une main de l'utilisateur), prit un certain retard jusqu'à septembre 2003. L'incertitude sur ces acquisitions et sur le choix d'un « gant de données » pour les modalités gestuelles manuelles conduisit au développement artisanal d'un dispositif de sémiotique manuelle indiquant les "poses" déictiques et "plan" de la main de l'utilisateur. Ce dispositif fait appel à un microcontrôleur informant les événements de modification de "poses" manuelles au travers d'une

liaison série. La récupération des informations issues de ce dispositif s'effectue à l'aide d'une application Java de communication série.

La première étape de développement de ce dispositif a consisté à réaliser une application démonstrative de sa faisabilité, afin également d'en vérifier la pertinence d'un point de vue conception et d'en améliorer la robustesse dans la durée impartie (le prototype fut réalisé en une semaine). Cette application consistait à réaliser une vidéo du gant en fonctionnement devant l'affichage du résultat sur un moniteur classique, dont trois extraits sont visibles Figure 123.

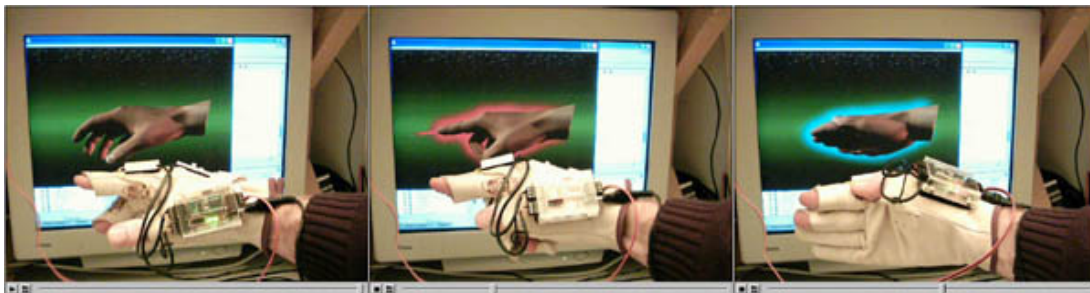


Figure 123 . Trois extraits de la vidéo montrant le gant en fonctionnement. La seconde figure montre le signalage nécessaire du geste de désignation d'une direction.

Ce « gant minimal de données » fut réalisé à l'aide d'un gant de jardinage privé des extrémités de certains doigts afin de permettre, lors de l'écriture des programmes, l'usage du clavier sans impliquer son retrait systématique. On trouvera davantage d'information à son sujet dans l'annexe.



Figure 124. Second prototype du gant, de petite taille pour les mains des participants féminins.

Le contrôle à distance de l'application depuis la machine à disposition du magicien nécessita également un développement pour éviter tout perçage de cloison séparant le laboratoire en salle d'expérimentation et salle d'observation (voir Figure 135 page 191). Un dispositif de transmission radio de signaux de communication fut donc également développé, permettant une éventuelle relocalisation ultérieure du poste de magicien.

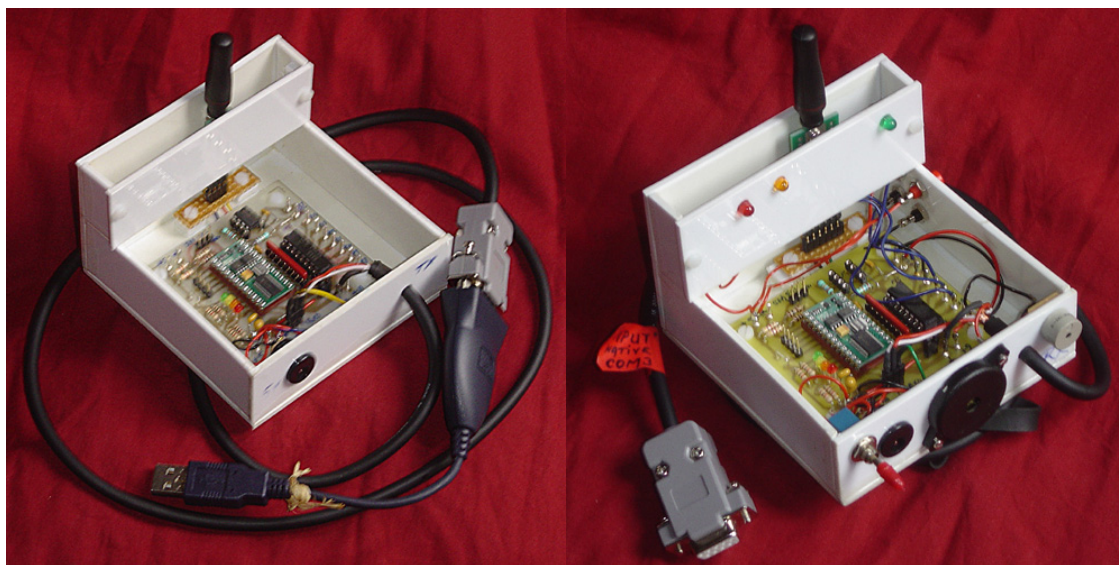


Figure 125. Transmetteur et récepteur radio (RF) de commande digitales de la plateforme à partir de la machine « ZardOz » à disposition du « magicien ».

Le récepteur de commande (Figure 125, à droite), est muni d'un beeper sur la face avant afin de permettre une signalisation audible par le participant sous forme d'un « beep » de 250 millisecondes. Cette présentation d'un « signal de départ » sur un tel canal distinct contrôlé séparément s'avéra, pour obtenir ce résultat, plus simple à mettre en œuvre que tenter de partager les ressources audio en sortie de la « *machine d'entrée* » entre l'application de synthèse vocale et une seconde application qui aurait fourni un tel signal sur cette machine. L'application de synthèse vocale semble en effet mobiliser les ressources audio d'une machine sans qu'il ait, semble-t-il, été prévu de pouvoir la partager aisément.

On peut également apercevoir, sur le panneau vertical en retrait du récepteur, trois LEDs de signalisation. La LED verte à droite signale que le module de réception est en fonction, la LED orange du milieu ne signale que les réceptions de signaux de communication. La LED rouge, à gauche sur le panneau, signale lorsqu'elle s'allume qu'un enregistrement est en cours. Ceci permet au « *magicien* » de constater visuellement l'état de la fonction d'enregistrement de l'application au cours de l'expérience.

Une autre fonction de sortie initialement prévue sur ce composant est le contrôle de sorties dites « à collecteur ouvert »¹⁴. Ce contrôle fut mis en place pour deux raisons. La première concerne une mise en service de l'éclairage du participant sous la forme d'un faisceau produit par un spot halogène basse tension suspendu au plafond de la salle d'expérimentation. Un tel éclairage était prévisible en raison de la nécessité de filmer le participant plongé dans la pénombre ou soumis à un rétro éclairage insuffisant par la scène 3D projetée. Une caméra d'enregistrement nécessite en effet un minimum de lumière pour enregistrer des images qui ne présentent pas de granulations gênant leur interprétation. Ce qui justifiait la mise en place d'une telle source de lumière disposée dans l'environnement proche du participant de manière à ne pas influencer l'expérience. La seconde raison, similaire, consistait à commander, depuis ce composant, la mise en service d'une caméra d'observation du participant et d'un émetteur radio VHF transmettant les images à distance. Fonction qui se révéla par la suite malcommode suite à des problèmes de fonctionnement de cette caméra qui fut remplacée par une caméra MiniDV.

Une fonction de reconnaissance vocale fut également mise en place durant la même période (juillet-août 2003). Elle fut programmée en Java au-dessus d'une application de reconnaissance vocale commerciale, et permet de stocker en tant qu'événements datés d'une part les termes reconnus au sein d'un énoncé, d'autres part les unités grammaticales (commande, sélection d'objet, locus, etc.) détectées au sein de celui-ci. Toutefois, cette application ne fut pas mise en œuvre pour cette première expérience avec les utilisateurs. D'une part à cause du fait que nous ne connaissons pas d'avance, à ce stade, l'étendue du sous-ensemble du langage que sont susceptibles de produire les participants, sous-ensemble qu'il est préférable d'inclure dans le dictionnaire de cette application « multilocuteur » en mode dit de « commande ». D'autre part et alternativement, dans le cas d'une reconnaissance du type « diction » permettant une reconnaissance plus « large » mais ne fonctionnant que pour un unique locuteur, parce qu'il faut alors « entraîner » le module vocal à reconnaître la voix du participant, ce qui peut prendre une trentaine de minutes. Sachant d'autre part qu'une telle application ne dispose d'aucune reconnaissance de « contexte »¹⁵, cette alternative suppose également que l'on connaît à l'avance, pour « entraîner » le module vocal, les termes que ce locuteur est susceptible de produire ultérieurement. Ce qui d'une part n'est pas encore le cas, et qui d'autre part reviendrait, on le verra plus loin, à introduire un apprentissage de l'utilisateur participant qui n'est pas forcément souhaitable selon l'expérience envisagée.

¹⁴ Le terme « Sortie à collecteur ouvert » désigne un transistor intégré fonctionnant à saturation et pouvant drainer des courants plus importants que des transistors TTL intégrés classiques.

¹⁵ Le premier essai de reconnaissance vocale fut l'énoncé suivant : « Déplace-moi de dix mètres vers la droite ». Le résultat s'afficha ensuite sous la forme : « Des places mois 2 dis maître vert la droite ».

Scène 3D

La scène que nous avons ensuite construite afin d'être proposée aux utilisateurs sujets consiste en la projection d'une vue fixe sur l'environnement. Cette vue sur l'environnement correspond à une fenêtre de 2m de large offrant une ouverture de quelques 49° horizontalement et 26,8° verticalement (Figure 128). Cet environnement simplifié expose un décor comportant des indices de profondeur (plancher de référence, texture granulaire, volumes homogènes fournissant des repères, etc.).

le gradient de texture constitue un indice visuel de profondeur (Sheehan N., 1995) qui a motivé cette apparence des dalles du décor. Le choix du fond se base sur une recommandation (Ware *et al.*, 1993), relative au « *vection background* », pour éviter la sensation de vide caractérisé par une absence de repères visuels. Les repères utilisés (nuages) sont des indices connus qui situent les dimensions de la scène.

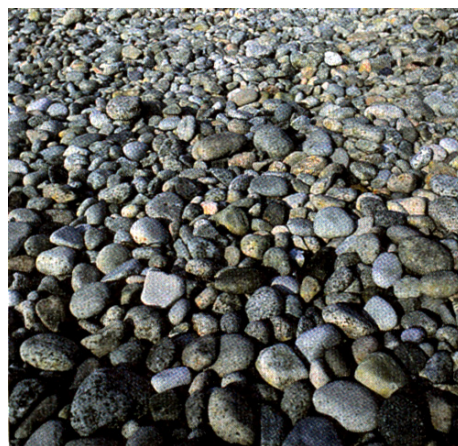


Figure 126. Exemple d'indice de profondeur, le gradient de texture.

Quelques objets distincts mais de nature homogène destinés aux opérations de déplacement et d'orientation sont disposés de part et d'autre de l'environnement sur le plancher. Ces objets devaient, de préférence, éviter de pouvoir être nommés, et présenter des formes géométriques les distinguant. Le choix s'est porté sur six des douze pièces du jeu de « *pentamino* »¹⁶, qui rappellent les motifs géométriques 3D utilisés par Shepard & Cooper (1986), pour leurs études des rotations mentales.

Un collimateur de visée piloté par la direction de la tête de l'utilisateur est représenté en permanence, ainsi qu'un retour visuel sur les poses que peut adopter la main de l'utilisateur au cours de la séquence d'interaction.

Ce dernier retour visuel n'est pas initialement conçu pour le participant, cette main virtuelle introduisant en effet une représentation distante adressant le « self ». Il n'est présent que pour informer le magicien qui observe également la scène, qu'un geste déictique est produit par l'utilisateur (raison pour laquelle celle-ci est figurée dans une « bulle », voir Figure 127).

¹⁶ Jeu sous forme de problème gigogne qui commence avec le nombre de figures qu'il est possible de former à l'aide de cinq carrés juxtaposés, figures ne différant ni par réflexion, ni par rotation ou retournement. Notamment mis en scène dans un roman de A.C. Clarke, « *Imperial Earth* », « Terre, planète impériale ».



Figure 127. Représentation de la « main » avec les trois états qu'elle peut adopter.

Parmi les trois états que cette « main » peut adopter, deux seulement sont utilisés pour l'expérience avec utilisateurs. La main au repos (Figure 127, à gauche), et en état « déictique », clairement identifiable à l'aide d'une couleur rouge (Figure 127, Figure 127 au milieu).



Figure 128. Scène 3D présentée aux utilisateurs sous forme de projection, sur laquelle se distinguent les objets sélectionnables pour subir des mouvements.

Sur cette représentation de la scène, (Figure 128), le "plancher" est damé de dalles de 2m de côté. Les cubes de repère disposés sur le plancher font tous un mètre d'arête. Les objets destinés à l'interaction (couleurs pastel) sont eux aussi constitués de cinq volumes cubiques d'un mètre d'arête.

Une information de distance apparaissant en réponse à une présélection d'objet, fournissant un indice supplémentaire déjà évoqué, fut par ailleurs envisagée et mise en oeuvre.

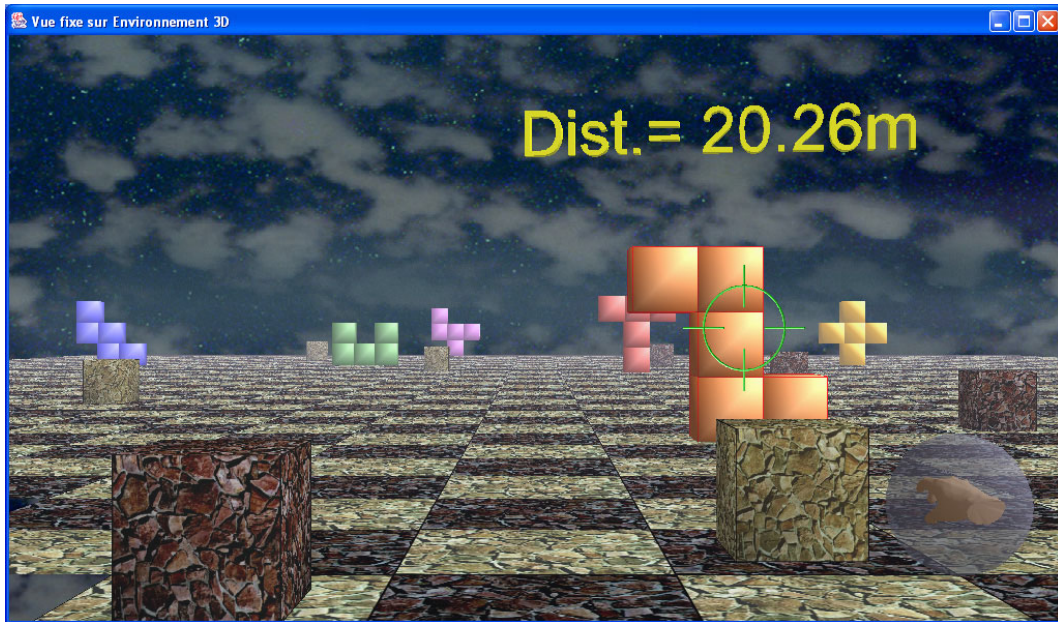


Figure 129. Présentation de la distance de l'objet à l'utilisateur en réponse à une présélection par visée.

Un affichage textuel de cette distance fut d'abord envisagé au-dessus de l'objet de la présélection (Figure 129, ci-dessus). L'affichage respectant ainsi la distance virtuelle, sa lisibilité s'avéra parfois malaisée (Figure 130), voire impossible en subissant une éventuelle obstruction (Figure 131).

Ces défauts de conception conduisirent à disposer la présentation de cette information à une distance constante à proximité du viseur (Figure 128, sous le collimateur de visée).

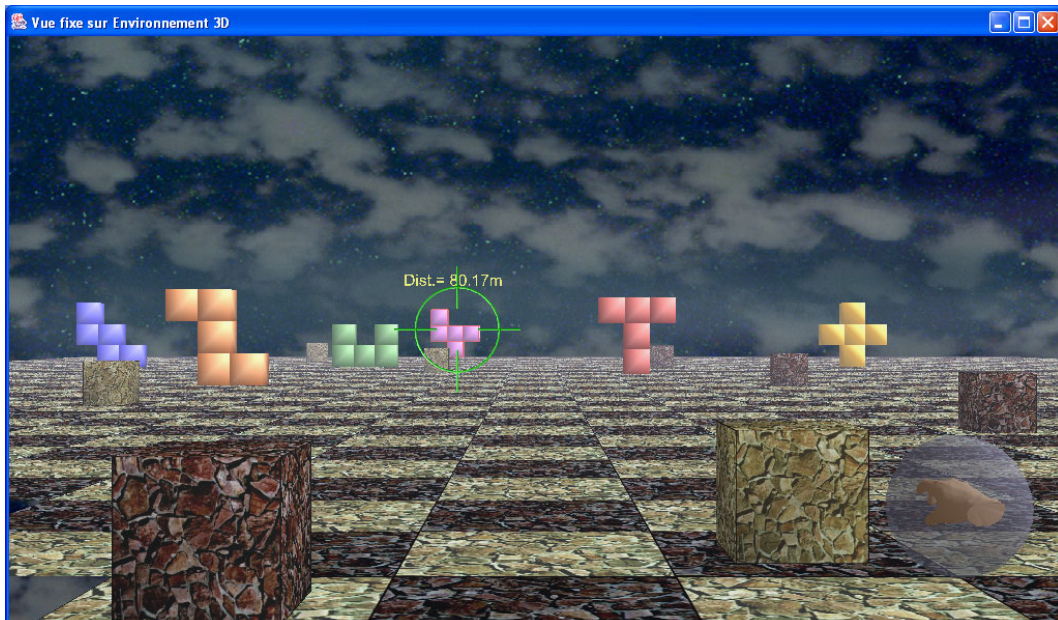


Figure 130. Difficulté de lecture dénonçant une présentation inadéquate de la distance de l'objet à l'utilisateur.



Figure 131. Obstruction de la présentation de la distance de l'objet à l'utilisateur.

Afin de simplifier les opérations de contrôle du magicien, et ce, conformément à l'objectif à atteindre pour chaque tâche proposée, certains des objets destinés à l'interaction furent munis d'un comportement de déplacement prédéfini activable à distance depuis la machine dont le « magicien » dispose. Ces comportements prédéterminés sont définis par les séquences d'interaction prévues pour la mise en scène et en fonction de l'affichage employé (en l'occurrence, un projecteur offrant un champ de vision couvrant 49° horizontalement et $26,8^\circ$ verticalement, impliquant par conséquent des déplacements circonscrits pour un résultat de déplacement visible par le participant).

Plateforme « magicien »

Le « *magicien* » dispose d'un panneau de commande qui se présente tel que figuré sur la Figure 132 page 187. Ce panneau est constitué d'un ensemble de « boutons » graphiques produisant une commande à distance et modifiant localement leur aspect à la manière d'un bistable en fonction de l'état dans lequel ils sont positionnés par le « *magicien* ». La ligne supérieure de boutons de contrôle généraux en comprend trois que nous allons expliciter en particulier.

En haut à gauche du panneau se trouve la commande d'enregistrement caractérisée par un disque rouge rappelant la signalétique correspondante du domaine audio-vidéo grand public.



Le bouton intitulé « **A** » commande la mise en rotation simultanée de tous les objets de la scène. Cette commande, initialement installée afin de permettre au magicien de vérifier le bon fonctionnement du système, s'avéra constituer une amélioration sensible de la mise en scène aux yeux des participants en début de séance d'expérimentation. Cette rotation simultanée des

objets sélectionnables de la scène était ensuite figée pour procéder à l'expérimentation.



Figure 132. Panneau de contrôle dont dispose le « magicien » pour piloter l'application répartie.

Le long bouton vert « *Commande non reconnue* » en haut du panneau fut conçu en tant que retour principal du magicien destiné à l'utilisateur et actionnant la synthèse

vocale d'un texte prédéfini sur la machine d'entrée des informations

de modalités. Retour qui se traduisait, auprès de l'utilisateur, par l'énoncé synthétique de la phrase « *Commande non valide, veuillez répéter* ».

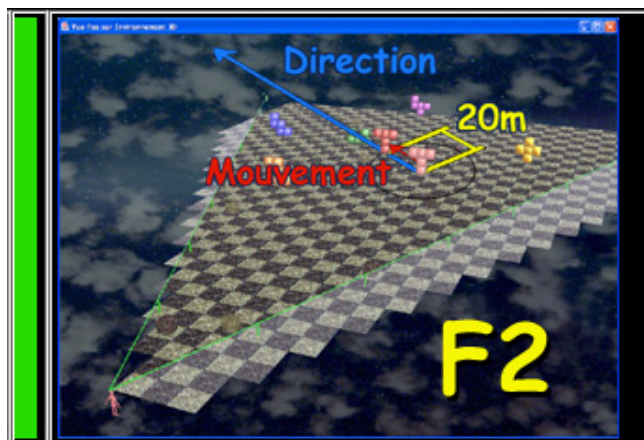


Figure 133. Détail des deux « boutons » d'une tâche.

Les autres boutons du panneau principal concernent plus particulièrement les tâches proposées. Pour chaque tâche (ex. Figure 133, pour la tâche « F2 » de déplacement exocentré), deux boutons de commande sont destinés l'un directement à la « *machine d'entrée* » des informations de modalités, l'autre indirectement à la

machine « 3D » gérant la scène. La Figure 133 montre le détail de l'un de ces couples de boutons. Le rectangle vertical vert sur la gauche représente un bouton dont l'action par un clic de souris engage, sur la « *machine d'entrée* » distante, la séquence de présentation de la tâche correspondante (d'abord vocale, puis graphique en affichant ensuite la représentation de la tâche correspondante). Le bouton de droite, de dimensions plus importantes, figure une miniature de la représentation de la tâche soumise au participant, et actionne le mouvement correspondant sur la scène 3D face au participant en réponse simulée par le système sous le contrôle du « *magicien* ». Une seconde action sur le bouton de droite de la tâche réinitialise la scène 3D et produit une synthèse vocale informant le participant quant à l'action en cours (« *Réinitialisation de la scène!* »). Cette réinitialisation s'avéra nécessaire afin de conserver pour chaque tâche une même situation de départ permettant leur présentation de façon homogène.

Un neuvième couple de boutons dont l'un représente une tâche non testée apparaît également sur le panneau de contrôle (reproduit Figure 134). Cette éventuelle tâche initialement baptisée « F8 » ne fut pas mise en œuvre pour cette première expérience, les commandes possibles étant identiques pour ces deux tâches distinctes de positionnement orthogonal angulaire.

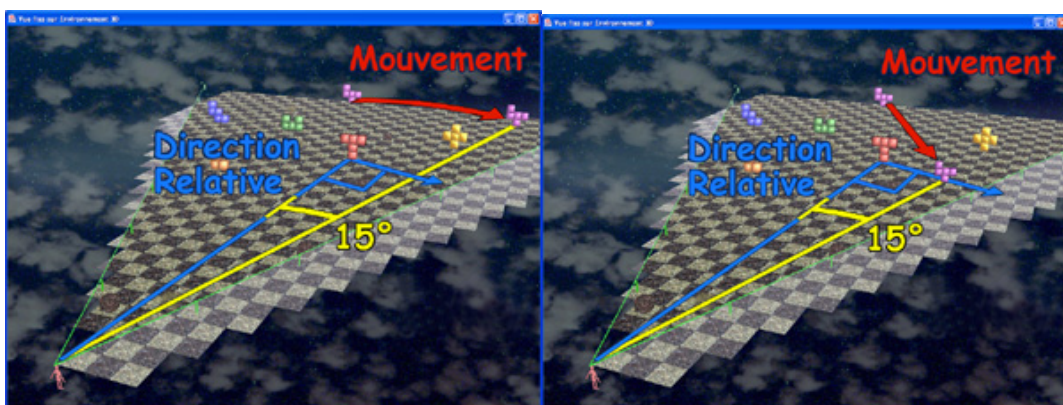


Figure 134. Images alternatives du « bouton » inemployé dans l'expérience pour la tâche initiale F8 de giration relative référencée à un second objet. (En bas au milieu du panneau de contrôle du magicien).

Parallèlement à ce panneau de contrôle de la machine « *ZardOz* », le « *magicien* » dispose également d'une régie d'enregistrement audio. En effet, afin d'entendre les énoncés synthétiques de la présentation de la tâche et le « beep » caractéristique disposé sur le récepteur RF pour suivre l'ensemble des opérations, le « *magicien* » dispose d'une réception audio radiodiffusée des événements sonores se produisant dans la salle d'expérimentation. D'autre part, afin d'entendre clairement les paroles du participant, une seconde réception audio radiodiffusée sur une fréquence distincte ne transmet que les productions d'énoncés de l'utilisateur à partir du micro-casque disposé sur sa tête. C'est cet usage de deux canaux de radiotransmission audio distincts qui justifia l'emploi d'une table de mixage, ainsi que la nécessité de disposer du signal résultant en sortie d'une part sur un casque stéréophonique porté par le

« *magicien* », et d'autre part en entrée de la machine « *ZardOz* » du « *magicien* » pour les seuls énoncés utilisateurs qui sont alors enregistrés localement sur disque dur.

Plate forme « d'entrée » d'informations de modalités

Le rôle de la « *machine d'entrée* » consiste d'abord à centraliser toutes les informations circulant pour en enregistrer certaines, en redistribuer d'autres.

Les informations enregistrées concernent les échantillons de position et d'orientation régulièrement produits par les deux capteurs magnétiques utilisés auxquels sont ajoutés un identifiant et une date système en millisecondes. Elles concernent aussi les informations de « poses » manuelles qui identifient un éventuel geste « déictique » de la main.

Cette machine reçoit d'autre part, par l'intermédiaire du récepteur RF, les commandes du « *magicien* » qui peut ainsi piloter l'ensemble de l'application. C'est à partir d'une information d'enregistrement du récepteur de commande que l'enregistrement de ces informations s'effectue ainsi que les mesures de durée des expressions pour assister « l'expérimentateur-magicien » qui peut alors opérer seul.

La « *machine d'entrée* » a également pour rôle de présenter les tâches à exécuter aux participants. Pour ce faire, elle dispose d'une synthèse vocale de textes prédéfinis guidant le participant de manière reproductible et dispose également d'un moniteur d'affichage des tâches proposées aux utilisateurs. Ce moniteur, baptisé « *écran d'instruction* » pour les besoins de l'expérience, est disposé à droite de l'écran affichant la scène 3D, et supporte les haut-parleurs utilisés pour reproduire la synthèse vocale.

Note

Si l'on compte également les systèmes à microcontrôleur au nombre des machines sur lesquelles est répartie l'application, nous avons également le « *récepteur de commande* » qui dispose de modalités primitives de présentation en sortie gérées par un processeur (« beeper » donnant le signal de départ au participant pour exprimer sa commande, et LED rouge signalant au « *magicien* » un enregistrement en cours). Ce récepteur contrôle également, à l'aide de sorties à collecteur ouvert, la mise en service de l'éclairage du participant à l'aide d'un faisceau étroit produit par un spot halogène basse tension suspendu au plafond de la salle d'expérimentation. Il contrôlait par ailleurs, initialement, la mise en service de la caméra vidéo transmettant ses images par radio.

L'émetteur de commande quant à lui dispose du même processeur mais n'effectue que des transmissions à l'instar d'une télécommande.

« L'application », dispositifs périphériques non comptés, repose donc en fait sur quatre machines de complexité distincte.

Equipement matériel

Techniquement, la « *machine d'entrée* » est un Mini-PC de type Shuttle SN41G2 équipé d'un processeur AMD Athlon XP2000+ et doté de 1 Go de RAM.

La machine « *3D* » est un Mini-PC de type Shuttle SB51G équipée d'un processeur Intel Pentium IV 2,4GHz et doté également de 1 Go de RAM.

La machine « *ZardOz* » est un PC portable Packard Bell équipé d'un processeur AMD Athlon XP2000+ et doté de 512Mo de RAM.

Le dispositif de capture magnétique est un « *Flock of Birds* » *d'Ascension Technology*, comprenant deux unités satellite « *Bird* » de capture dont l'une des deux tient le rôle de « *Master* » contrôlant un pulseur magnétique « *XMTR* ».

Le projecteur utilisé est un « *Data Projector* » PLUS U2-1110, technologie DLP (miroirs).

Le moniteur utilisé pour « l'écran d'instruction » est un Trinitron P900 Compac.

La camera vidéo à transmission radio est une microCam AV à objectif réglable émettant sur 2412MHz, conçue selon notice pour couvrir des distances de l'ordre de la centaine de mètres. Le récepteur vidéo correspondant est un récepteur AV à 4 canaux à sortie vidéo composite 1Vcc sur prise RCA. (Le convertisseur Vidéo composite – S-VHS utilisé est un modèle courant).

La caméra MiniDV est une « *Digital Handycam* » SONY DCR-VX1000E.

Les deux émetteurs audio radiodiffusé sont des émetteurs professionnels Monacor TXS-822LT (864,8MHz) et TXS820LT (863,05MHz). Les récepteurs audio correspondants sont des récepteurs Monacor TXS-802 (864,8MHz) et TXS-800 (863,05MHz).

La table de mixage est une Promix 100 à 6 canaux, dont 2 entrées LINE, 2 entrées Microphone, et sorties séparées.

Le casque utilisé par le magicien est un Denon AH-D550.

Les haut-parleurs utilisés pour la production de messages vocaux de synthèse sont des mini-enceintes actives multimédia Altec Lansing ACS43.

Les dispositifs additionnels sont tous basés sur la mise en œuvre de microcontrôleurs de type Basic Stamp II de Parallax.

Les logiciels multimedia utilisés sont Adobe Première 6, Photoshop 5, Cool Edit 2000.

6.3.5. Montage expérimental

Le montage expérimental se présente visuellement de la manière représentée sur la Figure 135.

Le laboratoire est constitué de deux pièces séparées par une baie vitrée semi transparente. Dans la « *salle d'expérimentation* », (de l'autre côté de la baie vitrée sur cette Figure 135), l'utilisateur, équipé des dispositifs de suivi de mouvements et d'un microphone, fait face à une projection de la scène 3D sur grand écran. A sa droite se trouve un « *écran d'instruction* » sur lequel s'affichent les objectifs à atteindre.



Figure 135. Montage expérimental mis en oeuvre pour étudier les expressions utilisateur.

De l'autre côté de la baie vitrée, en premier plan à gauche sur cette même Figure 135, le « *magicien* » observe le déroulement des opérations et des expressions du participant, et pilote l'application en conséquence à l'aide de la machine « ZardOz » décrite et de l'équipement audio et RF qui lui font face (devant lui, à gauche sur le montage, se trouve également l'émetteur de radiotransmission des commandes du système). Le boîtier bleu et le moniteur sur la table contiguë à celle qu'occupe le « *magicien* » figure l'équipement initial d'enregistrement vidéo, transmis par radio (qui fut, par la suite remplacé au profit d'une caméra miniDV).

Sur cette Figure 135, l'équipement de l'utilisateur est figuré en bleu (câbles débordant du dossier de la chaise). L'équipement magnétique est relié au travers de ces câbles (bleus) à deux boîtiers de traitement de ces informations (empilés, de couleur crème). Ces boîtiers transmettent ensuite les position-orientations des

capteurs magnétiques à la « machine d'entrée » (signalée par l'appellation « INPUT » sur la figure).

A côté de ces deux boîtiers empilés en figure un troisième, monté sur un mât, qui constitue l'émetteur de champ magnétique.

Le « *gant de données* » porté par l'utilisateur est, quant à lui, directement raccordé à cette dernière machine. Cette « *machine d'entrée* » supporte par ailleurs un boîtier qui figure le récepteur RF de commande. Le câble vert (vidéo) qui sort de la « *machine d'entrée* » (« INPUT ») contrôle « *l'écran d'instruction* » d'affichage des tâches et les haut-parleurs produisant les messages synthétiques audibles. Elle transmet d'autre part (câble rouge) les informations de position-orientation à la machine « 3D » qui elle-même contrôle le projecteur affichant la scène 3D.

« *L'écran d'instruction* » supporte, sur cette figure, deux haut-parleurs de type enceintes actives multimedia qui encadrent une caméra d'enregistrement.

6.3.6. Fonctionnement de l'ensemble

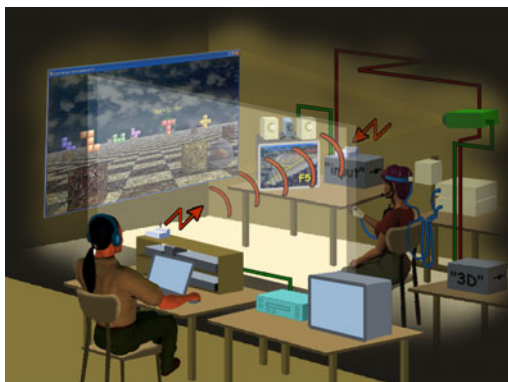


Figure 136. Emission de commande.

Nous allons détailler brièvement le déroulement d'une séquence expérimentale avec un participant.

Le « *magicien* » démarre la séquence de présentation de la tâche en cliquant sur le bouton vert de la tâche du panneau de commande (Figure 136), lequel bouton devient rouge sur le panneau. La « *machine d'entrée* » débite alors sur haut-parleurs son message vocal signalant au participant qu'une tâche va lui être présentée (Figure 137).

"Affichage d'une tâche sur l'écran d'instruction!".

Le système demande ensuite préalablement au participant de prendre le temps d'examiner la situation qui va être exposée.

"Veuillez observer le mouvement à obtenir!".

Il demande d'autre part à celui-ci de signaler lorsqu'il sera prêt à exprimer sa commande. *"Lorsque vous êtes prêt, dites - Je suis prêt! - . Vous pourrez ensuite formuler votre expression après le bip sonore!"*.

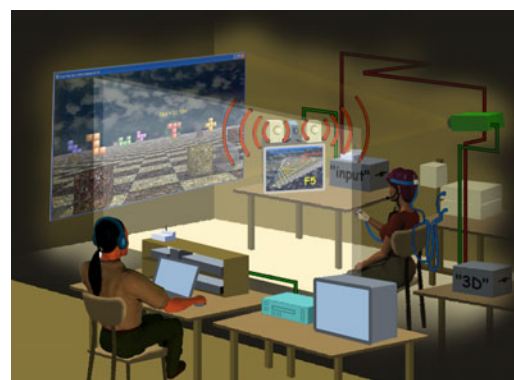


Figure 137. Emission d'un message vocal de synthèse du système.

Le « *magicien* » reste alors à l'écoute et en attente du message convenu par lequel le participant signalera qu'il est prêt.

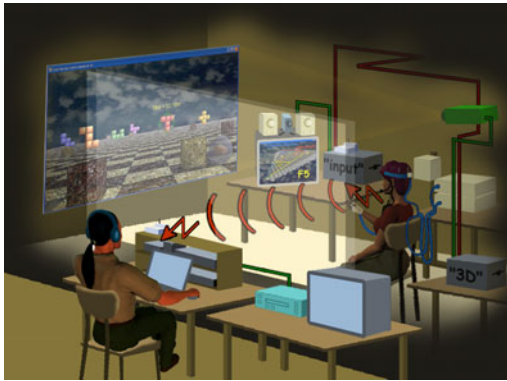


Figure 138. Transmission des messages vocaux du participant par radiotransmission.

Alors seulement la « machine d'entrée » affiche-t-elle la représentation de la tâche sur « l'écran d'instruction » que va examiner le participant.

Lorsqu'est prononcé le message attendu, ("*Je suis prêt*", transmis au « *magicien* » par l'émetteur audio raccordé au micro-casque du participant, Figure 138), le « *magicien* » actionne la commande de mise en service de l'enregistrement. Enregistrement dont il peut vérifier la mise en service en observant le boîtier de réception RF qui allume sa LED

rouge, et qui émet un « beep » long destiné à signaler au participant qu'il peut s'exprimer.

A partir de ce moment, la « machine d'entrée » enregistre les mouvements de l'utilisateur. Pendant que la machine du « *magicien* » enregistre, de son côté sur disque, les énonciations vocales du participant, transmises sous forme analogique par radiotransmission.

D'où il se trouve, le « *magicien* » peut donc contrôler le déroulement des opérations, observer ce que fait l'utilisateur, ce que ce dernier vise sur la projection de la scène 3D, déceler sur la représentation de la scène tout geste déictique de l'utilisateur (« main » affichée devenant rouge dans sa « bulle »), et entendre ses énoncés.

En fin d'énoncé, si la commande lui semble pertinente (complète, décrivant effectivement la tâche à réaliser, etc.), le « *magicien* » actionne le bouton d'exécution de la tâche correspondante, le mouvement de l'objet 3D de la tâche est exécuté. Il actionne ensuite le bouton de mise en service des enregistrements, ce qui a pour effet d'arrêter ceux-ci et d'éteindre la LED rouge sur le récepteur RF de commande. Ensuite, il actionne de nouveau le bouton du panneau correspondant à l'exécution de la tâche, la « machine d'entrée » débite alors le message vocal : "*Réinitialisation de la scène!*", pendant que la machine graphique (« 3D ») repositionne l'objet 3D à sa place initiale. Pour terminer la séquence, le « *magicien* » actionne finalement de nouveau le bouton de commande de présentation de la tâche sur « l'écran d'instruction » (lequel, initialement vert, était passé à une couleur rouge, redevenant désormais vert), ce qui produit un retrait de l'image de présentation de la tâche sur « l'écran d'instruction », remplacée par l'image initiale montrant au participant où il se situe dans la vue aérienne (cf. Figure 112 page 173).

Déroulement d'une passation

Les utilisateurs sont accueillis au laboratoire où leur est soumis un questionnaire (« Questionnaire », en annexe), les informant des dispositions du recueil de données les concernant (notamment la confidentialité des informations individuelles, et l'autorisation de pratiquer divers enregistrements), réunissant des informations d'âge, sexe, problèmes de vue ou d'audition, expérience avec des affichages ou des dispositifs d'interaction non conventionnels. Il leur est par ailleurs demandé de tirer successivement les huit cartes d'un jeu préalablement mélangé par leurs soins, jeu dont chaque carte porte l'identifiant d'une tâche utilisateur (« F1 » à « F7 » + « F9 »).

Il leur est ensuite remis un document de présentation de l'expérience (« Consigne aux participants », en annexe).

Ce document leur présente l'environnement d'interaction en décrivant la multimodalité en entrée et en insistant sur les tâches atomiques de présélection et de sélection, en restant toutefois le plus abstrait possible en ce qui concerne les énonciations qu'ils seront amenés à produire. Ce document insiste par ailleurs sur la nécessité de produire une expression de commande complète.

La durée de consultation de ce document n'est pas limitée.

Les participants sont ensuite équipés par l'expérimentateur des dispositifs d'entrée et placés dans les conditions d'interaction décrites dans le document de consignes.

L'expérimentateur se retire ensuite, prétextant la nécessité de contrôler le bon déroulement des opérations d'enregistrement, afin d'endosser de plus le rôle du « *magicien* » tenant le rôle de contrôleur de dialogue et d'orchestrer le bon déroulement de l'expérimentation, secondé en cela par le système en ce qui concerne la présentation des tâches utilisateur.

6.4 Mesures

Afin de déterminer la pertinence des associations de modalités à permettre effectivement l'exécution de diverses tâches, les mesures envisagées sont essentiellement des mesures de performance utilisateur (pourcentages de réussite dans la réalisation de tâches, pourcentage d'échecs, durée de complétion, etc.), ainsi que des mesures de satisfaction pour vérifier l'acceptabilité perçue auprès de l'utilisateur.

Certaines mesures sont effectuées par le système digital d'enregistrement des événements de modalités transmis par les dispositifs d'entrée, tant en ce qui concerne par exemple les durées de complétion que le comptage automatique des durées de gestuelles déictiques et mimétiques.

Les enregistrements étant placés sous le contrôle du « *magicien* », l'application d'enregistrement est conçue de manière à produire, dans le fichier XML d'enregistrements, des commentaires récapitulatifs des comptages de durées totales d'enregistrement et « d'exposition » de l'utilisateur à la présentation de la tâche avant qu'il ne se déclare « prêt » à s'exprimer. Cette durée est donc incluse dans la durée

d'enregistrement, dont on peut alors déduire une « durée d'expression » approximative qui comprend également le temps de réaction du « *magicien* » pour arrêter l'enregistrement.

Pour illustrer ces comptages des durées par le système, nous allons prendre un exemple. Lorsque le « *magicien* » initie la séquence de présentation d'une tâche, la « machine d'entrée » (qui est également celle qui va enregistrer les informations digitales) commence par prévenir oralement le participant qu'une tâche va lui être présentée. Lorsque cette machine termine son discours en demandant au participant de bien vouloir prévenir lorsqu'il sera prêt, elle commence un comptage de durée. Pendant ce temps, l'utilisateur observe « *l'écran d'instruction* » pour s'informer de la tâche à accomplir.

Lorsque le participant s'estime prêt à s'exprimer, il annonce « *Je suis prêt!* », le « *magicien* » actionne alors la commande d'enregistrement. Le récepteur de contrôle émet son « beep » signalant au participant qu'il peut commencer à s'exprimer. Dans le même temps, la « *machine d'entrée* » commence à effectuer l'écriture des informations de capture provenant des dispositifs d'entrée, alors que la machine « *ZardOz* » du « *magicien* » enregistre de son côté les signaux audio provenant de l'utilisateur.

6.4.1. Mesures effectuées par le système

Examinons ensuite la Figure 139. Le chronogramme qui figure sur cette illustration est un exemple constitué à l'aide d'un enregistrement d'un énoncé. On voit d'abord, à gauche sur cette figure, le « beep » produit par le récepteur de commande. Ce signal sonore succède donc à l'énonciation « *Je suis prêt!* » du participant (par conséquent non représentée sur ce chronogramme). Ce même signal sonore est présent sur cette figure car l'enregistrement proprement dit a déjà commencé.

On voit ensuite s'écouler une durée après le « beep » (commentée en fond bleu « durée de préparation »), pendant laquelle on estime que l'utilisateur prépare sa formulation.

On constate ensuite des vocalisations qui correspondent à l'énonciation séquentielle de la commande (commentée sur fond vert).

Lorsque l'utilisateur termine son énoncé, le « *magicien* » doit estimer si la commande est complète et arrêter l'enregistrement le cas échéant (dans l'autre cas, l'enregistrement se poursuit par une demande de reformulation). Il s'ensuit une durée non négligeable de réaction du « *magicien* » (commentée sur fond jaune sur cette Figure 139) qui s'additionne au sein de l'enregistrement.

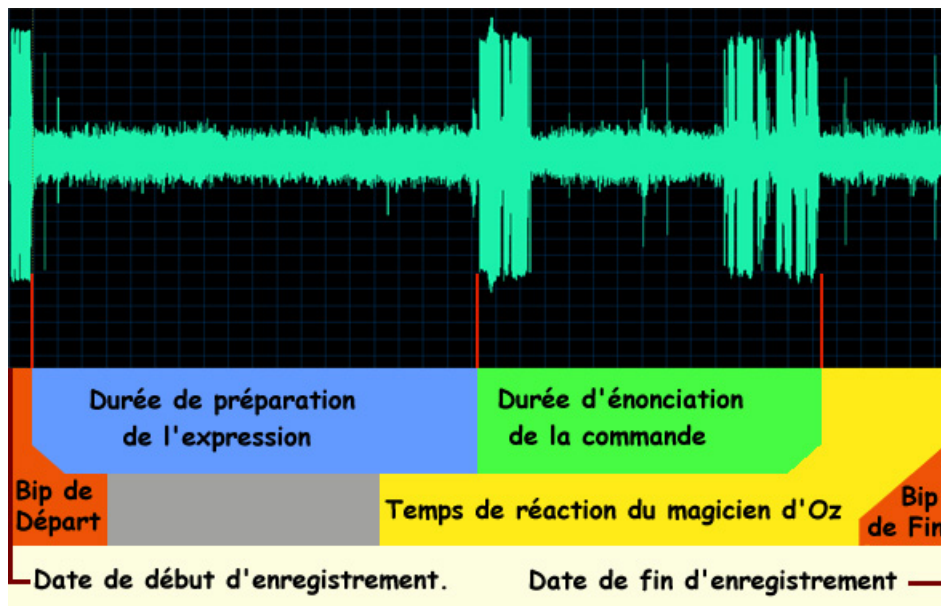


Figure 139. Illustration commentée d'un énoncé utilisateur au cours d'un enregistrement.

Voyons maintenant un extrait d'un enregistrement en ce qui concerne ces informations.

```
<!-- Manip n° = N4 -->
<!-- "F2" -->
<!-- Exposition Dtes= 1091018671453 1091018698203 Duree= 26750 ms -->
<!-- Enreg Dtes= 1091018698203 1091018707218 Duree= 9015 ms -->
<!-- Nb deict= 1 -->
<!-- Deict 1 Dte= 1091018703406 1091018705296 DRelDeb= 5203
DRelFin= 7093 Duree= 1890 ms -->
```

Figure 140. Extrait d'un enregistrement XML, illustrant le chronométrage effectué par le système.

L'extrait au format XML sur la Figure 140 expose des commentaires enregistrés à la fin de chaque enregistrement de tâche au sein du fichier.

Ces champs de commentaires sont provisoirement conçus à des fins d'éditations manuelles de la part de l'expérimentateur. Ils rappellent d'abord le quantième de la tâche (*Manip n° =*), puis son identifiant ("*F2*"), et fournit ensuite des dates et durées.

Les dates « d'exposition », exprimées en millisecondes, correspondent respectivement à l'instant de l'affichage de la tâche sur « l'écran d'instruction », et au moment succédant à la commande de début d'enregistrement de la part du « *magicien* ». Cette dernière date constitue également la date de début d'enregistrement (*Enreg*), cependant que la seconde date est l'instant d'interruption de l'enregistrement. Ces trois informations de dates fournissent deux durées, l'une « d'exposition » de l'utilisateur à la présentation de la tâche, et l'autre, d'enregistrement des informations digitales, qui contient également le temps de

réaction du « *magicien* ». Il est par la suite apparu que ces informations ne présentaient qu'une utilité relative et qu'il restait préférable d'estimer les durées d'énoncés à l'aide d'un éditeur audiographique (CoolEdit 2000).

A la suite de ces informations de durée, se trouve un dernier champ de commentaires qui concerne cette fois les « poses » adoptées par la main de l'utilisateur au cours de la période d'enregistrement. Le système étant informé des signaux produits par le « gant de données », il est en effet en mesure de les dater et d'en estimer la durée comme précédemment. Les dates système étant peu représentatives, on a ajouté une transcription en dates relatives de début et de fin d'indication de direction manuelle par rapport au début de l'enregistrement (*DRelDeb* et *DRelFin*). On constate ainsi sur cet exemple que l'utilisateur a commencé un geste déictique 5,2 secondes après le début de l'enregistrement, et que ce geste a duré presque deux secondes.

6.4.2. Retranscription d'informations enregistrées

L'automatisation des mesures constituant un développement en soi, la majorité des mesures sont effectuées *a posteriori* sur la base des enregistrements pratiqués.

Ainsi, l'une des premières opérations consiste à retranscrire textuellement les énoncés des utilisateurs. On conçoit ici l'intérêt d'une mise en œuvre d'une reconnaissance vocale fonctionnelle simultanément aux productions d'énoncés, ou encore à laquelle soumettre ultérieurement les enregistrements, qui impliquerait d'être multi-locuteur sans nécessiter d'entraînement préalable. En l'absence d'une telle application, les retranscriptions sont exécutées de manière traditionnelle pour chaque tâche et chaque énoncé, afin d'obtenir une distribution syntaxique des termes énoncés. Cette manière traditionnelle consiste à réécouter les énoncés enregistrés, puis à en écrire les contenus. Le Tableau 7 ci-dessous montre un exemple d'une telle transcription pour cinq utilisateurs effectuant la tâche F2 de déplacement avec indication manuelle de direction exocentrée.

D'autres informations sont ensuite reportées dans deux tableaux sous forme binaire sur la base des transcriptions ainsi obtenues, l'observation des vidéos ainsi que celle des restitutions multimedia (décrites plus loin).

Le Tableau 8 page 198 montre un tel tableau utilisé à cet effet. Le compléter revient à répondre à une somme de questions telles que « l'expression a-t-elle conduit au succès ? », « Si elle était incomplète, l'opération manquait-elle ? », etc.

Utilisateur / F2	Opération	Sélection Objet d'opération	Préposition + Valeur + Unité	Direction	Nbre d'essais avant Succès
SIO3D_MM-F2	Déplace	cet objet	20m	dans cette direction.	0
SIO3D_CV-F2	déplacer,	Cet objet,	de 20m,	dans cette direction.	1
SIO3D_BC-F2	Déplacer	cet objet	de 20m	à gauche.	0
SIO3D_BJ-F2	déplacer,	Cet objet,	de 20m,	ici	1
SIO3D_LS F2	Déplacer	cet objet	de 20m	dans cette direction	1

Tableau 7. Exemple de distribution syntaxique des productions orales de quelques participants.

Un second tableau binaire est également utilisé pour caractériser le comportement gestuel du participant au cours de son expression pour chaque tâche (reproduit Tableau 9 page 199).

Une expression de commande est ainsi caractérisée à l'aide d'un « profil » sous forme d'une suite de nombres binaires.

Tous les profils obtenus pour une même tâche par tous les participants des deux groupes sont ensuite insérés dans une feuille de calcul. Facilitant ainsi l'établissement de sommes, moyennes, variances, écarts-types, et t-test le cas échéant.

Tâche	Deict	DEG	EIC	EIP	EIQ	EID	ECA	ECI	VS00	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn	NbS	
F1											0										
F2		0																			
F3																					
F4																					
F5																					
F6																					
F7											0										
F9		0																			

Tableau 8. Tableau binaire reflétant les caractéristiques des énoncés d'un participant.

DEG: Direction manuelle indiquée de visu par l'utilisateur mais sans usage correct du dispositif de capture (défaut de conception du "gant" ou geste déictique sans contact du pouce sur la deuxième phalange du majeur). Le cas échéant, ce défaut du gant ne devrait apparaître que pour les opérations impliquant un acte déictique de la main droite, (c'est à dire F2 et F9).

EIC: Enoncé incomplet: Opération absente

EIP: Enoncé incomplet: Absence de **préposition**

EIQ: Enoncé incomplet: Quantité de mouvement absente

EID: Enoncé incomplet: Direction absente ou incomplète

ECA: Enoncé complet: Mais correspondant à une **autre tâche** (En cas d'incompréhension ou de confusion sur la tâche proposée, ou de confusion en cours d'énonciation suite à surcharge

ECI: Enoncé complet: Sans Identification possible d'une destination claire.

VSOO: Variation d'ordre syntaxique: Inversion de l'ordre syntaxique d'indication de l'**opération** et de la **sélection de l'objet** du mouvement.

VSQD: Variation d'ordre syntaxique: Inversion de l'ordre syntaxique d'indication de la **quantité** de mouvement et de la **direction**.

ALDP: Abus de langage: Usage de **verbes** dédiés au **déplacement** en lieu et place de verbes dédié au **positionnement** (stricto sensu).

Cet abus ne saurait apparaître que pour les tâches de positionnement non ambiguës (F1, F7).

Toutefois, F4 peut, par son absence de spécification de quantité de mouvement, apparaître aussi bien comme positionnement que comme déplacement.

ALDI: Abus de langage: Usage d'une **partie** seulement de l'énoncé permettant cependant l'indication complète d'une **direction**.

UFPI: Confusion d'usage de la langue française: Usage d'une **préposition inappropriée**.

UFVI: idem: Usage de **verbes inappropriés** pour la tâche proposée, verbe signifiant une autre opération cependant possible.

UFLI: idem: Usage de **latéralité inversée** (usage de "à droite" pour "à gauche" et inversement).

TID: Addition de termes inutiles à l'énoncé: Usage d'un acte déictique manuel non requis pour l'expression. Les actes déictiques manuels d'indication d'une direction ne sont strictement nécessaires que pour les tâches F2 et F9.

NF: Erreurs d'usage de la langue française: (l'énoncé n'est pas conforme à la langue)

InS: Insuccès. Pas d'énoncé permettant de situer la destination du mouvement.

TIn: Tâche incomprise. L'observation de la tâche proposée n'a pas permis de comprendre le mouvement à effectuer.

NbS: Nombre d'essais avant succès par énoncé d'une expression permettant de situer la destination du mouvement.

Composition non verbale de l'expression.

Dans l'expression d'une tâche à accomplir, l'utilisateur devrait avoir mis en oeuvre :

Tâche	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de destination par visée	Indication manuelle de direction de destination	Indication d'un second objet par visée	Indication de second objet s'accompagnant d'une Indication manuelle de direction
F1	1		1			
F2	1			1		
F3	1					
F4	1		1			
F5	1					
F6	1				1	
F7	1				1	
F9	1			1		

Tableau 9. Tableau binaire reflétant le comportement gestuel manifesté au cours de l'exécution d'une tâche.

6.4.3. Restitutions multimedia

Parallèlement aux mises au point de l'expérience, une application de restitution multimedia 3D fut développée conjointement dans le cadre d'un stage de DEA d'un étudiant en « Communication Multimédia » du CNAM.

L'objectif de ce stage consistait initialement à obtenir *a posteriori* une représentation simultanée des productions des participants sous forme visuelle segmentée par modalité d'entrée et modalité d'observation enregistrées.

Ainsi, le système de restitution pouvait-il être envisagé pour restituer de manière synchrone, concernant la passation d'une tâche d'un utilisateur, une reproduction de la scène 3D augmentée d'une représentation de l'utilisateur et de ses actions gestuelles, la vidéo et l'enregistrement audio réalisés au cours de la tâche utilisateur, et dans l'idéal, une présentation graphique textuelle des énoncés au cours du déroulement de l'enregistrement.

Il s'avéra rapidement qu'une machine classique de type PC manquait nettement de puissance pour gérer à la fois ne serait-ce qu'une scène 3D, une vidéo et une restitution sonore. L'effort porta donc sur une orientation « production multimedia » du projet d'application.

Les deux composants principaux de cette application sont un parser XML et l'application d'affichage de la scène 3D modifiée.

Le *parser* XML a pour rôle de lire les enregistrements digitaux réalisés par la « *machine d'entrée* » au cours des passations des utilisateurs, et de les transmettre à une scène 3D modifiée qui les interprète pour représenter sur l'affichage les événements de gestuelle des participants.

Cette application reproduit la situation expérimentale dans laquelle se trouve l'utilisateur au cours de sa production d'expression en vue de la réalisation d'une tâche proposée sur l'écran d'instruction sur sa droite.

La scène 3D originale (*cf.* Figure 128 p.184), est en effet modifiée en sorte de pouvoir y représenter l'utilisateur sous forme d'une marionnette (Figure 141 ci-après, p. 201).

La marionnette occupe en effet la position virtuelle de l'utilisateur, cependant que le point d'observation de la scène est reculé et rehaussé de façon à obtenir une vue plongeante à la fois sur la marionnette en premier plan, tout en conservant une vue sur la scène 3D en arrière-plan. La marionnette, privée de son bras droit afin d'éviter les effets d'obstruction visuelle, permet de suivre la position de la main de l'utilisateur.

La tête et la main droite de cette marionnette sont alors animées au gré des informations datées de positions et d'orientations enregistrées au cours de l'expérience.

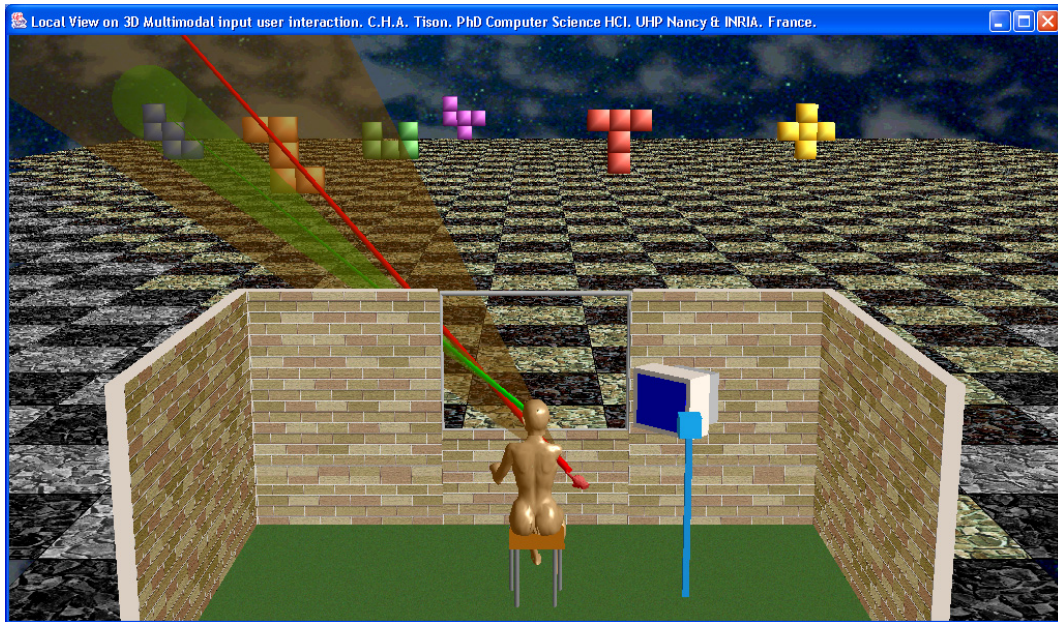


Figure 141. Restitution multimedia des séquences d'expression utilisateur.

L'utilisateur est ainsi représenté de dos face à une « fenêtre » lui donnant vue sur l'environnement qui lui est présenté par projection dans la réalité de l'expérience (les murs matérialisent l'enceinte du laboratoire). La direction de la tête de l'utilisateur est représentée de façon constante sous la forme d'un faisceau de couleur verte. Lorsqu'un geste déictique est réalisé avec la main droite, un second faisceau, plus évasé afin de représenter l'imprécision plus importante d'une direction indiquée manuellement, est également figuré en rouge.

A droite de la « fenêtre sur le monde » est également représenté « l'écran d'instruction » (bleu royal), que l'on peut voir l'utilisateur consulter de nouveau après une expression incomplète, lorsqu'on lui demande de reformuler son expression. L'extrait particulier illustrant cette « restitution » (Figure 141 ci-dessus) montre, lors d'une sélection de l'objet le plus à gauche de la scène (de couleur bleue), la redondance entre sa désignation à l'aide de la direction de la tête et l'indication supplémentaire réalisée à l'aide d'un geste déictique de la main.

L'application de restitution se voit donc pourvue des informations enregistrées au cours de l'expérience. Ces informations étant datées, il devient possible de n'envoyer à cette nouvelle scène 3D que les événements de gestuelles produits par l'utilisateur par intervalles de 40 millisecondes, puis d'effectuer une copie de l'image obtenue de la scène.

La répétition de ce processus par périodes successives de 40 millisecondes jusqu'à la fin du fichier d'enregistrement permet alors d'obtenir une succession de représentations visuelles de l'utilisateur en interaction avec la scène 3D, régulièrement espacées dans le temps.

Un montage de ces représentations visuelles à raison de 25 images par secondes à l'aide d'un éditeur multimédia (tel qu'Adobe « Première »), fournit finalement une

animation à laquelle peut être ajoutée le fichier sonore des énoncés correspondant à la même tâche utilisateur.

Le résultat obtenu peut ensuite être restitué autant de fois que nécessaire à l'aide d'une application « *player* » multimédia, et constitue ainsi une aide à l'observation ultérieure des comportements des participants au cours de chaque tâche utilisateur.

Chapitre 7. Résultats

7.1 Performance globale des deux groupes d'utilisateurs aux tâches proposées.

D'un point de vue quantitatif, réunir par groupe et pour chaque tâche au sein d'un même tableau le nombre d'expression ayant donné lieu à un « succès » (un "succès" se définissant par la production d'une spécification complète et correspondant effectivement à la tâche proposée) nous permet d'estimer les résultats obtenus par chaque groupe.

Tâches	Groupe Informaticiens	Groupe Non-Informaticiens
F1	8	3
F2	6	3
F3	8	5
F4	6	8
F5	7	6
F6	8	3
F7	8	5
F9	7	2
Somme	58	35
Moyenne (nb/10)	7,250	4,375
ET	0,886	1,995
Student (p)	0,004192	

Tableau 10. Effectif des expressions donnant lieu à succès pour chaque tâche proposée et pour chacun des deux groupes de dix utilisateurs informaticiens et non-informaticiens.

Test de l'hypothèse selon laquelle il n'y a pas de différence entre les deux groupes pour l'obtention de succès aux tâches proposées.

L'hypothèse H0 selon laquelle il n'y a pas de différence entre les deux groupes pour l'obtention de succès aux tâches proposées est rejetée, il apparaît une différence significative entre les deux groupes « Informaticiens » et « Non Informaticiens ».

Les 72,5% d'expressions correctes chez les informaticiens montrent en particulier, et par opposition aux résultats obtenus par le groupe des « non informaticiens » (43,75%), qu'une telle interaction est exploitable sans formation particulière en ce qui les concerne.

Rappelons que ce résultat est obtenu en l'absence de période de familiarisation avec le système ou d'exemples concrets ayant facilité la production de commandes pertinentes. Les utilisateurs, pour cette expérience, ne reçoivent qu'un ensemble de consignes volontairement abstraites pour éviter toute influence sur leurs énoncés (mises à part les opérations atomiques de l'interaction telles que sélection d'objet ou de direction).

L'examen du Tableau 10 de répartition des « succès » par groupe d'utilisateurs permet de constater que l'une des deux tâches pour lesquelles le groupe des informaticiens obtient le moins de résultats conduisant au succès (F4) est aussi celle pour laquelle le groupe des non informaticiens obtient le meilleur score, seule tâche pour laquelle ces derniers surpassent d'ailleurs les informaticiens.

On note d'autre part que pour chacun des deux groupes, les scores obtenus pour certaines tâches sont situés en deçà d'un écart autour de la moyenne. C'est le cas chez les informaticiens pour les tâches F2 et F4. C'est également le cas chez les non informaticiens pour la tâche F9 de rotation d'un objet.

Pour les non informaticiens, trois tâches situées dans l'étendue au dessus de la moyenne présentent un point commun qui explique peut-être une similitude de résultats (F3, F5 et F7). Ces trois tâches impliquent en effet la mise en oeuvre de directions relatives énoncées oralement, ce qui militerait pour l'hypothèse d'une "économie de modalités" mise en oeuvre au sein des expressions en faveur des modalités vocales. Ce résultat ne concerne toutefois que le groupe des non informaticiens.

On sait, à la lecture des commentaires post-expérimentaux formulés par les participants, que les tâches F4 et surtout F5 (cf. Figure 142 et Figure 143) ont subi l'influence d'une présentation des tâches induisant une erreur d'interprétation chez au moins trois utilisateurs ayant exprimé une commande de mouvement orbital centré utilisateur, ce biais expérimental permet de relativiser quelque peu les résultats obtenus à ces tâches.

Parmi ces trois utilisateurs, deux sont informaticiens. Le troisième est non informaticien.

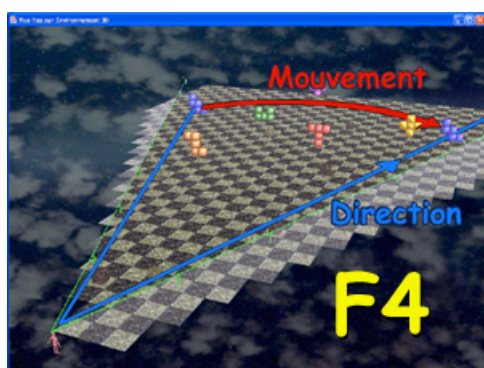


Figure 142. Représentation de la tâche « F4 ».

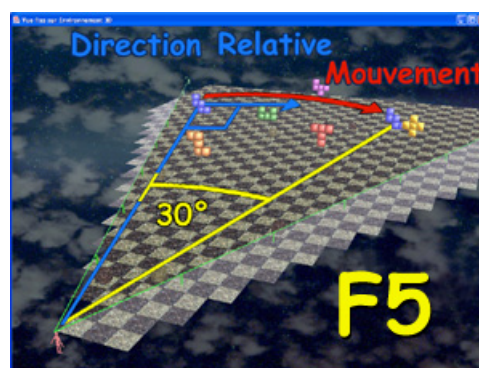


Figure 143. Représentation de la tâche « F5 ».

La tâche F4, présumée la plus déroutante au départ de la conception de l'expérience (du fait de l'absence de quantité requise pour l'énoncé de sa commande), est *a contrario* celle qui, globalement d'une part, en particulier pour le groupe des non informaticiens d'autre part, obtient le plus de réussite.

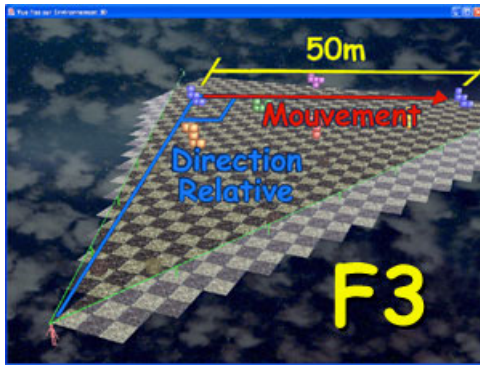


Figure 144. Représentation de la tâche « F3 ».

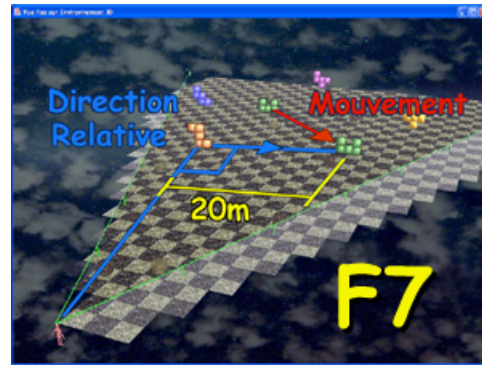


Figure 145. Représentation de la tâche « F7 ».

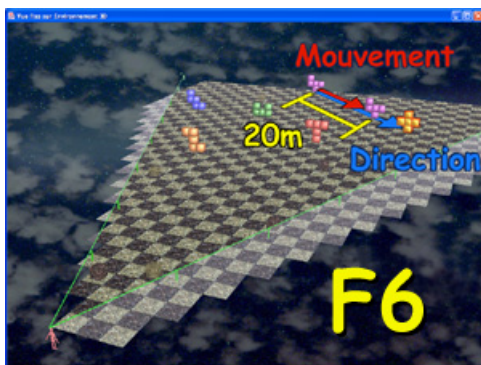


Figure 146. Représentation de la tâche « F6 ».

Nous savons d'autre part, à la lecture des commentaires post-expérimentaux, qu'au moins trois utilisateurs (et deux de plus qui les sous-entendent), ont cherché à résoudre les tâches les plus complexes telles que F6 et F7 (faisant référence à un second objet référent du mouvement, *cf.* Figure 146 et Figure 145), en plusieurs sous tâches de déplacement plus simples visant à obtenir un résultat équivalent.

Nous savons également, que pour F2 (*cf.* Figure 147), quand bien même son utilité est discutable, l'absence de retour visuel de la direction indiquée par un geste déictique manuel a constitué une gêne importante chez plusieurs utilisateurs. Cette absence de retour visuel peut s'avérer être à l'origine d'une certaine incompréhension de sa mise en oeuvre.

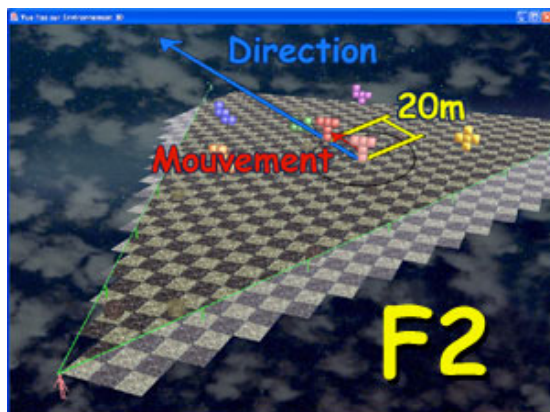


Figure 147. Représentation de la tâche « F2 ».

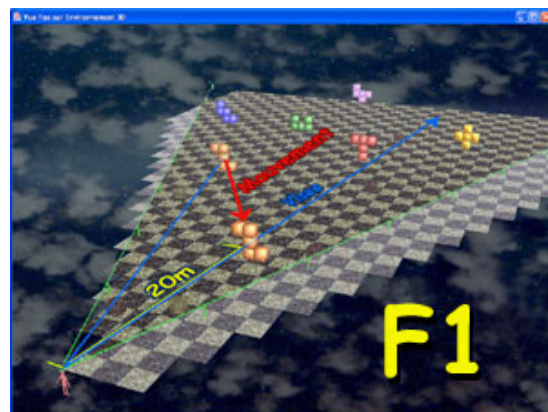


Figure 148. Représentation de la tâche « F1 ».

Quatre utilisateurs informaticiens ainsi que deux non informaticiens ont en effet exprimé une gêne par cette absence dans leurs commentaires post-expérimentaux,

soit 6 utilisateurs sur 20 ayant réussi à le verbaliser après avoir vu une démonstration post-expérimentale de l'application de "restitution" des enregistrements digitaux¹⁷. Ceci laisse présumer qu'un nombre plus important d'entre eux ont subi une telle gêne.

7.1.1. Durées des expressions de commande.

En ce qui concerne les durées d'expressions pour chaque tâche, on regroupe d'abord (Tableau 11), pour chacun des deux groupes d'utilisateurs, les durées des expressions relevées donnant lieu à succès pour en établir leur moyenne, ainsi que les plus courtes durées d'expression observées pour chaque tâche.

Tâches	Meilleurs Temps		Groupe Info.		Groupe Non-Info.	
	Groupe Info.	Groupe Non-Info.	Moyenne	ET	Moyenne	ET
F1	4940		12457	6070	10272	429
F2	5030		12925	5877	11180	4090
F3	3270		8526	4724	10808	3909
F4		4030	10387	4189	12864	11137
F5	3040		10010	5942	11942	4927
F6		8070	14710	6225	12783	4590
F7	5020		14622	13883	10602	4540
F9	5950		9187	2928	18234	8337

Tableau 11. Répartition des durées d'expression les plus courtes et des durées moyennes d'expression par groupe et par tâche, en millisecondes.

Les durées d'expressions respectives pour chacun des groupes ne présentant guère de différences en dehors des tâches F7 et F9, on peut simplifier le tableau précédent pour y rassembler les valeurs relatives à toutes les expressions ayant donné lieu à succès (Tableau 12).

Tâches	Meilleurs Tps	Moyenne	ET
F1	4940	11861	5183
F2	5030	12343	5150
F3	3270	9504	4389
F4	4030	11802	8669
F5	3040	10902	5364
F6	8070	14184	5670
F7	5020	13076	11110
F9	5950	11198	5570
Moyenne	4918	11859	
ET	1606	1415	

Tableau 12. Répartition des durées d'expression les plus courtes et des durées moyennes d'expressions pour l'ensemble des utilisateurs, exprimées en millisecondes.

¹⁷ L'application de restitution met en oeuvre un tel retour visuel destiné à l'expérimentateur (cf. Figure 141 page 201).

On constate alors une certaine uniformité dans les durées d'expression, voisinant la dizaine de secondes pour exprimer une commande de déplacement d'objet. Les durées d'expression les plus courtes, qui restent toutes inférieures à la dizaine de secondes, pouvant indiquer des durées auxquelles s'attendre après familiarisation avec le système ou avec ce style d'interaction.

7.1.2. Effectifs des apparition d'erreurs d'énoncé.

On appellera ici « erreur d'énoncé » toute formulation orale défectueuse. On distingue deux catégories. Les erreurs d'énoncé les plus évidentes constituant la première catégorie d'erreurs d'énoncé sont d'abord celles que nous regrouperons sous le terme « *d'erreurs d'omission* ». Omission du verbe d'opération (notée EIC), de la préposition précédant la quantité (EIP), absence de valeur de quantité (EIQ), ou encore direction de la destination incomplète ou absente (EID). Viennent ensuite dans une seconde catégorie ce que nous regrouperons sous le terme « *d'erreurs d'usage* » telles que la mise en œuvre d'une préposition inappropriée (UFPI), voire d'un verbe d'opération inapproprié ou désignant une autre opération non abordée dans cette étude (UFVI).

Variabilité des énoncés d'opération

Voyons, pour chaque tâche, la variabilité d'énoncé de l'opération. A toute fin utile, on trouvera les productions des utilisateurs dans le document annexe.

La tâche de positionnement égo-centré F1 (Tableau 13 ci-dessous), n'utilise globalement que trois types de termes similaires. 75% des énoncés « *déplacent* » un objet à une position définie. 10% de ces énoncés « *positionnent* » effectivement l'objet de la tâche (« *placer* », « *mettre* »), et 10% utilisent un terme générique de mouvement ne précisant aucune catégorie de mouvement (« *bouger* »).

Énoncé / F1	<i>Bouger</i>	<i>Place</i>	<i>Met</i>	<i>Déplace</i>	<i>Déplacer</i>
Effectif	2	1	1	6	9
Rapport (%)	10	5	5	30	45

Tableau 13. Répartition des termes d'opération utilisés pour la tâche F1. Note : l'un des énoncés ne contenait aucun groupe verbal désignant une opération.

Pour la tâche F2 de « déplacement » d'un objet d'une certaine distance dans une direction *exocentrée* (Tableau 14), 90% des énoncés mettent en œuvre un terme correspondant à la nature du déplacement, cependant que 5% d'entre eux utilisent un terme générique de mouvement ne précisant aucune catégorie (« *bouger* »).

Énoncé / F2	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Rapprocher</i>
Effectif	1	8	10	1
Rapport (%)	5	40	50	5

Tableau 14. Répartition des termes d'opération utilisés pour la tâche F2.

La tâche F3 de déplacement exocentré dans une direction relative (Tableau 15), voit 75% de ses énoncés mettre en œuvre un terme correspondant à la nature du déplacement, et 15% utiliser un terme générique de mouvement ne précisant aucune catégorie (« *bouger* »).

Énoncé / F3	<i>Bouge</i>	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Emmener</i>
Effectif	1	2	8	7	1
Rapport (%)	5	10	40	35	5

Tableau 15. Répartition des termes d'opération utilisés pour la tâche F3. Note : l'un des énoncés ne contenait aucun groupe verbal ou désignant une opération.

Pour la tâche approximative de repositionnement angulaire égocentrée F4 (Tableau 16), 75% des énoncés invoquent un déplacement. 5% utilisent un terme générique de mouvement, 10% mettent en œuvre un groupe verbal désignant plus particulièrement un mouvement circulaire.

Énoncé / F4	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>amener</i>	<i>Tourne</i>	<i>Faire tourner</i>
Effectif	1	8	7	1	1	1
Rapport (%)	5	40	35	5	5	5

Tableau 16. Répartition des termes d'opération utilisés pour la tâche F4. Note : L'un des énoncés produits ne contient pas de groupe verbal.

C'est maintenant pour la tâche de giration F5 (Tableau 17), traduisant un mouvement circulaire autour d'un axe distant que les énoncés confirment l'indéterminisme du seul groupe verbal non seulement pour le mouvement à exécuter, mais également quant à la catégorie de mouvement adressée.

Énoncé / F5	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Faire tourner</i>	<i>Faire une rotation</i>
Effectif	2	8	7	1	1
Rapport (%)	10	40	35	5	5

Tableau 17. Répartition des termes d'opération utilisés pour la tâche F5. Note : l'un des énoncés ne contenait aucun groupe verbal ou désignant une opération.

On constate que 10% des énoncés mettent en œuvre un terme réputé générique, que 10% seulement des énoncés contiennent un groupe verbal qui détermine effectivement un mouvement circulaire, mais que 75% de ceux-ci mettent en œuvre pour cette tâche un verbe que l'on aurait dédié aux seules translations.

Sur ces 15 énoncés utilisant le terme de translation (« *déplace* », « *déplacer* »), 13 d'entre eux lui ajoutent une direction linéaire (« vers la droite »), ce qui concerne alors 65% des énoncés utilisant le verbe générique de translation.

Il s'ensuit que pour une majorité d'expressions, le verbe utilisé n'effectue plus de distinction fiable entre les deux catégories prédéfinies de mouvements que forment

les translations et les mouvements circulaires. Il s'ensuit également que l'usage du verbe « *déplacer* », très largement mis en œuvre, n'exprime qu'un changement de position qui ne précise pas de catégorie de mouvement. Nous constatons également que c'est alors, dans ce cas, la nature de l'unité de mesure sélectionnée ultérieurement au sein de l'énoncé, qui complète la détermination de l'opération à effectuer.

F6 désigne la tâche de déplacement ciblé. Pour cette tâche (Tableau 18), 75% des énoncés « *déplacent* » l'objet de la tâche et 15% de plus utilisent un terme générique. On notera également l'usage des termes spécialisés « *avancer vers* » et « *rapprocher de* » qui déterminent mieux la tâche, en restant cependant d'un usage restreint.

Énoncé / F6	<i>Bouge</i>	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Avancer</i>	<i>Rapproche</i>
Effectif	1	2	7	8	1	1
Rapport (%)	5	10	35	40	5	5

Tableau 18. Répartition des termes d'opération utilisés pour la tâche F6.

La tâche F7 se rapporte à un positionnement relatif à un objet référent. Pour effectuer ce positionnement (Tableau 19), 85% des énoncés font appel à un « déplacement » contre seulement 5% qui positionnent effectivement avec un verbe correspondant à la nature de l'opération, cependant que 5% supplémentaires utilisent un terme générique.

Énoncé / F7	<i>Bouger</i>	<i>Met</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Avancer</i>
Effectif	1	1	7	10	1
Rapport (%)	5	5	35	50	5

Tableau 19. Répartition des termes d'opération utilisés pour la tâche F7.

Pour la tâche de rotation (F9), nous obtenons les productions suivantes :

Énoncé / F9	<i>Tourne</i>	<i>Tourner</i>	<i>Fais tourner</i>	<i>Pivote</i>	<i>Rotation</i> (*)	<i>Déplace</i>
Effectif	3	2	7	2	4	2
Rapport (%)	15	10	35	10	20	10

Tableau 20. Répartition des termes d'opération utilisés pour la tâche F9. (*) Les mises en œuvre de l'appellation du mouvement impliquant le terme « *rotation* » se déclinent sous les formes « *Effectuer une rotation* », « *faire une rotation* », « *doit faire une rotation* » et « *faire un mouvement de rotation* ».

70% des énoncés désignent donc le mouvement de rotation à l'aide d'un verbe dédié synonyme. 20% énoncent l'appellation du mouvement sous sa forme nominale. Les 10% restant utilisent un terme que l'on aurait dédié aux seules translations.

Ceci confirme en partie la carence déjà constatée du groupe verbal d'un énoncé à déterminer à lui seul l'opération à exécuter.

Regroupons maintenant ces résultats par catégorie de tâches sous forme de tableaux. Les « *positionnements* » concernent les tâches F1 et F7. Les « *déplacements* » sont constitués des tâches F2, F3 et F6. Les girations ne sont représentées que par la tâche F5 (Tableau 17), et les rotations par la tâche F9 (Tableau 20).

Énoncé	<i>Bouger</i>	<i>Place</i>	<i>Met</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Avancer</i>
Effectif	3	2	2	13	19	1
Rapport (%)	7,5	5	5	32,5	47,5	2,5
Cumuls (%)				80		

Tableau 21. Répartition des termes d'opération utilisés pour les tâches de « *positionnement* ».

Énoncé	<i>Bouge</i>	<i>Bouger</i>	<i>Déplace</i>	<i>Déplacer</i>	<i>Avancer</i>	<i>Rapproche</i>	<i>Emmener</i>
Effectif	2	5	23	25	1	2	1
Rapport (%)	3,33	8,33	38,33	41,67	1,67	3,33	1,67

Tableau 22. Répartition des termes d'opération utilisés pour les tâches de « *déplacement* ».

On constate, dans l'ensemble, une variabilité relativement faible au profit de quelques termes seulement qui rassemblent l'essentiel des expressions.

Un examen systématique des énoncés pour les tâches **F1** et **F7** (tâches de positionnement strict, cf. Tableau 19), nous permet de constater un usage de la langue pas tout à fait « conforme à la règle ». Ainsi, pour **F1** (cf. Figure 148), tâche de positionnement absolu centré utilisateur pour laquelle un terme de "**placement à**" serait plus exact eu égard à la langue française, **75%** des utilisateurs demandent au système de "**déplacer**" un objet "**à**" un endroit particulier. En ce qui concerne la tâche **F7**, similaire de ce point de vue, ce pourcentage s'élève à **85%**. Soit une moyenne de **80%** des 40 énoncés qui mettent en oeuvre un "**déplacement**" quelque peu abusif d'un objet "**à**" un endroit déterminé.

Ces rapports approximatifs contiennent également des énoncés erronés correspondant à d'autres tâches suite par exemple à une interprétation des présentations des tâches F1 et F7 autre que celles pour lesquelles elles furent respectivement conçues¹⁸. Cependant, les pourcentages respectifs rapportés aux seuls énoncés retenus comme pertinents pour les tâches considérées, s'élèvent encore à

¹⁸ 11 énoncés sur 40 pour les deux tâches **F1** et **F7**, produits par 9 utilisateurs différents, dont 2 informaticiens et 7 non-informaticiens. (Inf.MC-F1, "Vers moi". NInf.BJ-F1, "Devant moi". NInf.RC2-F1, "Vers la droite". NInf.AC-F1, "Vers cet objet". NInf.SI-F1, "à droite". Inf.CI-F7, "vers la droite". NInf.CV-F7, "dans cette direction". NInf.BC-F7, "de 20° à droite". NInf.GE-F7, "à la place de celui-ci". NInf.RC2-F7, "de 90° vers la droite". NInf.SI-F7, "de 20m, par ici").

63,63% et 84,61%, soit une moyenne approximative de 74,12% des 25 énoncés retenus.

Le verbe « *déplacer* », de loin le plus utilisé, sert à exprimer 80% des tâches de « *déplacement* » telles que nous les avons définies. Ce même verbe entre également dans 80% des énoncés pour les tâches de « *positionnement* », exprime 75% des « *girations* », et participe encore à 10% de la tâche de « *rotation* ».

On en retiendra que l'on ne peut se fier au verbe d'opération utilisé au cours de l'énoncé par les utilisateurs pour en déduire si l'opération à effectuer par le système sera un positionnement strict, un véritable déplacement, ou encore un mouvement de giration.

Tâche	EIC	EIP	EIQ	EID	UFPI	UFVI	Σ
F1	1	1	0	1	4	0	7
F2	0	2	0	3	2	1	8
F3	1	2	0	0	4	0	7
F4	1	0	0	0	0	2	3
F5	1	2	1	2	0	2	8
F6	0	1	0	1	1	1	4
F7	0	1	1	3	6	1	12
F9	0	1	1	9	1	0	12
Σ	4	10	3	19	18	7	61
%	6,55	16,39	4,91	31,14	29,50	11,48	100

Tableau 23. Effectifs des erreurs d'omission au sein de l'énoncé et des erreurs d'usage de la préposition ou du verbe adéquat sur l'ensemble des énoncés pour chaque tâche proposée. EIC : omission du verbe d'opération. EIP : omission de la préposition précédant la quantité. EIQ : absence de valeur de quantité. EID : direction de destination incomplète ou absente. UFPI : usage de préposition inadéquate. UFVI : verbe d'opération inapproprié.

Les erreurs d'énoncé sont réunies pour l'ensemble des 160 énoncés dans le Tableau 23 ci-dessus. Les pourcentages figurés en dernière ligne sont rapportés au nombre total d'erreurs. Les erreurs les plus fréquentes pour l'ensemble des utilisateurs sont donc l'emploi inadéquat de la préposition et les directions de destination incomplètes ou absentes.

Reportons maintenant les mêmes erreurs d'énoncé par groupe d'utilisateur.

Les erreurs d'énoncé du Tableau 24 ci-après concernent le groupe des informaticiens, et affectent 21 des 80 énoncés produits par ce groupe. Le type d'erreur le plus fréquemment produit concerne les « erreurs d'usage » cumulant 70% des erreurs d'énoncé (UFPI+ UFVI).

Tâche	EIC	EIP	EIQ	EID	UFPI	UFVI	Σ
F1	0	0	0	0	1	0	1
F2	0	0	0	0	2	1	3
F3	0	0	0	0	2	0	2
F4	0	0	0	0	0	2	2
F5	0	1	1	0	0	1	3
F6	0	0	0	0	1	1	2
F7	0	0	0	1	3	1	5
F9	0	0	0	3	0	0	3
Σ	0	1	1	4	9	6	21
%	0	4,76	4,76	19,04	42,85	28,57	100

Tableau 24. Effectifs des erreurs d'omission au sein de l'énoncé et des erreurs d'emploi de la préposition ou du verbe adéquat constatés dans les énoncés produits par le groupe des informaticiens.

Voyons maintenant les erreurs d'énoncé produites par le groupe des « non-informaticiens ».

Tâche	EIC	EIP	EIQ	EID	UFPI	UFVI	Σ
F1	1	1	0	1	3	0	6
F2	0	2	0	3	0	0	5
F3	1	2	0	0	2	0	5
F4	1	0	0	0	0	0	1
F5	1	1	0	2	0	1	5
F6	0	1	0	1	0	0	2
F7	0	1	1	2	3	0	7
F9	0	1	1	6	1	0	9
Σ	4	9	2	15	9	1	40
%	10	22,5	5	37,5	22,5	2,5	100

Tableau 25. Effectifs des erreurs d'omission au sein de l'énoncé et des erreurs d'emploi de la préposition ou du verbe adéquat constatés dans les énoncés produits par le groupe des non-informaticiens.

On remarque dans le Tableau 25 que pour le groupe des « non-informaticiens », les trois quarts des erreurs d'énoncé sont cumulée par les omissions de termes (EIC+EIP+EIQ+EID), contre un quart seulement d'usages inadéquats de la préposition ou des verbes d'opération.

7.1.3. Usage des modalités

Huit tâches ont été retenues pour tester la pertinence de l'association de trois groupes de modalités d'entrée. Quatre d'entre elles impliquent la mise en œuvre d'énonciations, de visée à l'aide de la tête, et d'un geste de la main, les quatre autres ne nécessitent que des énonciations et la direction de la tête. Après lecture de la

« consigne aux participants » (en annexe) leur indiquant en particulier la mise en oeuvre des opérations de bas niveau (présélections, sélections) ainsi que celle d'un énoncé complet, les utilisateurs conservent le choix d'exprimer librement la production d'une commande. Le choix des modalités est donc restreint au niveau le plus bas. Les redondances sont cependant admises et l'ordre de sélection des composants de la commande n'est pas imposé.

Expressions monomodales

D'un point de vue monomodal, la modalité vocale d'énonciation reste la seule sur laquelle le système mis en place permet aisément des comparaisons d'enregistrement. Les directions indiquées par visée de la tête ou indiquées avec le gant ne permettent de comparaisons que lorsqu'un énoncé de sélection est employé simultanément. Nous allons rappeler ici le rôle de chaque modalité seule avant d'aborder davantage l'usage des couples de modalités. Rappelons également que l'usage simultané des trois modalités de visée de tête, d'indication de direction manuelle et d'un énoncé vocal n'apparaît pas ici, car il trouve peu d'exemple de mise en oeuvre (exemple, la sélection d'un point d'intersection entre visée et direction manuelle dans l'espace rapproché autour de l'utilisateur, non étudié ici).

L'affectation des modalités participant à des sous-tâches est la suivante :

Direction de visée avec la tête

La direction de visée avec la tête ne permet à elle seule que deux opérations.

La première opération est la présélection d'un objet de la scène en vue de sa sélection, ou afin d'en connaître la distance. La seconde opération possible est la présélection d'une direction centrée utilisateur si elle reste circonscrite au champ visuel offert.

Indication manuelle de direction

L'indication manuelle de direction, utilisée seule, ne reçoit que deux affectations. La première est la présélection d'une direction exocentrée dont l'origine est localisée à un objet sélectionné de la scène. La seconde est l'indication d'un axe et sens de rotation par mouvement de description d'un tracé circulaire.

Modalité vocale seule

L'expression d'une commande, d'un point de vue monomodal, commence en principe avec la sélection vocale de l'opération à effectuer (on verra plus loin que ce n'est toutefois pas toujours le cas). Il arrive toutefois que le verbe d'opération soit omis de l'énoncé par certains utilisateurs. C'est le cas pour quatre énoncés produits (par deux utilisateurs) sur l'ensemble des 160 expressions produites. Nous avons répertorié cette omission pour chaque groupe d'utilisateurs et pour chaque tâche dans le Tableau 26 suivant.

Tâches (EIC)	Groupe Informaticiens	Groupe Non- Informaticiens
F1 (%)	0	14,28
F2 (%)	0	0
F3 (%)	0	20
F4 (%)	0	50
F5 (%)	0	25
F6 (%)	0	0
F7 (%)	0	0
F9 (%)	0	0
Somme	0	109,28
Moyenne (%)	0	13,66
ET	0	17,87
Student (p)	0,067428289	

Tableau 26. Effectifs des apparitions d'erreurs d'omission du verbe d'opération rapportées au total des « insuccès » (ou expressions n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Ces valeurs représentent le pourcentage des occurrences de l'erreur constatée, rapporté au nombre d'expressions infructueuses produites par le groupe pour la tâche considérée. Elles figurent ainsi les proportions de participation de l'erreur par rapport à l'effectif des expressions erronées. Les quelques 14,3% concernant la tâche F1 pour le groupe des non informaticiens représentent ainsi une omission du verbe d'opération sur sept expressions infructueuses produites par le groupe pour cette tâche. La répartition de cette omission du verbe dans l'énoncé n'est donc pas significative, mais reste néanmoins, toutes tâches confondues, l'exclusivité du groupe des « Non Informaticiens ».

Sélection de quantité (valeur, unité)

Nous avons déjà évoqué précédemment le cas de l'absence de préposition introduisant la valeur ou quantité que nous allons maintenant examiner. L'absence de quantité n'affecte que 3 énoncés sur l'ensemble des 140 expressions produites impliquant la spécification d'une quantité.

Ce dernier Tableau 27 ne représente pas la tâche F4, pour laquelle aucune quantité n'étant nécessaire, son omission ne saurait constituer une erreur.

On y constate que bien que les omissions des quantités n'apparaissent pas pour chaque groupe pour les mêmes tâches (ce qui pourrait constituer une différence supplémentaire distinguant les deux groupes), leurs moyennes respectives toutes tâches confondues sont équivalentes.

Tâches (EIQ)	Groupe Informaticiens	Groupe Non- Informaticiens
F1 (%)	0	0
F2 (%)	0	0
F3 (%)	0	0
F5 (%)	33,33	0
F6 (%)	0	0
F7 (%)	0	20
F9 (%)	0	12,5
Somme	33,33	32,5
Moyenne (%)	4,76	4,64
ET	12,59	8,21
Student (p)	0,983694189	

Tableau 27. Effectifs des apparitions d'erreurs d'omission de valeur de quantité rapportées au total des « insuccès » (ou expression n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Sélection de direction relative orthogonale

La sélection d'une direction relative par énonciation concerne trois tâches avec un ensemble de 60 énoncés.

Tâches	EID	
	Groupe Informaticiens	Groupe Non- Informaticiens
F3	0	0
F5	0	2
F7	1	2
Σ	1	4

Tableau 28. Effectifs des erreurs d'incomplétude ou d'absence d'indication de direction de destination pour les tâches F3, F5 et F7 impliquant l'indication d'une direction relative orthogonale pour chacun des deux groupes d'utilisateurs.

Le Tableau 28 ci-dessus rapporte les erreurs d'incomplétude ou d'absence d'indication de direction de destination pour les tâches F3, F5 et F7 impliquant une direction relative orthogonale (« *vers la droite* », « *à droite de* », etc.). On constate que 5 erreurs ayant été commises, l'indication d'une direction relative de destination fut correcte pour 55 des énoncés (91,6%).

Direction de la tête et vocal

Sélection d'un objet de la scène pour l'opération

En ce qui concerne la sélection de l'objet de la commande, celui-ci n'a jamais manqué d'être désigné. Tout au plus un utilisateur a-t-il, pour la tâche F6, sélectionné un autre objet que celui pour lequel la tâche était proposée. Nous ne constatons donc

aucune erreur quant à la sélection d'un objet de la scène. Quelques utilisateurs ont cependant manifesté une certaine difficulté, observable lors de l'expérimentation, à contrôler précisément la position du viseur.

Sélection d'une direction centrée utilisateur

Cette forme d'indication de direction n'apparaît qu'en restant circonscrite au champ visuel offert par l'affichage de la scène. F1, F4.

Tâches	EID	
	Groupe Informaticiens	Groupe Non- Informaticiens
F1	0	1
F4	0	0
Σ	0	1

Tableau 29. Effectifs des erreurs d'incomplétude ou d'absence d'indication de direction de destination pour les tâches F1 et F4 impliquant l'indication d'une direction centrée utilisateur pour chacun des deux groupes.

Deux tâches impliquent l'indication d'une direction de destination prenant son origine à la position occupée par l'utilisateur et pointant sur la scène. Nous allons, pour ces 40 énoncés, détailler davantage ce qui a été constaté. Il s'agit d'abord de la tâche F1 qui consiste à disposer un objet à une certaine distance de l'utilisateur, dans une direction située sur la scène à partir de l'utilisateur. Deux utilisateurs du groupe des informaticiens produisirent des expressions qui correspondaient à une autre tâche. Parmi les huit autres expressions, sept contenaient une indication de direction de destination par visée avec la tête, la huitième exprimait une direction indiquée avec la main. Cinq utilisateurs du groupe des non informaticiens produisirent des expressions qui correspondaient à d'autres tâches. Parmi les cinq autres expressions, une fut infructueuse pour défaut d'indication de direction (erreur figurant à ce titre dans le Tableau 29 ci-dessus), et quatre contenaient une indication de direction de destination par visée avec la tête. Soit, pour 13 tâches réputées « comprises » par les utilisateurs, 11 indications de direction adéquates avec la tête (84,6%).

Examinons maintenant la tâche F4, qui consiste à désigner une direction de repositionnement d'un objet dans une autre direction pointant sur la scène que celle dans laquelle cet objet se trouve par rapport à l'utilisateur (sans modification de sa distance à l'utilisateur). Sur les 10 expressions produites par le groupe des informaticiens, quatre correspondaient à une autre tâche. Cinq contenaient une indication de direction de destination visée avec la tête, et la dernière contenait une indication manuelle de direction de destination. Deux des 10 expressions du groupe des non informaticiens concernaient une autre tâche, les huit autres expressions contenaient une indication de direction de destination visée avec la tête. Soit, pour 16 tâches réputées impliquer une direction égocentrée, 11 indications de direction adéquates avec la tête (81,25%).

Nous obtenons alors une moyenne de 82,9% de mise en œuvre de la direction de la tête pour des tâches impliquant une indication de direction à partir de la position virtuellement occupée par l'utilisateur, ce qui milite en faveur d'une telle association de deux modalités pour la sélection de direction égocentrée.

Indication manuelle de direction et vocal

Sélection d'une direction exocentrée

L'origine d'une telle direction est localisée à un objet de la scène (par exemple pour une translation latérale dont la direction ne peut être indiquée avec la visée).

La tâche F2 implique l'indication, à l'aide d'un geste déictique manuel, d'une direction dont l'origine est localisée à la position de l'objet de la commande. Elle consiste à déplacer un objet, depuis la position qu'il occupe initialement, d'une certaine distance dans une direction à indiquer.

Quatre utilisateurs ne mettent pas de geste déictique manuel en œuvre pour indiquer une direction de translation, et manifestent une direction relative vocale (« à gauche »). Un autre indique avec la tête la direction de la destination.

Un sixième cherche à « rapprocher » l'objet de la commande d'un second objet cible situé approximativement dans la direction voulue, d'une distance de 20m. La tâche paraît donc incomprise, ou bien la mise en œuvre du gant pour cet usage est problématique.

L'un des utilisateurs présente un problème d'interprétation avec un geste déictique approximatif qui semble pointer la destination plutôt que la direction de translation.

Davantage au titre de la variabilité d'expression, deux utilisateurs indiquent une direction manuelle correcte, l'un en manifestant oralement une direction relative (« à gauche »), l'autre en la validant du terme « ici », plutôt que « dans cette direction », ce qui prête à confusion car le terme vocal employé désigne un locus. Une troisième expression présente un cas intéressant, l'utilisateur déplace en effet l'objet de 20m « sur la gauche dans cette direction » avec un geste déictique manuel indiquant clairement une direction de translation.

(Note : l'un des utilisateurs indique correctement la direction voulue, mais sans activer la zone tactile du gant, cet utilisateur effectue les gestes déictiques sans refermer le pouce sur le majeur).

Six des 20 utilisateurs n'ont pas compris la représentation de la tâche à accomplir.

L'une des expressions pose un problème d'interprétation.

Trois énoncés apportent une manifestation de direction relative qui pourrait être attribuée à une variabilité d'expression, mais dont l'usage impliquerait un conflit d'affectation des modalités.

Pour résumer, sur 14 utilisateurs qui ont manifestement compris la manœuvre à effectuer, 10 ont employé un geste déictique indiquant une direction de translation correspondant à la tâche (71,4%), 3 ont augmenté leur expression d'une direction relative ou d'une référence verbale à une direction de destination, et un seul semble avoir mis en œuvre un geste déictique manuel pointant la destination.

Sélection d'un axe et sens de rotation

Un axe et un sens de mouvement circulaire peuvent être simultanément indiqués à l'aide d'un geste décrivant un tracé circulaire.

La tâche de rotation d'un objet présente une variabilité importante dans les résultats. Du point de vue de l'indication d'un axe et sens de rotation de l'objet, neuf des vingt expressions produites par l'ensemble des utilisateurs laissent apparaître un problème d'indication de direction (45%).

Une telle indication fut définie par la description d'un geste mimétique de rotation avec un énoncé vocal lui faisant référence sur la base d'exemples d'expression courante. D'autres expressions sont sans doute possibles pour peu qu'elles définissent complètement le mouvement sans ambiguïté.

Sur ces neuf expressions n'indiquant pas le mouvement à effectuer de façon bimodale claire ou complète, six d'entre elles manifestent un geste mimétique manuel correspondant à la tâche, c'est alors la référence vocale à ce mouvement qui fait généralement défaut, traduisant éventuellement une tentative de manipulation directe.

Le mouvement mimétique indiquant une rotation est donc effectué au cours de 17 des 20 expressions enregistrées pour cette tâche (85%).

Autres usages constatés

Une redondance entre la direction de la tête et un geste déictique manuel supplémentaire peut apparaître au cours d'une expression, celle-ci peut concerner la désignation d'un objet de la scène ou l'indication d'une direction depuis la position occupée par l'utilisateur.

Sur les 160 expressions produites impliquant la sélection d'un objet, l'indication manuelle de la direction d'un objet de la commande à sélectionner avec la direction de la tête est apparue 19 fois (11,9%). Cette redondance paraît donc relativement faible, d'autant plus que 14 de ces désignations trimodales sont le fait de deux utilisateurs seulement, lesquels représentent donc 10% de l'ensemble des utilisateurs. Les cinq autres désignations redondantes sont distribuées sur quatre tâches différentes.

Lorsqu'il s'agit de sélectionner un second objet de la scène au cours de l'expression (cas de 40 expressions pour l'ensemble des utilisateurs), cette redondance apparaît 10 fois, soit cette fois dans 25% des expressions concernées. Quatre de ces

redondances sont le fait des deux mêmes utilisateurs que précédemment, les six autres étant distribuées sur quatre utilisateurs supplémentaires.

L'indication redondante d'un geste déictique manuel de direction depuis l'utilisateur ne peut être considérée que pour deux tâches pour lesquelles elle est facultative (F1 et F4). Il s'agit des tâches impliquant une direction depuis l'utilisateur vers la scène, qu'une direction de la tête suffit à exprimer. Dans ce cas, le phénomène est cette fois plus important avec 20 des 40 expressions qui dénotent une redondance d'indication de direction centrée utilisateur.

La redondance d'un mouvement manuel au cours de la désignation d'objet ne présente pas de conflit sachant que cette tâche est clairement dévolue avec succès à la direction de la tête. En ce qui concerne l'indication d'une direction quelconque sur la scène depuis l'utilisateur, cette manifestation pose davantage de problèmes du fait que les deux modalités gestuelles mises en œuvre participent toutes deux à des rôles cependant différents d'indications de direction.

7.1.4. Comparaisons individuelles d'erreurs ou de particularités.

Voyons maintenant s'il nous est possible de trouver une explication concernant la différence précédemment constatée en terme de succès des expressions produites par les deux groupes d'utilisateurs étudiés.

Malgré la faiblesse de l'échantillon, nous avons constaté une différence significative entre les deux groupes d'utilisateurs. Nous pouvons supposer que l'expérience des informaticiens avec différents systèmes d'interaction (ex. WIMP vs Ligne de commande) leur apporte une éventuelle capacité d'adaptation à un nouveau système d'interaction. Nous pouvons également supposer que le groupe de non informaticiens, qui est essentiellement formé d'utilisateurs dont 80% ne disposent d'une expérience qu'avec un unique modèle d'interaction de type WIMP, cherchent éventuellement à reproduire ce modèle en présence d'un nouveau système d'interaction.

Si les deux groupes commettent les mêmes types d'erreur dans des proportions similaires, on pourra incriminer une cause liée à une erreur de conception du système. Dans le cas contraire, l'observation d'une différence notable entre les erreurs commises par chaque groupe conduira à établir des regroupements d'erreurs caractérisant chacun des deux groupes d'utilisateurs.

D'un point de vue « comportement » expressif des utilisateurs, on peut tout d'abord établir, pour chaque tâche et par groupe, une comparaison du nombre d'inversions de l'opération avec l'objet de l'opération au cours de l'énoncé (ex. « *Cet objet, déplace...* »), qui fut constatée chez de nombreux utilisateurs. (Note : cette caractéristique de certains énoncés n'a pas, dans le cadre de l'expérimentation, constitué une « erreur » donnant lieu à insuccès, toujours en vertu du principe selon lequel si une fraction importante de la population s'exprime de la sorte, alors il faudra l'admettre comme l'une des différentes façons de s'exprimer, pour peu qu'elle ne soit

pas ambiguë, afin d'obtenir l'exécution de la commande correspondante). Cette caractéristique d'inversion affecte en tout 29 énoncés (voir Tableau 30).

Tâches (VSOO)	Groupe Informaticiens		Groupe Non-Informaticiens	
	Σ	Moyenne	Σ	Moyenne
F1	0	0	3	0,3
F2	0	0	4	0,4
F3	0	0	3	0,3
F4	0	0	2	0,2
F5	0	0	3	0,3
F6	0	0	5	0,5
F7	0	0	4	0,4
F9	0	0	5	0,5
Somme	0		29	
Moyenne	0		3,625	
ET	0		1,0606	
Student (p)	2,67401E-05			

Tableau 30. Effectifs et moyennes des inversions du verbe et de l'objet de l'opération au sein des énoncés pour chacun des deux groupes et pour chaque tâche.

On notera qu'aucun des utilisateurs du groupe des informaticiens ne s'est exprimé en effectuant une telle inversion du terme exprimant l'opération (le verbe) avec l'objet de l'opération. Cette caractéristique « comportementale » reste donc le fait d'utilisateurs du groupe des « Non informaticiens », et affecte 29 énoncés sur 80, soit plus d'un tiers des expressions produites par ce groupe d'utilisateurs constituant une différence significative entre les deux groupes.

Autre caractéristique similaire à celle que nous venons de mentionner, mais constituant cette fois une « erreur » conduisant à l'insuccès, celle qui consiste à produire un énoncé ne comportant pas de préposition précédant la valeur ou quantité au sein de l'énoncé. Ex. typique, « *Déplacer cet objet vers la droite, 20m.* » qui ne permet pas de déterminer s'il s'agit d'un positionnement "à" 20m de l'utilisateur, ou d'un déplacement "de" 20m à partir de l'endroit où se trouve l'objet. Cette forme d'expression quelque peu « encodée », donnant lieu à des énoncés ambigus, fut également constatées chez plusieurs utilisateurs.

Le Tableau 31 rapporte les pourcentages d'erreurs constitués par l'absence de préposition au sein de l'énoncé, pour chaque tâche, par chacun des deux groupes. Dans ce dernier tableau ne figure pas la tâche F4 pour laquelle aucune quantité n'est en principe nécessaire (« en principe », car un utilisateur informaticien scrupuleux vis-à-vis de la syntaxe conseillée, dérouté par l'absence de valeur sur la représentation de la tâche, en a tout de même ajouté une à son énoncé, avec succès).

Tâches (EIP)	Groupe Informaticiens	Groupe Non- Informaticiens
F1 (%)	0	14,28
F2 (%)	0	28,57
F3 (%)	0	40
F5 (%)	33,33	25
F6 (%)	0	14,28
F7 (%)	0	20
F9 (%)	0	12,5
Somme	33,33	154,64
Moyenne (%)	4,76	22,09
ET	12,59	9,91
Student (p)	0,015	

Tableau 31. Effectifs des apparitions d'erreurs d'omission de la préposition rapportées au total des « insuccès » (ou expressions n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Une expression pouvant révéler plusieurs erreurs conduisant à l'échec, chacun de ces pourcentages ne constitue pas forcément à lui seul la cause de l'échec de l'expression, ils représentent en fait la fréquence d'apparition de l'absence de préposition introduisant la valeur ou quantité du mouvement, rapportés au nombre d'expressions ayant conduit à l'échec.

On constate que l'absence de préposition introduisant la quantité dans l'énoncé n'est plus l'exclusivité du groupe des « Non Informaticiens », mais que ce dernier groupe se distingue tout de même par cette caractéristique « comportementale ».

D'autres absences peuvent ensuite caractériser les énoncés des utilisateurs, sans forcément caractériser l'un ou l'autre groupe.

Ainsi, dans l'ordre d'énonciation, il arrive que le verbe d'opération soit omis.

La répartition de cette omission du verbe dans l'énoncé sur le Tableau 26 p.214 n'est pas complètement significative et son effectif sur le nombre total d'énoncés reste minimale (2,5%, soit 5% des énoncés du groupe des non informaticiens). Cette erreur reste néanmoins l'exclusivité du groupe des « Non Informaticiens ».

Nous avons déjà évoqué précédemment le cas de l'absence de préposition introduisant la valeur ou quantité que nous allons maintenant examiner.

Dans le Tableau 27 p.215 ne figure pas la tâche F4, pour laquelle aucune quantité n'étant nécessaire, son omission ne saurait constituer une erreur.

On y constate que bien que les omissions des quantités n'apparaissent pas pour chaque groupe pour les mêmes tâches (ce qui pourrait constituer une différence supplémentaire distinguant les deux groupes), leurs moyennes respectives toutes tâches confondues sont pratiquement égales.

Pour aborder la dernière omission ou incomplétude possible des commandes, examinons maintenant le cas des indications de direction de destination. Il arrive en

effet que l'indication d'une direction soit absente, incomplète, ou inappropriée à la tâche proposée, conduisant à l'échec de l'expression (Tableau 32).

Tâches (EID)	Groupe Informaticiens	Groupe Non- Informaticiens
F1 (%)	0	14,2857
F2 (%)	0	42,8571
F3 (%)	0	0
F4 (%)	0	0
F5 (%)	0	50
F6 (%)	0	14,2857
F7 (%)	50	40
F9 (%)	100	75
Somme	150	236,4285
Moyenne (%)	18,75	29,5535
ET	37,2011	26,6795
Student (p)	0,5164	

Tableau 32. Effectifs des apparitions d'erreurs d'incomplétude ou d'absence d'indication de direction de destination rapportées au total des « insuccès » (ou expressions n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Sauf à considérer les tâches individuellement, on n'y remarque guère de différences globales. Chacun des groupes peut ainsi commettre, dans ses erreurs conduisant à l'échec de l'expression produite, une omission au sein de l'expression d'une direction de translation ou de positionnement.

Pour terminer les erreurs d'omission ou d'incomplétude des commandes, voyons ce que donne le regroupement de ces « *erreurs d'omission* » en une seule catégorie dans le Tableau 33 ci-dessous.

Un tel regroupement de toutes les erreurs d'omission possibles présente une moyenne de ces erreurs traduisant des énoncés incomplets (absence d'opération, de préposition, de valeur...). On observe sur une telle présentation une différence qui devient significative en caractérisant le groupe des « Non Informaticiens ».

Voyons maintenant un autre type d'erreur s'exprimant lui aussi dans les énoncés de plusieurs utilisateurs. Il s'agit, bien que la tâche semble effectivement avoir été comprise, d'une part de l'usage incorrect d'une préposition introduisant une valeur de quantité, et d'autre part de l'usage incorrect d'un verbe pour la tâche proposée.

Tâches (Moy % EI)	Groupe Informaticiens	Groupe Non- Informaticiens
F1	0	10,7142
F2	0	17,8571
F3	0	15
F4	0	12,5
F5	16,6666	25
F6	0	7,1428
F7	12,5	20
F9	25	25
Somme	54,1666	133,2142
Moyenne %	6,7708	16,6517
ET	9,9446	6,5212
Student (p)	0,036577357	

Tableau 33. Effectifs des apparitions des « erreurs d'omission » rapportées au total des « insuccès » (ou expressions n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

L'usage correct de la préposition relève bien sûr d'un certain respect de la langue, cependant, sachant que les trois quarts des utilisateurs n'apportent pas de distinction entre un positionnement et un déplacement par l'usage précis du verbe d'opération correspondant dans leurs énoncés, l'usage correct des prépositions relève aussi et surtout du fait que la préposition constitue alors le seul élément de l'énoncé qui distingue le type de mouvement auquel l'utilisateur fait référence (positionnement ou déplacement). Le Tableau 23 p.211 totalise 18 usages inadéquats de la préposition (« UFPI »), soit environ 13% du nombre total des 140 énoncés concernés (on ne tient pas compte ici de la tâche F4).

En ce qui concerne l'usage incorrect de la préposition, nous allons citer un exemple pour plusieurs tâches afin de clarifier ce type d'erreur.

Pour la tâche F1, qui consiste à positionner un objet à une certaine distance dans une direction centrée sur l'utilisateur qui en occupe le point d'origine, l'exemple type d'énoncé erroné est le suivant : « *Déplacer cet objet, dans cette direction, **de** 20m.* », cependant que l'utilisateur indiquera clairement la direction proposée à l'aide d'une voire plusieurs modalités gestuelles d'indication de direction. Le plus étonnant à ce sujet reste que certains utilisateurs ont produit deux fois ce même énoncé pour deux tâches différentes (F1 et F2).

Pour la tâche F2, qui consiste à déplacer un objet **d'**une certaine distance dans une certaine direction à partir de la position que cet objet occupe initialement, ce sera l'inverse. Un énoncé typique de cette erreur sera : « *Déplace cet objet **à** 20m dans cette direction* » avec une direction manuelle exocentrée clairement indiquée.

La tâche F3 rencontre la même préposition incorrecte pour déplacer un objet d'une certaine quantité vers la droite, ex. « *Déplace cet objet **à** 50m, vers la droite.* ».

Expression qui pourrait indiquer un positionnement de l'objet à droite de l'utilisateur à la distance indiquée.

La tâche F5 est similaire à F3, et par extension, nous pourrions nous attendre à un énoncé erroné du type « *Déplace cet objet à 30°, vers la droite.* », qui n'est cependant pas apparu.

Pour la tâche F6, qui consiste à rapprocher un objet d'un autre d'une certaine quantité, cette erreur d'usage de la préposition est d'autant plus problématique qu'il donne une autre signification à l'expression. Ainsi, l'énoncé « *Déplace cet objet à 20m dans la direction de cet objet* » n'évoque-t-il pas un déplacement d'une certaine distance dans une certaine direction telle que proposée pour la tâche F6, mais plutôt un positionnement à une certaine distance d'une cible désignée.

Tâches (UFPI %)	Groupe Informaticiens	Groupe Non- Informaticiens
F1	50	42,8571
F2	50	0
F3	100	40
F5	0	0
F6	50	0
F7	150	60
F9	0	12,5
Somme	400	155,3571
Moyenne (%)	57,1428	22,1938
ET	53,4522	24,9833
Student (p)	0,153488819	

Tableau 34. Effectifs des apparitions de l'usage inadéquat d'une préposition rapportées au total des « insuccès » (ou expression n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Bien que la différence entre les deux groupes au sujet de l'usage de la préposition ne soit pas significative (Tableau 34), elle est néanmoins plus fréquemment observée au sein du groupe des informaticiens avec une moyenne deux fois et demie plus importante.

Examinons maintenant l'éventuel usage incorrect du verbe d'opération.

Comme déjà signalé, la représentation des tâches F4 et F5 ayant induit plusieurs utilisateurs en erreur en les incitant à invoquer un mouvement orbital d'un objet distant autour de leur position à l'aide du terme « *Tourner* », il n'en sera pas tenu compte ici, car cet usage incrimine clairement un défaut de conception de l'expérience.

Un terme inapproprié pour la tâche F2 consiste par exemple à « *Rapprocher* » un objet là où un déplacement serait plus indiqué (le « *Rapprochement* » trouvant plutôt un éventuel usage pour la tâche F6 qui implique de « *déplacer* » un objet d'une certaine distance en direction d'un second objet cible, ce qui revient au même).

Tâches (UFVI %)	Groupe Informaticiens	Groupe Non- Informaticiens
F1	0	0
F2	25	0
F3	0	0
F6	50	0
F7	50	0
F9	0	0
Somme	125	0
Moyenne (%)	20,83	0
ET	24,57	0
Student (p)	0,0925	

Tableau 35. Effectifs des apparitions d'erreurs d'usage d'un verbe d'opération inapproprié rapportées au total des « insuccès » (ou expressions n'ayant pas conduit au succès) pour chacun des deux groupes d'utilisateurs et pour chaque tâche.

Pour F6, on retiendra par exemple l'énoncé « *Avancer cet objet vers cet objet d'une distance de 20m.* », qui, bien que relativement correct, suppose un objet de la scène muni d'une face « avant » qui n'a pas été envisagée pour ces déplacements ou positionnements.

La répartition de ce type d'erreur (Tableau 35 ci-dessus) sur l'emploi du verbe d'opération apparaît clairement constituer l'exclusivité du groupe des informaticiens.

7.1.5. En résumé

Pour résumer ces informations, nous observons deux populations significativement différentes quant à la production d'énoncés conduisant au succès.

Du point de vue des erreurs commises, le groupe des « Informaticiens » prédomine par un usage inapproprié de la préposition introduisant les quantités ou valeurs et reste le seul à user parfois de verbes d'opération inappropriés en regard de la tâche.

Le groupe des « Non Informaticiens » produit des erreurs qui se distinguent de façon significative par la production d'énoncés incomplets apparaissant dans plus d'un tiers de ses énoncés, en particulier en ce qui concerne les prépositions introduisant les quantités ou valeurs pour 9 des 80 énoncés, soit 11,25% des expressions (*cf.* Tableau 25 p.212, colonne « UIP »). De telles erreurs pourraient indiquer une surcharge mentale chez l'utilisateur, bien que le nombre d'empans à mettre en œuvre soit généralement inférieur à 7 (Miller, 1956), que la préposition ne se situe pas en fin d'énoncé, et qu'en outre, une éventuelle surcharge devrait également affecter le groupe des informaticiens. Ces erreurs pourraient tout autant indiquer une gêne liée à des effets de contexte lorsque l'utilisateur non informaticien est confronté à un nouvel environnement technique face auquel il croit par exemple devoir s'adapter.

Par ailleurs, et sans qu'il s'agisse d'une « erreur », mais plutôt d'un usage inhabituel du langage articulé, ce même groupe se distingue par des inversions, au sein de l'énoncé, du verbe d'opération avec l'objet de l'opération.

Les deux groupes semblent effectivement produire des erreurs différentes, indiquant *a priori* des erreurs davantage liées à des caractéristiques de chacun des deux groupes qu'à un défaut de conception de l'interaction (qui fournirait alors des erreurs similaires).

7.1.6. Hypothèse

Le groupe des informaticiens produisant nettement moins d'erreurs que le groupe des « Non Informaticiens », et les erreurs respectives des deux groupes paraissant assez différentes, on peut suspecter l'existence d'un phénomène conduisant le groupe des « Non Informaticiens » à en produire davantage.

A défaut de meilleure explication qui nécessiterait d'autres investigations, on peut s'interroger sur le résultat que fournit un énoncé cumulant plusieurs des erreurs ou particularités propres aux « Non Informaticiens ».

L'inversion de l'objet de la commande avec le verbe d'opération, ainsi que l'absence de préposition introduisant une quantité, tels qu'au sein de l'énoncé :

« *Cet objet, déplacer, à droite, 20m* » rappelle en effet, avec une segmentation caractéristique des éléments qui le composent, le style d'interaction mis en œuvre avec des interfaces WIMP (« Windows-Icon-Menu-Pointing »), largement expérimenté par la plupart des utilisateurs d'informatique contemporaine « Non Informaticiens », (ce qui est le cas de l'ensemble des utilisateurs du groupe des « Non Informaticiens » étudié).

Cette structure de l'énoncé rappelle la sélection préalable de la représentation visuelle d'un objet (l'icône), suivi de la sélection de l'opération à lui appliquer, avant réglage des divers paramètres au sein d'une nouvelle fenêtre généralement formatée en zones syntaxiques préétablies n'offrant aucune place pour quelque préposition.

Il est donc possible d'envisager que les utilisateurs du groupe des « Non Informaticiens », abordant un nouveau style d'interaction pour la première fois et sans familiarisation ou formation préalable, cherchent à « dupliquer » un comportement précédemment acquis au cours de l'apprentissages des interfaces GUI¹⁹ avec interaction WIMP²⁰ auxquels ils sont accoutumés.

Une telle explication aurait l'avantage de permettre plus rapidement aux utilisateurs « non informaticiens » de produire autant d'expressions correctes que les informaticiens lorsqu'ils abordent ce style d'interaction, en leur laissant par exemple

¹⁹ « GUI » : Graphical User Interface, ou Interface Graphique Utilisateur.

²⁰ « WIMP » : Windows-Icon-Menu-Pointing, ou pointage sur fenêtres présentant icônes et menus.

la possibilité d'interagir avec le système à l'aide d'un menu contraint et progressif similaire aux énoncés à produire oralement, ceux-ci constituant ensuite, en quelque sorte, des « raccourcis multimodaux » d'interaction (cf. Figure 149).

Mais encore une fois, d'autres investigations seraient nécessaires, d'une part pour confirmer cette éventuelle cause d'erreur chez les « Non informaticiens », mais aussi, d'autre part, pour identifier d'autres causes éventuelles d'erreur.

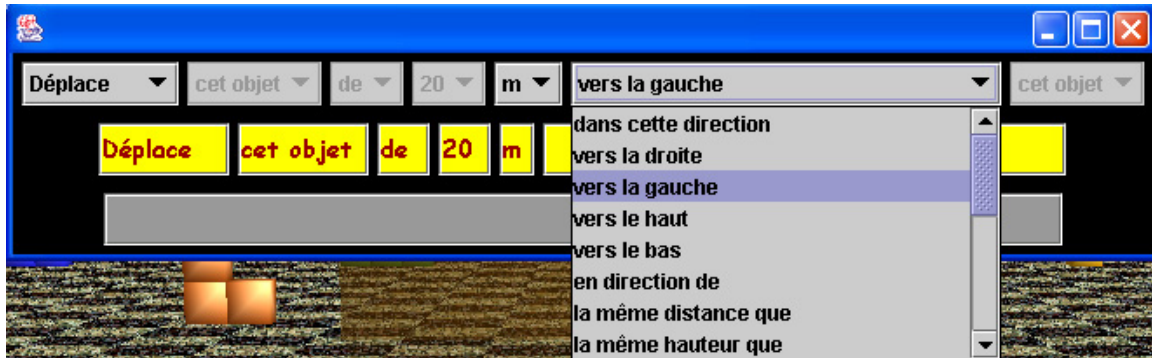


Figure 149. Exemple de menu contraint progressif permettant l'usage de modalités classiques d'entrée du système, et menant l'utilisateur à la production d'énoncés similaires à ceux d'une interaction interpersonnelle en situation de commande.

7.2 Conclusion au vu des résultats.

Avant de rappeler brièvement les éléments de réponse aux questions à l'origine de cette étude, rappelons que cette expérimentation avec des utilisateurs visait à renseigner trois points :

1. d'une part la pertinence d'une interaction orientée communication incluant des actes non verbaux pour effectuer des tâches de mouvements d'objets 3D au sein d'un environnement virtuel,
2. d'autre part tester à cet effet la syntaxe que nous avons définie pour l'usage des modalités gestuelles aux seuls niveaux primitifs de l'interaction (présélections, sélections) ainsi que l'orientation « *langages de commande* » pour la mise en œuvre combinée des trois modalités retenues afin de permettre à l'utilisateur d'effectuer des tâches de mouvements d'objets 3D,
3. et enfin obtenir davantage d'informations sur l'usage de la modalité vocale et son impact d'ensemble sur la formulation d'expressions de commande de ces mouvements d'objets 3D.

7.2.1. Trois interrogations et leurs éléments de réponse

Pertinence d'une interaction orientée communication.

En ce qui concerne le premier point ci-dessus, s'agissant d'une mise en situation d'interaction des utilisateurs avec un nouveau style d'interaction, sans formation ni familiarisation préalable avec le système, les 72,5% d'énoncés conduisant au succès du seul groupe des 10 informaticiens, sinon les quelques 58% de succès pour les 20

utilisateurs, semblent indiquer un résultat qui milite en faveur de la pertinence d'une interaction utilisateur orientée communication avec un environnement virtuel. Du moins en ce qui concerne les mouvements d'objets 3D.

Un autre élément en faveur de ce style d'interaction concerne les durées de mise en œuvre des commandes. Sur l'ensemble des expressions et toutes tâches confondues, ces durées avoisinent en moyenne $12 \pm 1,4$ secondes. Toutefois, la moyenne des meilleurs temps obtenus toutes tâches confondues peut donner une meilleure indication de la durée moyenne de commande à laquelle s'attendre après familiarisation avec le système. Celle-ci avoisine alors quelques $4,9 \pm 1,6$ secondes. On retiendra qu'au cours de cette durée moyenne l'utilisateur spécifie, selon la tâche, trois à cinq paramètres (ou groupes de paramètres) constituant la commande.

Affectation des modalités aux tâches.

Pour le deuxième point évoqué qui vise à déterminer l'adéquation des affectations des combinaisons de modalités aux tâches atomiques, les résultats obtenus ainsi que les erreurs attribuables à un défaut de conception de l'environnement (ex. retour visuel absent) ou relatives à l'incertitude concernant la définition d'une sélection (ex. axe et sens de rotation) confirment l'adéquation de telles associations de modalités pour effectuer les tâches de bas niveau. Rappelons que, pour ce faire, nous avons d'abord distingué deux catégories d'actes utilisateur, d'une part les actes continus éventuellement involontaires voire moteurs mais ajustables « d'indication », ne produisant que des « présélections », et d'autre part les actes utilisateur volontaires, moteurs, ponctuels ou de courte durée conduisant à une interprétation immédiate, produisant des « sélections », seuls ou associés avec une « présélection » pour les tâches à caractère spatial. Cette distinction nous a permis d'adresser un plus large spectre de manifestations utilisateur, et notamment, pour les tâches à caractère spatial, les spécifications de directions.

Ainsi, en ce qui concerne l'association de la direction de la tête et de la parole à laquelle on a affecté la sélection d'une direction sur la scène depuis la position virtuellement occupée par l'utilisateur et la sélection d'objet par visée, nous avons constaté 100% de succès quant à la sélection d'un objet de la scène, et plus de 80% d'usage de cette association pour indiquer une direction égocentrée pour l'ensemble des tâches de positionnement égocentré (F1) et de repositionnement approximatif (F4).

En ce qui concerne l'association d'une Indication manuelle de direction avec la parole, nous lui avons affecté la sélection d'une direction exocentrée (F2) et la sélection d'un axe et sens de rotation par mouvement mimétique de description manuel d'un tracé circulaire (F9). Nous avons constaté que malgré un défaut de présentation d'un retour visuel explicitant une métaphore de relocalisation de l'origine d'une telle direction exocentrée à l'objet sélectionné pour la tâche, plus de 70% des utilisateurs ayant manifestement compris l'objectif de la tâche ont employé

un geste déictique indiquant une direction de translation adéquate. Nous avons d'autre part constaté qu'en reconsidérant le mouvement mimétique indiquant les paramètres géométriques d'axe et de sens de rotation complémentaires à la valeur comme une sélection implicite (au même titre qu'un énoncé, et non plus une simple indication initialement à valider de manière vocale), ce mouvement monomodal de sélection implicite est effectué au cours de 85% des tâches impliquant une telle spécification.

Les sélections de quantité, de direction relative, et d'opération restent monomodaux, mais en relation avec des tâches atomiques multimodales auxquelles elles contribuent à donner une interprétation aboutissant à une commande. Nous avons déjà évoqué l'incertitude due à la variabilité du langage en ce qui concerne la sélection d'opération et nous y reviendrons plus loin au sujet du troisième point à considérer. Pour l'indication d'une direction relative verbale, nous avons constaté qu'elle fut mise en œuvre pour 91,6% des tâches qui la requerraient. Enfin, l'indication de valeur numérique et d'unité géométrique, qui constitue une sélection de quantité géométrique, ne fut omise que pour 3 des 140 tâches impliquant une telle spécification.

Ces résultats semblent confirmer l'intérêt d'une telle approche se démarquant de l'usage de la spécification d'un point en tant que tâche atomique « universelle » (par exemple à l'aide de quelque « pointeur 3D ») au profit de spécifications de « directions » selon nous plus atomiques, sur la base desquelles construire les tâches de bas niveau à caractère spatial (présélections, sélections), jusqu'à leur association en séquences permettant d'adresser les différentes catégories de mouvement d'un objet 3D au sein d'une interaction utilisateur avec un environnement 3D orienté « communication ».

Du « bon » usage de la parole...

En ce qui concerne le troisième et dernier point pour lequel nous cherchons à préciser l'usage de la parole pour structurer les expressions de commande, nous avons établi que le terme d'opération définissant l'un des mouvements considérés, quand il existe, n'est pas couramment employé de manière précise. Nous avons au contraire constaté qu'un nombre restreint de termes imprécis d'opération prédominent pour adresser des sous-ensembles de mouvements de manière générique, cependant que l'usage d'un terme adéquat reste marginal. Nous en déduisons que la notion précise « d'opération à exécuter » ne se détermine pas à l'aide d'un unique terme d'opération mais se trouve distribuée sur plusieurs termes de l'énoncé d'une commande. En particulier, les termes pressentis pour déterminer l'opération à exécuter sont des combinaisons différentes de deux à trois éléments composant l'énoncé. Ces éléments comprennent le verbe utilisé, la préposition introduisant une quantité géométrique, et l'unité de mesure géométrique mise en œuvre au sein de cette quantité.

Cette constatation n'est cependant pas limitative et ne saurait conduire à négliger l'usage correct d'un terme d'opération précis et adéquat dont l'usage permet notamment une économie des termes mis en œuvre, donc de la durée d'énonciation de la commande par l'utilisateur. (Ex. « *Déplace cet objet de deux mètres dans ma direction* » versus « *Rapproche cet objet de deux mètres* »).

Par ailleurs, nous constatons l'absence d'un ordre séquentiel bien établi dans la mise en œuvre des unités syntaxiques. La sélection de l'objet de la tâche peut ainsi précéder celle de l'opération, de même que la sélection d'une direction qui peut elle aussi précéder la quantité.

7.2.2. Amélioration de l'expérience (magicien d'OZ) suite aux problèmes constatés

Laboratoire

Le laboratoire constitue une enceinte fermée d'environ 25m³ qui, en cours d'expérience, contient plusieurs machines en fonctionnement dont deux PC et un projecteur, sources notables de chaleur. Ce laboratoire est par ailleurs exposé au sud sous une toiture de shingle vert kaki, couleur complémentaire des radiations dans le domaine du rouge étendu au moins à une partie des infrarouges solaires. La chaleur régnant dans l'enceinte du laboratoire un après-midi du mois de juillet ensoleillé a contribué à limiter la passation des participants au matin seulement. Outre la chaleur éventuellement incommode pour les participants, une autre raison de cette limitation fut indirecte et liée au bon déroulement de l'ensemble de l'expérience. En effet, alors qu'aucun participant ne s'est plaint de la chaleur, c'est la transpiration de la main des utilisateurs dans le gant qui a partiellement motivé cette limitation. Il fut constaté, lors des pré-expériences de mise au point certains après-midi dans l'enceinte réchauffée par le soleil et les machines, que la transpiration de l'utilisateur au niveau de la main enfermée dans le gant, imprégnait le cuir dont celui-ci est fait jusqu'à affecter le fonctionnement des capteurs par conduction (la sueur contient en effet du sel qui la rend conductrice). Cette constatation relativise la qualité de conception d'un tel « gant de données » artisanal, mais adresse également une éventuelle modification technique du laboratoire (ex. conditionnement d'air).

Une autre amélioration de la conception du laboratoire concerne son impact sur la conception de l'expérience. Le perçage de passages de câbles entre la salle d'expérimentation et la salle d'observation fut exclu car portant atteinte à l'intégrité du bâtiment (et donc soumis à approbation) ainsi qu'au laboratoire lui-même (cette cloison est isolante d'un point de vue sonore, et vraisemblablement constituée d'au moins deux cloisons de parement de forte densité encadrant sans doute un matériau amortisseur de vibrations).

Deux solutions furent envisagées, la première faisant appel au réseau du site de Rocquencourt (parfois encombré, pouvant de ce fait induire des durées de latence), et une solution de transmission radio à travers la baie semi-transparente, qui fut retenue. Le laboratoire pourrait bénéficier à moindre frais soit d'un perçage étudié de passe-cloison isolant, soit d'un aménagement de passage de câble contournant cette cloison par les combles. Ces câbles pourraient ainsi se voir réunis de part et d'autre de la cloison sur des panneaux de connecteurs propres à transmettre diverses sources audio, vidéo, éthernet, optique, etc. dans les deux sens.

Déclinaisons du retour d'information produit par le « magicien » auprès de l'utilisateur.

Nous avons évoqué le rôle, en haut du panneau de commande du « *magicien* », du bouton vert « *Commande non reconnue* » qui fut conçu en tant que retour principal destiné à l'utilisateur sous la forme d'un énoncé synthétique de la phrase « *Commande non valide, veuillez répéter* ».

Cet unique retour fut conçu de manière générique, et s'avéra insuffisant dans plusieurs cas.

- " **Commande Non Reconnue !** " -

Ainsi, un utilisateur (peut-être absorbé par la scène) n'avait pas compris, au discours du système, qu'on en attendait une action.

Après l'énoncé d'un autre participant comportant une erreur d'omission, ce retour sonore invitant à reformuler la commande de manière plus complète fut interprété comme un accès de surdité du système, qui n'eut pour résultat que la reformulation de l'énoncé à l'identique toutefois teinté d'exaspération.

Le déroulement d'une passation invitait les participants à prévenir avant formulation de leur expression de commande, afin notamment de laisser le temps au « *magicien* » de démarrer l'enregistrement ("*Lorsque vous êtes prêt, dites - Je suis prêt! - . Vous pourrez ensuite formuler votre expression après le bip sonore!*"). Toutefois, un autre utilisateur, oubliant cette consigne, formula sa commande à plusieurs reprises sans prévenir, avec pour seul moyen disponible pour obtenir une reformulation de cette « *Commande non valide* », d'interrompre le participant de manière peu adaptée, suivi du « *beep* » produit par le récepteur de commande consécutif à la mise en service immédiate de la fonction d'enregistrement.

Ces trois cas cités mériteraient chacun leur propre retour vocal destiné au participant.

Ces améliorations posent toutefois un problème eu égard à la tâche du « *magicien* ». En effet, celui-ci devant à la fois gérer le déroulement de l'expérience (tâche suivante planifiée) ainsi que la progression de l'expérience au niveau de l'affichage de la machine « *ZardOz* » dont il dispose, (se souvenir de l'étape courante

de l'état du système ou la retrouver rapidement sur l'affichage), il doit également gérer les étapes de l'interaction distribuée sur la scène, « *l'écran d'instruction* » et le récepteur de commande. Sans oublier de mentionner le plus important, l'utilisateur. Le « *magicien* » se trouvant par ailleurs parfois confronté à une situation inattendue qui occasionne alors passagèrement une légère surcharge, ajouter trois « boutons » dédiés à des retours sonores spécialisés demandera sans doute plus d'entraînement à « *l'apprenti sorcier* ».

Cependant, certaines des tâches évoquées ci-dessus peuvent être simplifiée par une reprise du panneau de commande limitant la possibilité de commettre une erreur de « bouton » lorsqu'une tâche est engagée, par exemple en limitant l'accès d'abord aux seules commandes de démarrage d'une présentation, puis l'accès aux seules commandes relatives à la tâche en cours. Ce qui allègerait la tâche du « *magicien* » de la crainte de commettre une erreur telle qu'un déplacement intempestif de plusieurs objets de la scène sous le regard incrédule du participant.

Les problèmes de représentation des tâches.

Après constatation d'une éventuelle difficulté à interpréter la présentation graphique des tâches aux utilisateurs, une discussion informelle permet de dégager une éventuelle amélioration de cette forme graphique de présentation.

La conception des représentations graphiques des objectifs à atteindre pour chaque tâche fut en effet réalisée en sorte d'occuper au mieux une surface rectangulaire d'affichage sur moniteur.

Le défaut éventuellement constaté pour certains utilisateurs consiste en ce que cette présentation implique à la fois une triple translation et une double rotation mentale de la représentation afin d'en obtenir un équivalent transposable à la scène 3D affichée (cf. Figure 150).

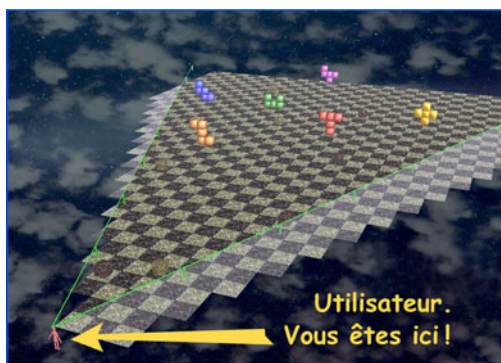


Figure 150. Format de présentation graphique des tâches aux participants.

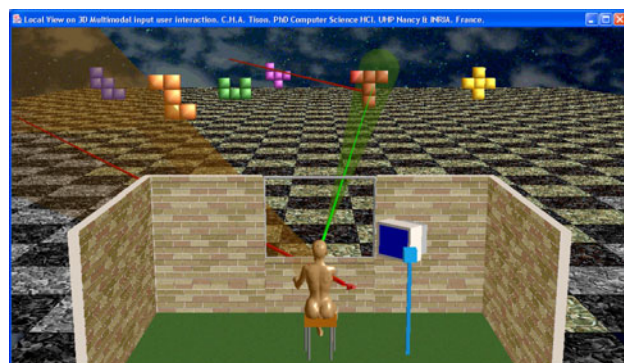


Figure 151. Format de représentation graphique de la restitution des expressions de commande des participants.

Une éventuelle amélioration consisterait à représenter les objectifs à atteindre sous une forme similaire mais n'impliquant que deux translations et une seule rotation à l'image de la représentation choisie pour l'application de restitution multimédia (cf. Figure 151).

Les résultats du comptage des durées

Les résultats du comptage des durées au cours des enregistrements peuvent être automatisés au sein d'un parser XML adressant tous les fichiers d'enregistrement, au lieu d'en éditer les commentaires manuellement sous forme textuelle comme cela fut pratiqué.

7.2.3. Amélioration du système

Solution au manque de retour visuel lors d'une indication de direction exocentrée.

Au moins six des vingt utilisateurs ont signalé une gêne quant au manque de retour visuel lors de l'usage du gant pour indiquer une direction exocentrée. Un tel retour visuel ne devrait pouvoir être représenté qu'à partir de l'objet de la tâche qu'il concerne. Ceci suppose un dialogue de plus haut niveau au cours même de l'expression de commande, lorsque le système est en mesure d'identifier sans ambiguïté la tâche (du moins la catégorie « mouvement »), ainsi que son objet. Cette critique conforte cependant quelque peu la notion de présélection dans son rôle d'état maintenu et ajustable en attente de validation. En effet, l'absence de retour visuel de direction exocentrée ne permet pas, d'une part, à l'utilisateur d'être informé de la prise en compte d'une quelconque indication de direction exocentrée, mais ne permet pas non plus d'ajuster plus précisément cette direction à l'aide d'un retour visuel. Une amélioration notable du système consisterait donc à lui ajouter la production d'un tel retour visuel de direction exocentrée.

Une autre solution éventuellement envisageable consisterait à fournir un tel retour visuel d'indication d'une direction exocentrée de la part de l'utilisateur lorsque celui-ci met en œuvre cette indication bimodale simultanée de présélection d'objet et d'indication simultanée de direction exocentrée. A condition que cela n'entre pas en conflit avec une même action combinée pour une autre tâche d'un autre groupe de tâches.

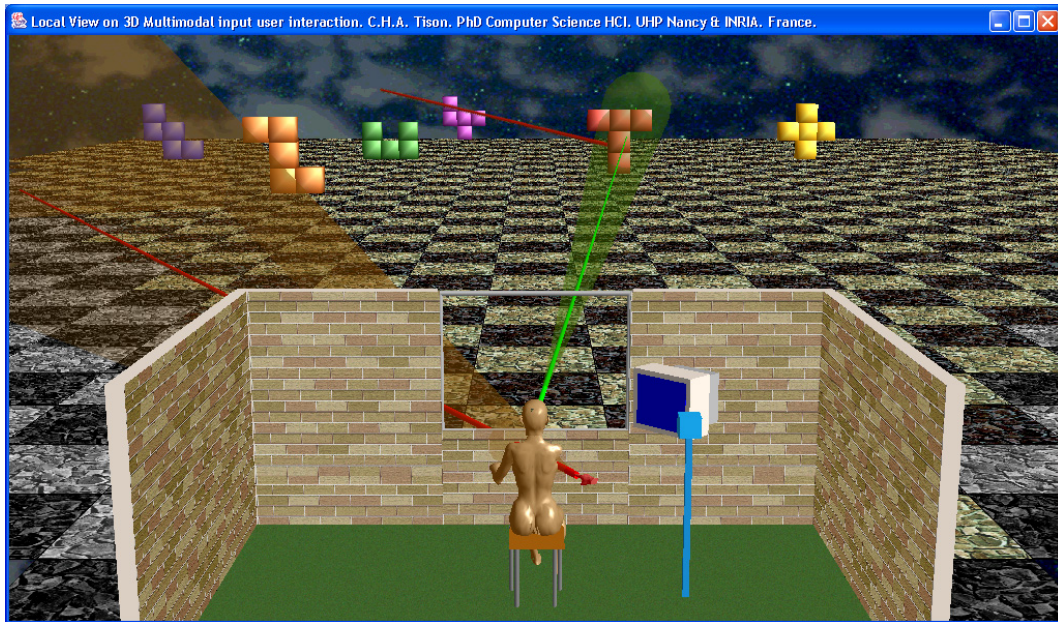


Figure 152. Exemple de représentation visuelle d'une indication de direction exocentrée (transposée ici par montage photo sur une représentation de l'application de restitution multimedia des enregistrements).

Automatisation de l'interprétation de gestuelle mimétique de la main

Rappelons ici certains des résultats obtenus pour la tâche de rotation d'un objet de la scène.

Sur les vingt expressions produites par l'ensemble des utilisateurs, neuf laissent apparaître un problème d'indication de direction (45%), laquelle direction fut définie comme devant être indiquée de manière mimétique, avec un énoncé lui faisant référence.

Sur ces neuf expressions n'indiquant pas le mouvement à effectuer de façon bimodale claire ou complète, six d'entre elles manifestent un geste mimétique manuel correspondant à la tâche.

Un compromis pourrait consister, afin de permettre une réponse du système, à autoriser la manipulation directe pour les seules tâches de rotation. Cependant, l'irruption de la manipulation directe irait ici à l'encontre du critère d'homogénéité/cohérence établi par Bach & Scapin (2003). Une autre solution plus adaptée pourrait toutefois consister à rendre la gestuelle manuelle plus « *proactive* », en partageant la priorité donnée arbitrairement jusqu'alors aux modalités vocales.

Cette solution répondrait, en la confortant, à la définition que nous avons retenue pour la notion de « validation » d'une « présélection » (ou indication ajustable d'un acte utilisateur continu éventuellement involontaire) à l'aide d'un acte utilisateur volontaire, moteur, ponctuel ou de courte durée conduisant à une interprétation immédiate.

Certains mouvements observables de la main acquerraient ainsi le même niveau de priorité que l'énoncé vocal d'une opération en constituant des sélections à elles seules. (Le traitement de ces éléments de gestuelle manuelle fut l'objet d'une proposition, en 2004, d'un stage de reconnaissance de formes resté sans suite).

En d'autres termes, une prise en compte automatique du seul mouvement manuel circulaire adoptant une pose déictique de la main sans référence vocale nécessaire, lorsqu'est évoquée une rotation sur un objet, réduirait le taux de cette « erreur d'indication de direction » des deux tiers. Réinterprétées individuellement en considérant ce mouvement mimétique comme une sélection, les productions des participants conduisant à un succès seraient ainsi portées à un effectif de 17 au lieu de 9, soit 85% et non plus 45% de succès pour cette tâche de rotation.

Chapitre 8. Conclusion

Le but de cette thèse consistait à établir, justifier et mettre en oeuvre une construction théorique permettant finalement à l'utilisateur d'interagir avec des objets graphiques 3D de manière "efficace".

D'un point de vue IHM pour les environnements 3D, l'expérience décrite adresse l'applicabilité d'une combinaison de modalités afin de permettre à l'utilisateur de communiquer des informations en entrée d'un système d'interaction 3D. Les tâches utilisateur retenues pour l'étude consistaient en des repositionnements d'objets 3D au sein d'une scène réactive affichée.

La détermination de modalités d'entrée reposa sur une approche dite « top-down » à partir d'interactions utilisateurs observées dans le domaine de la conception textile vestimentaire au cours d'une activité « *d'inspection de prototype* ».

Un ensemble de tâches utilisateur fut défini en respect des contraintes de l'activité. L'examen de ces tâches en vue d'en obtenir une définition du plus général au plus particulier, souleva d'abord un problème d'atomicité des tâches à considérer au bas niveau d'interaction. Ceci nous conduisit à nous démarquer de la notion de « *pointeur 3D* » au profit d'une notion plus atomique, à notre sens, de « *direction* » indiquée par l'utilisateur.

De telles indications de « *direction* » constituant ces tâches de bas niveau permirent de composer des tâches de niveau plus élevé, ou tâches de présélection et de sélection, participant de l'interaction.

L'examen des modalités observables qui composent les expressions de commande chez l'utilisateur au cours de son activité, afin d'en retenir les plus pertinentes, mena à considérer une interaction orientée communication d'information en entrée du système. Plusieurs « *gestuelles* » furent alors définies pour distinguer en particulier les mouvements de la tête et les mouvements des mains comme media d'expression complémentaires de la parole.

Ces trois modalités d'expression (parole, gestuelle de la tête, gestuelle des mains), apparurent en effet constituer le meilleur rapport entre effectif de modalités à mettre en oeuvre et spectre d'expression utilisateur qu'elles permettent.

L'examen de chacune des trois modalités retenues permit ensuite de définir les critères d'un profil théorique de dispositifs propres à mesurer et convoier les informations exprimées dans les modalités choisies. Un examen de la littérature visa parallèlement à situer, quant il existait, un dispositif répondant à ces critères pour chaque modalité retenue afin de permettre, de la part d'un utilisateur, sa mise en oeuvre au sein de tâches d'interaction définies.

Les trois modalités se virent alors affecter des rôles expressifs complémentaires en fonction de leurs compétences particulières. Ces rôles furent ensuite testés

séparément et de manière combinée au cours de scénarios d'interaction afin d'en homogénéiser le style de mise en oeuvre. Ceci nous conduisit à distinguer et préciser d'une part la notion de « *présélection* » en tant qu'acte utilisateur éventuellement continu, involontaire, mais ajustable et révélant éventuellement un objet « d'intérêt », de celle, d'autre part, de « *sélection* » réalisée à l'aide d'un événement ponctuel, du moins de courte durée produisant une interprétation immédiate, mis en oeuvre de façon volontaire, éventuellement motrice, pour effectuer la « *validation* » d'une « *présélection* » en cours.

Le style d'interaction résultant nécessitait alors un test visant d'une part à en confirmer la pertinence et à étudier la façon dont ce style serait mis en oeuvre par des utilisateurs, et, d'autre part, à en préciser la grammaire nécessaire à ce niveau d'interaction.

L'expérience proposée pour répondre à ces questions consista en un exemple de test utilisateur connu sous le nom de « *Magicien d'Oz* ». Cette expérience fut réalisée par la mise en oeuvre d'un test d'acceptance modifié et de techniques apparentées à celles de la post-production cinématographique, sous la forme de tests de performance à des tâches de manipulation d'objets 3D.

D'un point de vue informatique, l'expérience s'appuya sur la réalisation d'une application 3D interactive répartie, permettant l'enregistrement des comportements des utilisateurs au cours d'interactions multimodales avec des objets 3D. Elle s'appuya également sur la réalisation d'une seconde application de restitution 3D de ces enregistrements à des fins d'observation, destinée à l'expérimentateur.

Les résultats obtenus semblent, pour des tâches de manipulation d'objets dans l'espace, confirmer la pertinence d'une interaction orientée « Communication Homme-Machine » sur la base d'un nombre réduit de modalités d'entrée sélectionnées.

8.1 Travaux complémentaires à envisager

Les résultats obtenus militent pour l'approfondissement d'un tel style d'interaction, nous allons brièvement évoquer ici d'une part les travaux qui sont à réaliser pour obtenir davantage d'information afin d'aboutir à un système fonctionnel, et d'autre part les perspectives d'utilisation de ce style d'interaction proche de l'interaction « Homme-Homme » à statut social dissymétrique en situation de commande.

Outre les améliorations dont il a déjà été question en ce qui concerne l'interaction, l'étude de la désignation d'un *locus géométrique* à l'aide de la visée, par exemple au point d'intersection avec un plancher de référence, devrait être évaluée même dans le cas où son utilité s'avérerait mineure. Ce phénomène fut incidemment observé chez l'un des utilisateurs produisant l'énoncé « *Met cet objet, ici!* », en positionnant le viseur sur le lieu de destination approximatif à la surface du plancher. Cette expression était alors inattendue pour la tâche proposée et fut spontanément

corrigée, elle révèle néanmoins la richesse d'expression d'une telle multimodalité en entrée du système et mérite sans doute davantage d'attention. En effet, L'étude ayant été volontairement restreinte aux désignations d'objets graphiques et de destinations essentiellement composées de directions munies d'une quantité géométrique, cette catégorie supplémentaire de sélection d'un locus, qui n'a pas été envisagée, enrichirait le plus bas niveau d'interaction.

Au titre des « *positionnements* », les relocalisations d'un objet graphique de la scène par rapport à deux entités sélectionnables peut se révéler d'intérêt. Dans cette catégorie de placements se trouvent par exemple les énoncés tels que « *place cet objet entre cet objet et cet objet!* ». De tels positionnements impliquent alors la prise en compte de trois objets désignés au cours de l'expression de commande. Par ailleurs, un « *positionnement angulaire égocentré monomodal* » est envisageable quoiqu'éventuellement de moindre utilité. Une telle expression de commande reprendrait par exemple la formulation d'une direction telle que « *à dix heure* » par rapport à l'orientation de l'utilisateur vis-à-vis d'un plancher de référence (« *midi* » se trouvant toujours face à l'utilisateur).

D'autres « *déplacements* » sont également à étudier en ajoutant d'une part une caractéristique « *d'orientation* » de l'environnement lui-même, d'autre part une caractéristique « *d'orientation intrinsèque* » d'un objet graphique 3D.

Dans le premier cas, l'environnement étant globalement « *orienté* », par exemple avec un « *nord* », un « *zénith* » voire un « *azimut* », cela permet par exemple d'envisager des « *déplacements* » d'objets graphiques d'une certaine distance « *vers le sud* ».

Dans le second cas, les objets 3D, qui disposent pour nombre d'entre eux d'un « *haut* » et d'un « *bas* », peuvent également disposer d'une « *face avant* » qui permettrait au système de répondre à des sollicitations telles « *qu'avancer* » tel ou tel objet graphique d'une certaine distance.

Puisque nous évoquons maintenant la nature d'un objet graphique 3D (orienté avec une « *face avant* »), une autre étude à réaliser concerne la sélection monomodale d'un objet graphique de la scène. Nous avons vu que seule l'énonciation d'un nom de classe générique global fut envisagée pour valider la sélection d'un objet graphique. L'appartenance d'un objet 3D à une sous-classe « *d'objet* » (munie d'un nom de sous-classe) permettrait par l'énonciation de son appellation d'apporter une information supplémentaire de désambiguation dans un environnement densément peuplé. Ainsi, la sélection de « *cette chaise* » là où la ligne de visée pourrait désigner la « *table* » ou chacun des « *objets* » qui y sont disposés, se verrait effectuée de façon plus efficace. De même l'utilisateur pourrait-il tirer avantageusement parti d'une sélection impliquant l'examen d'un attribut de l'objet graphique tel que sa « *couleur* ». Une telle méthode de sélection permettrait par exemple également la désambiguation d'une désignation de « *cette voiture rouge* » sur un parking bondé.

Par ailleurs, en ce concerne les objets graphiques dits « composites », plusieurs méthodes de sélection proche d'une « exploration » d'un objet graphique sont à étudier. L'exemple évoqué à Rennes avec T. Duval en 2003 fut celui de la sélection du carburateur du moteur d'une voiture. En particulier, la mise en œuvre d'un accès direct au carburateur semble problématique.

Spin3D utilise par exemple à cet effet des « boîtes englobantes progressives » à comportement dynamique qui réagissent en fonction de la proximité du pointeur (voir Dumas, 1999).

Pour évoquer maintenant une autre catégorie importante de tâches utilisateur à envisager, une étude est à réaliser concernant la « *locomotion virtuelle* » de l'utilisateur au sein de l'environnement. Cette catégorie est conceptuellement différentes des relocalisations d'objets graphiques 3D, mais en reste proche du point de vue de sa mise en œuvre. Une telle considération est partiellement justifiée par les commentaires d'un utilisateur qui, pour la tâche de « *rotation* » d'un objet graphique « *d'un quart de tour vers la droite* », indiquait de cette manière le sens de rotation par identification au mouvement auquel il aurait procédé sur sa propre personne d'un point de vue locomoteur. D'où la possibilité d'envisager la prise en compte, par le système, d'un énoncé tel que « *Tourne moi d'un quart de tour vers la droite !* », ou encore « *déplace-moi de 10m dans cette direction !* ».

Munir l'utilisateur d'une « locomotion » au sein de l'environnement permettrait d'enrichir les tâches de relocalisation d'objets graphiques à l'aide de la métaphore du « *transporteur* ». Un objet graphique peut en effet être rendu solidaire de la position de l'utilisateur. Celui-ci, lors de mouvements de locomotion, peut alors déplacer l'objet ainsi « *arrimé* » au gré de ses propres mouvements virtuels. Toutefois, l'usage d'une telle métaphore pouvant se décliner en plusieurs solutions possibles de mise en œuvre, celles-ci mériteraient par conséquent une étude comparative de performances utilisateur.

D'un point de vue plus sociologique, l'étude équivalente de l'expression de ces relocalisations d'objets dans une autre langue (ex. anglais) permettrait d'étudier l'impact des différences culturelles sur l'usage d'une telle multimodalité d'entrée impliquant la parole. Une évaluation de cet impact permettrait alors de mesurer la capacité du système à être adapté à d'autres cultures.

Les tâches transversales dites « de service » doivent également être traitées en vue de pouvoir conserver tout travail de construction ou de modification d'une scène 3D. Ces tâches concernent l'addition d'objets à la scène, l'enregistrement des objets graphiques, de leurs caractéristiques, de leur disposition au sein de la scène, etc.

Deux aspects « informatiques » sont également à considérer. Il s'agit d'une part du noyau fonctionnel de l'application qui doit abriter les fonctions correspondantes aux différentes opérations à exécuter, et d'autre part la gestion de l'ensemble de l'interaction. Si les fonctions à exécuter sont relativement connues dans le domaine

du graphisme 3D, le mécanisme de fusion des informations de modalités en entrée en vue d'une sélection adéquate de l'opération à exécuter est plus problématique et mérite à lui seul une étude. Le style d'interaction choisi nécessite en effet une évaluation des unités syntaxiques décrites au fur et à mesure de leur production afin de permettre l'ajustement des présélections de bas niveau (ex. direction exocentrée), mais nécessite également dans certains cas une réévaluation de l'expression de commande complète afin par exemple de permettre au système de « proposer », le cas échéant et au cours d'un dialogue de haut niveau, plusieurs réponses à une expression de commande ambiguë. Dans les cas de mouvements d'objets que nous avons considéré, cela consisterait par exemple à éviter l'exécution immédiate d'une *rotation* (« *Tourne cet objet d'un quart de tour...* », avec indication mimétique manuelle d'axe et de sens de rotation) quand une « *giration exocentrée* » est indiquée (« *...autour de cet objet!* »). L'évaluation tardive de l'ensemble de l'expression permet effectivement de déterminer le mouvement demandé, alors qu'une évaluation automatique de chaque nouvelle unité syntaxique de l'expression de commande au fur et à mesure de leur production conduit à l'interprétation du cas particulier que constitue la rotation.

Enfin, l'expérience avec des utilisateurs peut maintenant également être envisagée à l'aide cette fois d'une familiarisation des utilisateurs avec le système afin de mesurer l'amélioration résultante de ce type d'interaction. Une amélioration sensible des résultats après une telle familiarisation confirmerait l'intérêt de ce style d'interaction avec des environnements 3D.

Une telle familiarisation peut prendre différentes formes. La première forme de familiarisation à court terme avec le système consisterait à mettre en oeuvre une documentation plus étendue contenant des exemples.

De tels exemples restaient peu envisageables au cours de cette première expérience d'une part du fait que le contenu des expressions à fournir restait hypothétique, et d'autre part du fait qu'il était nécessaire d'éviter d'influencer les expressions des utilisateurs.

Nous disposons maintenant d'un échantillon d'expressions de commande correspondant à différentes relocalisations géométriques dans le plan horizontal. Ces expressions peuvent donc désormais être utilisées afin de servir d'exemple pour ces mêmes mouvements.

En ce qui concerne une documentation écrite détaillant les exemples d'expressions de commande, on retiendra que le document abstrait et approximatif utilisé pour cette première expérience réalisée fut déjà considéré par les utilisateurs comme très riche en information.

Une autre catégorie de familiarisations consiste en la démonstration d'exemples de fonctionnement de l'interaction. On trouvera, dans cette catégorie, des vidéos d'exemples de mise en oeuvre de commandes, montrant par exemple simultanément

un utilisateur exprimant une commande ainsi que la scène graphique sur laquelle sera montré le résultat de la commande.

Deux catégories d'expériences peuvent donc maintenant être menées. La première consiste à mesurer, le cas échéant, l'amélioration des résultats obtenus par des utilisateurs non-informaticiens suite à une familiarisation de ceux-ci avec le système à l'aide d'exemples démonstratifs (ex. enregistrements vidéo de la mise en scène d'une séquence d'interaction reprenant les termes les plus courants).

Une seconde catégorie d'expérience consiste à familiariser les utilisateurs avec ce style d'interaction à l'aide des mêmes exemples, pour ensuite leur soumettre des objectifs à atteindre pour d'autres tâches non étudiées. De tels exemples permettraient ainsi une "appropriation" plus efficace du style d'interaction par les utilisateurs en vue d'obtenir des expressions pertinentes pour de nouvelles tâches à étudier.

Ces solutions impliquent une lourde mise à jour en cas de modification même mineure au sein de l'interaction, telle par exemple qu'une modification de forme ou de couleur du "viseur".

Toute documentation établie de cette manière est menacée d'obsolescence dès lors qu'on a affaire à un produit susceptible d'évoluer afin d'en améliorer l'efficacité. L'aspect visuel étant ici prépondérant, nombre de modifications liées à l'apparence de la présentation de l'environnement et des outils d'interaction impliqueraient une remise à jour complète de la documentation à seule fin de la maintenir conforme au style d'interaction pour les seuls besoins de l'expérience.

Une autre solution consiste alors à faire appel à un enseignant. L'expérimentateur peut être envisagé comme "expert" démonstrateur des commandes possibles avant l'expérience, cette solution implique alors une action de la part d'un tiers (l'expérimentateur) qui doit donc être disponible.

Une solution supplémentaire possible consiste à insérer, au sein de l'application, un exemple pour chaque tâche identifiée à l'aide d'une marionnette animée multimodale en sortie au sein de l'environnement de travail. L'utilisateur disposerait alors d'un menu d'exploration des exemples de tâches par catégorie (par exemple les déplacements d'objets) et assisterait alors à une interaction d'une représentation d'un utilisateur de synthèse en interaction avec le reste de l'environnement pour la tâche considérée. L'interaction restant basée sur des commandes sans contact avec les objets des tâches, une telle marionnette animée n'impliquerait qu'une animation en cinématique dite "avant" (pas de "cinématique inverse" plus coûteuse en calculs). Par ailleurs, l'essentiel de l'interaction au plus bas niveau restant basé sur la manifestation de directions de la tête et des mains, (comme on peut le voir sur les "restitutions" multimédia des expressions des utilisateurs), ces sous-ensembles graphiques de la marionnette seraient les seuls à contrôler avec soin. Les énoncés de commande vocale n'impliqueraient pour leur part qu'une synthèse vocale sommaire

de textes préprogrammés correspondant à la tâche, afin que l'utilisateur puisse en bénéficier. La commande du système s'effectuant pour sa part sur la base des termes de l'énoncé produit de la même manière qu'ils seraient reçus de la part de l'utilisateur.

L'intérêt d'une telle approche consisterait en ce que des modifications ultérieures de la présentation de l'environnement et de l'interaction n'affecteraient pas de telles démonstrations automatisées. La marionnette ne constituerait qu'un objet informatique et graphique doué d'un comportement spécifique à chaque exemple de tâche, faisant appel au même environnement présenté et aux mêmes commandes que celles mises à disposition de l'utilisateur. Ce qui reviendrait finalement à placer une marionnette aux commandes du système à la place de l'utilisateur.

De telles marionnettes existent pratiquement, désormais, de manière standardisée pour figurer des êtres humains au sein d'un environnement 3D, par exemple, le projet "H-Anim" (<http://h-anim.org/>), ou encore les marionnettes interactives que Ken Perlin présenta au salon Imagina 2004 (Perlin, 2004), dont une démonstration est disponible sous forme d'applet sur son site web.

Pour terminer, il peut être souhaitable de comparer la forme d'interaction développée au cours de ce travail avec celle obtenue à l'aide d'autres systèmes. En particulier, nous avons supposé, sans en avoir encore la confirmation, qu'une telle méthode d'interaction pouvait, pour effectuer des relocalisations d'objets, s'avérer plus avantageuse en terme de durée et de facilité de mise en oeuvre pour l'utilisateur, que d'autres méthode utilisant le pointeur 3D ou encore des dispositifs d'entrée différents impliquant une manipulation directe pour la relocalisation d'objets 3D.

L'une des principales difficultés d'une telle comparaison résidera dans le fait que chaque méthode d'interaction apportant des avantages pour certaines tâches spécifiques auxquelles elle donne accès, il pourra s'avérer malaisé de trouver des tâches équivalentes à comparer. Par exemple, comparer la sélection d'un objet graphique par visée avec une sélection équivalente par manipulation directe à l'aide d'un pointeur impliquera différents cas de figure en fonction de la distance de l'objet graphique à sélectionner et de ses dimensions (donc son diamètre apparent), de la présence d'autres objets situés à différentes distances dans la même ligne de visée, etc.

Cette remarque permet notamment d'envisager qu'il pourra éventuellement en résulter une complémentarité des techniques d'interaction centrées utilisateur avec celles de manipulation directes et celles qui mettent en oeuvre un pointeur 3D.

Enfin, une comparaison entre les tâches utilisateur expérimentées à l'aide des modalités d'entrée utilisées pour l'expérience décrite et les mêmes tâches réalisées à l'aide d'un équipement plus classique "clavier-souris" en entrée est également possible. Nous avons signalé la possibilité de mettre en oeuvre un menu graphique intermédiaire de commande (cf. Figure 149 page 227) pour les utilisateurs novices

afin de les orienter vers des expressions syntaxiques contraintes correspondant à la syntaxe générale de commande que nous avons définie. Cette technique intermédiaire, utilisant la même métaphore d'interaction ainsi que la même scène avec les mêmes objets graphiques permettrait d'effectuer à nouveau les mesures en ne faisant varier que les dispositifs d'entrée.

8.2 Perspectives

En ce qui concerne les perspectives d'utilisation proprement dites, nous avons convenu que l'étude portait initialement sur un environnement orienté « conception » dont l'évaluation par l'utilisateur s'effectuait de manière essentiellement visuelle. Les applications envisageables concernent donc directement les activités d'aménagement de l'espace ou encore de composition esthétique de l'environnement à des fins, par exemple, d'études préalables à un aménagement de l'environnement réel (urbanisme, architecture, paysagisme, restauration de monuments, etc.).

Par extension à la locomotion de l'utilisateur, ce style d'interaction concerne également les environnements à caractère ludique. Cette catégorie d'applications est en partie à considérer du fait qu'elle fut le principal promoteur du développement de l'équipement multimedia aujourd'hui standard d'un ordinateur grand public, en particulier en ce qui concerne les cartes graphiques actuelles.

La perspective d'une application ludique au sein de laquelle l'utilisateur peut explorer et modifier l'environnement à l'aide d'un comportement multimodal de commande similaire à une interaction Homme-Homme de type maître-élève, pourrait alors conduire à d'éventuelles applications d'enseignement d'interactions Homme-Homme (*team building*) au sein desquelles, par exemple, les participants étudieraient des stratégies à mettre en œuvre. De telles applications concernent par exemple les environnements réels pour lesquels une action coordonnée de plusieurs participants est nécessaire en vue de la réalisation d'un objectif (ex. lutte contre l'incendie, participation à une chaîne de commandement organisée de façon pyramidale, etc.).

Par extension aux environnements augmentés, la commande d'un véhicule automoteur informatisé est également envisageable au sein d'une réplique virtuelle d'un environnement réel, ne serait-ce qu'en tant « *qu'objet* » d'un « *déplacement* ». Dans une moindre mesure mais avec un plus large spectre d'applications, l'adressage de composants augmentés de l'environnement réel en vue d'en modifier l'état interne est à étudier. Un exemple d'application « domotique » de bas niveau consiste ainsi à modifier l'état d'un éclairage à l'aide de commandes simple (allumer, éteindre). Ce style d'interaction grand-public présente un intérêt évident pour les personnes souffrant d'une mobilité réduite au sein de leur propre environnement réel mais disposant toutefois de deux des modalités d'entrée choisies (parole, mouvements de la tête). De telles applications permettraient une gestion statique de l'environnement

domotique (éclairage, températures, ventilations, hygrométries, isolement tels que commande de volets, de verrouillages de portes, etc.), mais concerneraient également les applications industrielles de contrôle-commande (commande d'électrovannes, de mises en service d'actuateurs ou de dispositifs actionneurs divers, etc.). Il faudra toutefois, pour des applications dites « critiques », par exemple en milieu hostile, disposer pour ce faire d'une interaction particulièrement robuste.

Chapitre 9. Acronymes.

AR :	Réalité Augmentée (" <i>Augmented Reality</i> ").	p. 12
AV :	Virtualité Augmentée (" <i>Augmented Virtuality</i> ").	p. 12
CAD :	" <i>Computer Assisted Design</i> ".	p. 22
CAO :	Conception Assistée par Ordinateur.	p. 11
CSCW :	Collecticiel, (" <i>Computer Supported Cooperative Work</i> ").	p. 23
EMG :	Electro-MyoGramme, Electro-Myographie.	p. 68
EPM :	" <i>Extent of Presence Metaphor</i> ".	p. 14
EWK :	" <i>Extent of the World Knowledge</i> ".	p. 13
GPS :	" <i>Global Positioning System</i> ".	p. 61
GUI :	Interface Graphique utilisateur, (" <i>Graphical User Intereface</i> ").	p. 74
HMD :	Lunette d'affichage portables, (" <i>Head-Mounted Displays</i> ").	p. 46
HMT :	" <i>Head Magnetic Tracker</i> ", parfois aussi " <i>Hand Magnetic Tracker</i> ", (cf MT).	p. 62
HRTF :	" <i>Head Related Transfer Functions</i> ".	p. 50
HWD :	Synonyme de HMD, (" <i>Head-Worn Displays</i> ").	p. 46
IHM :	Interaction Homme-Machine.	p. 11
ILD :	" <i>interaural level difference</i> ".	p. 51
ITD :	" <i>interaural time difference</i> ".	p. 51
LED :	Diode Electro-Luminescente, (" <i>Light Emitting Diode</i> ").	p. 78
MR :	Réalité Mixte (" <i>Mixed Reality</i> ").	p. 12
MT :	" <i>Magnetic Tracker</i> ".	p. 62
NPR :	Rendu Non-Photorealiste, (" <i>Non Photorealistic Rendering</i> ").	p. 38
PACG :	" <i>Perceptually-Adaptive Computer Graphics</i> ".	p. 70
VB :	" <i>Virtual Body</i> ".	p. 15
VE :	Environnement virtuel (" <i>Virtual Environment</i> ").	p. 12
VRD :	Affichage par "projection rétinienne", (" <i>Virtual Retinal Display</i> ").	p. 47
WIMP :	" <i>Windows-Icon-Menu-Pointing</i> ".	p. 55
WoW :	" <i>window-on-the-world</i> ".	p. 14, 15

Chapitre 10. Références.

- Amicis R., Fiorentino M. & Stork A., 2001, "Parametric Interaction for CAD application in Virtual Reality Environment", *International Conference on Design Tools and Methods in Industrial Engineering*. Proceedings 2001.
- "Archeoguide", 2002, Projet Européen dont le premier prototype fut présenté à IMAGINA 2002. <http://archeoguide.intranet.gr/project.htm>.
- Arns, L., and Cruz-Neira, C., 2002, "A New Taxonomy for Locomotion in Virtual Environments", *Proc. of 4th Virtual Reality International Conference*, Laval, France, June 2002, pp.141-149.
- Austin J. L., 1962, "How to Do Things With Words". *Clarendon Press*. Trad. Française, "Quand dire c'est faire", Le Seuil, 1970.
- Azuma R., 1993, "Tracking Requirements for Augmented Reality", *Communication of the ACM*, Jul 1993/Vol.36, No.7 p.50-51.
- Azuma R.T. , 1999, "The Challenge of Making Augmented Reality Work Outdoors In Mixed Reality: Merging Real and Virtual Worlds", *Yuichi Ohta and Hideyuki Tamura (ed.), Springer-Verlag, 1999*. Chp 21 pp. 379-390. ISBN 3-540-65623-5.
- Bach C., 2000, "Inspection ergonomique d'un logiciel de CAO destiné aux modélistes de l'industrie du textile", *Rapport de recherche*, Inria, Rocquencourt.
- Bach & Scapin, 2003, « Recommandations ergonomiques pour l'inspection d'environnements virtuels ». *Rapport de contrat Eureka-Comedia*, INRIA, Rocquencourt.
- Ballegooij (van) A. & Eliéns A., 2001, "Navigation by query in virtual worlds". *Proc. of VRML2001*. Pages: 77 – 83.
- Baxter B., Scheib V., Lin M. C. & Manocha D., 2001, "DAB: interactive haptic painting with 3D virtual brushes", *Proc. of the 2001 conference on Computer Graphics*, SIGGRAPH 01, pp. 461 - 468.
- Y. Bellik, "Interfaces Multimodales : Concepts, Modèles et Architectures", *Thèse de Doctorat*, Université Paris XI, Orsay, 1995
- Bernsen N. O., 1994, "Foundations of Multimodal Representations: A Taxonomy of Representational Modalities", *Interacting with Computers 1994* v.6 n.4 p.347-371.
- Berthoz, A., 1997. "Le sens du mouvement". Odile Jacob, Paris, 1997. ISBN 2-7381-0457-6

-
- Biermann H., Kristjansson D. & Zorin D., 2001, "Approximate Boolean operations on free-form solids", *Proc. of the 2001 conference on Computer Graphics, SIGGRAPH 01*, pp. 185 - 194.
- Billingham M., Weghorst S. & Furness III T., 1997, "Wearable Computers for Three Dimensional CSCW", *Proc. of ISWC97* 97-10 .
- Billingham M., Poupyrev I., Kato H. & May R., 2000, "Shared Space: Mixed Reality Interface for Collaborative Computing", *Imagina 2000*.
- Bolt R.A., 1980, "Put-That-There", *Computer Graphics*, vol.14, pp. 262-270, 1980.
- Bowman, D. and Hodges, 1997a, L. "User Interface Constraints for Immersive Virtual Environment Applications", *Graphics, Visualization, and Usability Center Technical Report GIT-GVU-95-26*.
- Bowman D. & Hodges L., 1997b, "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments", *Proc. of the 1997 Symposium on Interactive 3D Techniques*, page 35.
- Bowman, D. Koller, D., and Hodges, L., 1998, "A Methodology for the Evaluation of Travel Techniques for Immersive Virtual Environments", *Virtual Reality: Research, Development, and Applications*, vol. 3, no. 2, 1998, pp. 120-131.
- Bowman, D. A., 2002, "Principles of the Design of Performance-oriented Interaction Techniques". In K. Stanney (Ed.), *Handbook of Virtual Environments* (pp.277-300). Lawrence Erlbaum.
- Bremer D. H. & Hugues J. F., 1998, "Rapid Approximate Silhouette Rendering of Implicit Surfaces", *Implicit Surfaces '98*. Voir aussi <http://www.cs.brown.edu/people/jfh/is/img/bunny3.html>.
- Butterworth J., Davidson A., Hench S. & Olano M.T., 1992, "3DM: a three dimensional modeler using a head-mounted display ", *SI3D92* Pages 135 - 138.
- Cadoz C., 1994, "Le geste, canal de communication homme/machine : la communication "instrumentale"; ", ACROE (Association pour la création et la recherche sur les outils d'expression), ministère de la Culture et de la Francophonie. LIFIA (Laboratoire d'informatique fondamentale et d'intelligence artificielle, unité de recherche associée au CNRS, IMAG), *Technique et science informatiques*. Volume 13 - n0 1/1994, pages 31 à 61.
- Carbonell, N., Dauchy, P., 1999, "Empirical data on the use of speech and gestures in a multimodal human-computer environment", in H.-J. Bullinger, J. Ziegler (Eds.), *Proc. of HCI International'99*, Munich, August 22-26 1999, Mahwah (NJ), London: Lawrence Erlbaum Associates, Vol. 1, pp. 446-450, 1999.
- Carbonell N., 2003, "Recommendations for the design of usable multimodal command languages", *Universal Access in the Information Society Journal (UAISJ)*, Special issue 'Multimodality, a step towards universal access'.

-
- Chalmé S., Briffault X., Denis M. & Gaunet F., 1999, "Spécifications d'interfaces multimodales de dialogue dans les systèmes d'aide à la navigation : expérimentations en conduite réelle et sur simulateur", ("Experiments for designing multimodal dialogue interfaces in navigational aid systems : real versus simulated driving situations") - *Proc. of the Driving Simulation Conference, 1999*. (DSC'99, Session 3 : Études perceptionnelles et ergonomiques / Perception and ergonomic studies).
- Coutaz J., 1992, "Multimedia and Multimodal User Interfaces: A Taxonomy for Software Engineering Research Issues", *Proc. of EWHCI 92*, p.229 .
- Coyne B. & Sproat R., 2001, "WordsEye: an automatic text-to-scene conversion system", *Proc. of the 2001 conference on Computer Graphics, SIGGraph 2001*, pp. 487 - 496.
- Danforth R., Duchowski A., Geist R. & McAliley E., 2000, "A Platform for Gaze-Contingent Virtual Environments", *Proc. of AAAI-2001*.
- Darken R. P., Cockayne W. R. & Carmein D., 1997, "The omni-directional treadmill: a locomotion device for virtual worlds", *Proc. of UIST 97*, pp. 213 – 221.
- Darken R., Allard T. & Achille, L., "Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces: an Introduction", *Presence*, Vol. 7, n°2, pp. 101-107.
- Dumas, C., 1999. "Un modèle d'interaction 3D: interaction homme-machine et homme-machine-homme dans les interfaces 3D pour la TCAO synchrone". Thèse de doctorat en informatique, Université des sciences et technologies de Lille, 1999.
- Eibl-Eibesfeldt I., 1984, "Ethologie - Biologie du Comportement", *Naturalia et Biologia, 1984* (3me édition).
- Esposito C., 1996, Taxonomy of gestures proposed by Nespoulous & Lecours, (1986), in " Toward a Vision-Based Hand Gesture Interface", Francis K.H. Quek, *Proc. of VRST94*.
- Eysenck M.W. & Keane M. T., 1998, *Cognitive Psychology. A Student's Handbook*, Psychology Press, 3rd Ed. 1998, ISBN 0-86377-375-3 .
- Favalora G., Hall D., Giovinco M., Napoli J. & Dorval R., 2000, "A Multi-Megavoxel 3-D Display System for Distributed Collaboration", IEEE Globecom 2000 workshop, "*Application of Virtual Reality Technologies for Future Telecommunication System*".
- Feiner S., 2002, Invited Speaker for VRIC'02 Session 3, "Mixed Reality", june 2002, the 19th, Laval, France.
- Ferber J., 1995, "Les Systèmes Multi-Agents, vers une intelligence collective", InterEditions, Paris, 1995 (ISBN 2-7296-0665-3).

-
- Forsberg A., Herndon K. & Zeleznik R., 1996, "Aperture based selection for immersive virtual environment", *Proc. of UIST'96*, pp. 95-96 (1996).
- Foxlin E., Harrington M., & Pfeifer G., 1998, "Constellation™: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications", *Proc. of SIGGRAPH 98*, pp. 371 - 378.
- Freudenberg B., Masuch M. & Strothotte T., 2001, "Walk-Through Illustrations: Frame-Coherent Pen-and-Ink Style in a Game Engine", *Proc. of EUROGRAPHICS 2001 / A. Chalmers and T.-M. Rhyne, Guest Editors, Volume 20 (2001), Number 3*.
- Fukatsu S., Kitamura Y., Masaki T. & Kishino F., 1998, "Intuitive control of "bird's eye" overview images for navigation in an enormous virtual environment", *Proc. of VRST98*, pp. 67 - 76.
- Gabbard, J., L., and Hix, D., 1997, "A Taxonomy of Usability Characteristics in Virtual Environments", *Deliverable to Office of Naval Research*, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, November 1997.
- Gao S., Wan H. & Peng Q., 2000, "An approach to solid modeling in a semi-immersive virtual environment", *Computers & Graphics 24 (2000)*, pp.191-202.
- Ghidary S., Nakata Y., Takamori G. T. & Hattori M., 2000, "Human Detection and Localization at Indoor Environment by Home Robot", *Proc. of IEEE SMC Conference, 2000*.
- Gooch A., 1998, "Interactive Non-Photorealistic Technical Illustration", *MSc Thesis*, University of Utah, Dept. of Computer Science, 1998.
- Gooch A. A. & Willemsen P., 2002, "Evaluating Space Perception in NPR Immersive Environments", *Proc. of NPAR'02*, pp. 105-110.
- Gourdol, A.P.J.; Nigay, L.M.B.; Salber, D. & Coutaz, J. , 1992, "Multimodal systems: aspects of events fusion and a taxonomy", *IFIP Transactions A, Computer Science and Technology*, (Netherlands), vol.A-12 p. 156-62.
- Gross M. & Staadt O., 2001, "The blue-c Project", *ERCIM News No. 44*, (January 2001), pp.41-42.
- Hamard J., 2000, "Inspection ergonomique d'un logiciel de CAO destiné aux stylistes de l'industrie du textile", *Rapport de recherche*, Inria, Rocquencourt.
- Hinckley K., Pausch R., Goble J. C. & Kassel N. F., 1994, "A Survey Of Design Issues In Spatial Input. ", *Proc. of UIST '94*, pages 213-222, November 1994.
- Hirose M. (The University of Tokyo) , 2000, "Immersive Projection Technology and Wearable Computers", *IMAGINA 2000*, Session "3-D and interactive techniques. When Reality meets Virtuality".

-
- Iwata H., 1990, "Artificial reality with force-feedback: development of desktop virtual space with compact master manipulator", *Proc. of SigGraph90*, pp. 165-170.
- Iwata H., Yano H., Nakaizumi F. & Kawamura R., 2001, "Project FEELEX: adding haptic surface to graphics ", *Proc. of the 2001 conference on Computer Graphics, SIGGRAPH 01*, pp. 469 - 476.
- Jacob R. J. K., 1991, "The use of eye movements in human-computer interaction techniques what you look at is what you get", *ACM Transactions on Information Systems*, Vol. 9 , Issue 2 (1991), Pages 152-169.
- Just, M. A. & Carpenter, P. A. , 1976, "Eye fixations and cognitive processes", *Cognitive Psychology*, 8 (1976), 441-480.
- Kaur K., 1998, "Designing Virtual Environment for Usability", *PhD Thesis*, London, City University, Center for Human-Computer Interface Design, 1998. Appendix 6A, p.84: "*Objects in the vicinity are parts of the VE seen to individually possess functionality and meaning, from the user's point of view.*"
- Kitamura Y., Konishi T., Yamamoto S. & Kishino F., 2001, "Interactive stereoscopic display for three or more users", *Proc. of the 2001 conference on Computer Graphics, SIGGRAPH 01*, pp. 231-240.
- Koller, Mine & Hudson, 1996, "Head-Trackled Orbital Viewing: An Interaction Technique for Immersive Virtual Environments", *Proc. of the UIST'96 Symposium on User Interface Software and Technology (UIST)*, 1996, pp. 81-82.
- Krueger M., 1993, "Environmental technology: making the real world virtual", *Communications of the ACM*, vol. 36, No. 7 (Jul. 1993), Pages 36 - 37.
- Larson J., Bass L., Chatty S., Gulliksen J., Holsti N., Krishnamurthy B., Lefebvre P., Nigay L., Stiegler H. & Ukelson J. , 1992, "Critical Issues: Multimedia and Multimodality", *IFIP Transactions A, Computer Science and Technology*, (Netherlands), vol.A-18 pp. 411-13.
- LaViola J., Feliz D., Keefe D., Zeleznik R., 2001, "Hands-free multi-scale navigation in virtual environments". *Proc. of the Symposium on Interactive 3D graphics 2001*, pp. 9–15.
- López de Ipiña D. & Lo, 2001, "Sentient Computing for Everyone", *DAIS'2001*.
- Lourdeaux D., 2001, "Réalité Virtuelle et Formation : conception d'Environnements Virtuels Pédagogiques", *Thèse "Informatique Temps Réel, Robotique et Automatique"*, Ecole des Mines de Paris, octobre 2001.
- Mann S., 2002, "Mediated Reality with implementations for everyday life", *Presence*, Vol. 2, Issue 5, oct. 2002.
- Marrin Nakra T., 2000, "Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture", *Ph.D. Thesis*, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology, (feb. 2000).
-

-
- Masliah M. R. & Milgram P., 2000, "Measuring the allocation of control in a 6 degree-of-freedom docking experiment", *Proc. of the CHI 2000 conference*, pp. 25-32.
- Mason A. H., Walji M. A., Lee E. J. & MacKenzie C. L., 2001, "Reaching movements to augmented and graphic objects in virtual environments", *Proc. of CHI 2001*, pp. 426 - 433.
- McKenna M., 1992, "Interactive Viewpoint Control and Three-Dimensional Operations", *Proc. of SI3D92 Symposium on Interactive 3D Graphics*, p.53.
- Mellet d'Huart D., 2001, "Former au-delà de la réalité, quand l'abstrait devient réel", *Proc. of VRIC'2001, Virtual Reality International Conference*, Laval, France, 2001.
- Milgram P., Takemura H., Utsumi A. & Kishino F., 1994, "Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum", *SPIE94 Vol. 2351, Telemanipulator and Telepresence Technologies*.
- Milgram P. & Kishino, F., 1994, "A Taxonomy of Mixed Reality Visual Displays", *IEICE Transactions on Information and Systems*, vol.E77-D, no.12 p. 1321-9, Dec. 1994.
- Miller, G., 1956, "The Magic Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information", *Psychological Review*, 63, 81-97.
- Mine M. R., Brooks F. P. & Séquin C.H., 1997, "Moving Objects in Space: Exploiting Proprioception In Virtual-Environment Interaction", *SIGGRAPH 97 Conference Proceedings CD-ROM*.
- Mulder A., 1996, "Hand Gestures for HCI. Hand Centered Studies of Human Movement Project", *Technical Report 96-1*, Research on human movement behaviour reviewed in the context of hand centred input, Axel Mulder, School of Kinesiology, Simon Fraser University, February 1996.
- Mulder J. & van Liere R., 2002, "The Personal Space Station: Bringing Interaction Within Reach", *proc. of VRIC2002, Virtual Reality International Conference*, Laval, France, 2002.
- Murakami T. & Nakajima N., 2000, "DO-IT: deformable object as input tool for 3-D geometric operation", *Computer-Aided Design*, vol. 32 (2000) p. 5-16.
- Murphy H. & Duchowski A. T., 2001, "Gaze-Contingent Level Of Detail Rendering", in *EuroGraphics 2001 (Short Presentations)*, September 4-7, 2001, Manchester, UK, EuroGraphics. <http://www.vr.clemson.edu/~andrewd/home.html#pubs>
- Mynatt E., Blattner D., Blattner M. M., Macintyre B. & Mankoff J., 1998, "Augmenting Home and Office Environments", *SIGCAPH 98 / p169*.
- Nespoulous, J.L & Lecours, A., 1986, "Gestures: nature and function". In: Nespoulous, J.L., Perron, P., Roch Lecours, A., *The biological foundations of gestures: motor*

and semiotic aspects, p 49-62, Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates.

- Nigay L., 1994, "Conception et modélisation logicielles des systèmes interactifs: application aux interfaces multimodales.", *Thèse d'Informatique de l'Université Joseph Fourier - Grenoble 1*, janvier 1994.
- Nigay L. & Coutaz J., 1993, "A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion", *Proc. of the ACM INTERCHI'93 Conference on Human Factors in Computing Systems 1993* p.172-178.
- Nishino H., Utsumiya K., Kuraoka D. & Yoshioka K., 1997, "Interactive Two-Handed Gesture Interface in 3D Virtual Environments". *Proc. of VRST'97*, pp. 1-8.
- Nishino H., Utsumiya K. & Korida K., 1998, "3D Object Modeling using Spatial and Pictographic Gestures". *Proc. of VRST'98*, pp. 51-58.
- Orr R. J. & Abowd G. D., 2000, "The Smart Floor: A Mechanism for Natural User Identification and Tracking", *GVU Technical Report GIT-GVU-00-02* (full paper). January 2000.
- Ohshima T., Yamamoto H. & Tamura H., 1996, "Gaze-Directed Adaptive Rendering for Interacting with Virtual Space", In *Proceedings of VRAIS'96*, 103–110. IEEE.
- Papin J.P., 2001, Propos recueillis lors de la conférence Virtual Reality International Conference (VRIC 2001) à Laval.
- Paradiso J. A. & Gershenfeld N., 1996, "Musical Applications of Electric Field Sensing", April 1996; *Computer Music Journal*.
- Paradiso J. A. & Sparacino F., 1997, "Optical Tracking for Music and Dance Performance", Media Laboratory, Massachusetts Institute of Technology, Presented at the *Fourth Conference on Optical 3D Measurement Techniques*, ETH, Zurich, September 1997.
- Patrick E., Cosgrove D., Slavkovic A., Rode J. A., Verratti T. & Chiselko G., 2000, "Using a large projection screen as an alternative to head-mounted displays for virtual environments", *Proc. of CHI 2000*, Pages 478 - 485.
- PBVE2000, 2000, "Environnements Virtuels à Base de Projections sur Grands Ecrans", *Séminaire INRIA PBVE2000*, novembre 2000.
- Perlin, K., 2004. "Emotive Virtual Actors", Imagina 2004, session "Quels jeux vidéo pour demain?". Applet 3D à l'adresse <http://mrl.nyu.edu/~perlin/experiments/emotive-actors/>.
- Peterson B., Wells M., Furness III T. A. & Hunt E., 1998, "The effect of the interface on Navigation in Virtual Environments", *Human Factors and Ergonomics Society 1998 Annual Meeting*.

-
- Pierce J., Forsberg A., Conway M., Hong S., Zeleznik R. & Mine M., 1997, "Image plane interaction techniques in 3D immersive environments", *Proc. of SI3D Graphics'97*, pp. 39-43 (1997).
- Pingali G., Tunali G. & Carlbom I., 1999, "Audio-Visual Tracking for Natural Interactivity", *ACM Multimedia '99*, pp. 373-382.
- Poupyrev I., Billinghurst M., Weghorst S. & Ichikawa T., 1996, "The go-go interaction technique: non-linear mapping for direct manipulation in VR", *Proc. of UIST'96*, pp. 79-80 (1996)
- Poupyrev I., Weghorst S., Billinghurst M. & Ichikawa T., 1998, "Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques", *EUROGRAPHICS' 98 / N. Ferreira and M. Göbel Volume 17, (1998), Number 3.*
- Preece J., 1994, "Human-Computer Interaction", *Addison Wesley*, 1998, ISBN 0-201-62769-8.
- Pryor H. L., Furness III T. A. & Viirre E., 1998, "The virtual Retinal Display: a new display technology using scanned LASER light", *HFES 98, Proceedings of the Human Factors and Ergonomics Society*, 1998, Vol.2.
- Quek F. K.H., 1994, "Toward a Vision-Based Hand Gesture Interface", *Proc. of VRST 1994*, May 11,1994.
- Raskar R., Welch G., Cutts M., Lake A., Stesin L. & Fuchs H., 1998, "The office of the future: a unified approach to image-based modeling and spatially immersive displays.", *Proc. of Siggraph'98*, pp. 179-188, 1998.
- Robbe, S., Carbonell N., Dauchy P., 2000, "Expression constraints in multimodal human-computer interaction", in H. Lieberman (Ed.), *Proc. of the 2000 International Conference on Intelligent User Interfaces (IUI'2000)*, New Orleans, January 9-12 2000, New York: ACM Press, pp. 225-229, 2000.
- Robinett W. & Holloway R., 1992, "Implementation of flying, scaling and grabbing in virtual worlds", *Proc. of the 1992 symposium on Interactive 3D graphics, SI3D 1992*, pp. 189 - 192.
- Rolland J. & Fuchs H., 2001, "Optical versus Video See-Through Head-Mounted Displays", Jannick P. Rolland & Henry Fuchs, in "*Fundamentals of Wearable Computers and Augmented Reality*", Ed. Barfield & Caudell, Lawrence Erlbaum Associates, Publishers, 2001, ISBN 0-8058-2901-6, pp.150
- Rosenfeld R., Olsen D. & Rudnicky A., 2001, "Universal Speech Interfaces", *Interactions*, nov.-dec. 2001.
- Salvucci D. D., 1999, "Inferring intent in eye-based interfaces: tracing eye movements with process", *Proc. of CHI 99*, Pages 254-261.

-
- Salvucci D. D. & Anderson J. R., 2000, "Intelligent gaze-added interfaces", *Proc. of CHI 2000*, Page 273.
- Sato M., 2001, "Spider", présentation sur le site de la conférence *VRIC2001*, Laval.
- Scapin, D., 1981. "Computer commands in restricted natural language: some aspects of memory and experience". *Human Factors*, 1981, 23(3), pp. 365-375.
- Scapin, D., 1982. "Generation effect, structuring and computer commands". *Behaviour and information technology*, 1982, vol.1, no.4, pp. 401-410.
- Schkolne S., Pruett M. & Schröder P., 2001, "Surface drawing: creating organic 3D shapes with the hand and tangible tools", *Proc. of CHI 2001*, pp. 261 - 268.
- Schumann J., Strothotte T., Raab A. & Laser S., 1996, "Assessing the Effect of Non-Photorealistic Rendered Images in CAD", *Proc. of CHI'96*.
- Searle J. R., 1969, "Speechs Acts". Cambridge University Press. ("Les actes de Langage").
- Searle J. R. ,1979, "Expression and Meaning". Cambridge University Press.
- Seminaire INRIA-Rocquencourt, nov. 2000, Geoclay Demo, *Séminaire sur les "Environnements virtuels à base de projections sur grands écrans"*, INRIA-Rocquencourt, novembre 2000.
- Sheehan N., 1995, "Physiology of Behavior". Carlson. Sixth ed. Allyn and Beacon. 1995. ISBN 0-205-27340-8. Ch.6, Vision. pp. 148-184. "Texture cues. Variations in texture can produce an appearance of distance".
- Shepard R. & Cooper L., 1986, "Mental Images and their Transformations". MIT Press, 1986, ISBN 0-262-69099-4
- Sheridan, T.B., 1996, "Further Musings on the Psychophysics of Presence", *Presence* 5(2): 241-246.
- Shilling RD & Shinn-Cunningham B. G., 2001, "Virtual auditory displays", in *Handbook of Virtual Environments Technology*, K. Stanney (ed.).
- Shinn-Cunningham B. G., 2001, "Spatial auditory displays", in *International Encyclopedia of Ergonomics and Human Factors*, W. Karwowski (ed.). London: Taylor and Francis, Ltd.
- Shneiderman B., 1998, "Designing the user interface" Addison Wesley, third Edition. ISBN 0-201-69497-2.
- Sibert L. E. & Jacob R. J. K., 2000, "Evaluation of eye gaze interaction ", *Proc. of CHI 2000*, pp. 281 - 288 .
- Slater M. & Usoh M., 1994 , "Body Centred Interaction in Immersive Virtual Environments", in N. Magnenat Thalmann and D. Thalmann (eds.), *Artificial Life and Virtual Reality*, John Wiley and Sons, 125-148.

-
- Slater M., 1997, "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments", *Presence* 6(6):603–616.
- Smith J. R., 1996, "Field mice: Extracting hand geometry from electric field measurements", *IBM Systems Journal*, Vol 35, Nos 3&4, 1996.
- Starker I. & Bolt R. A., 1990, "A gaze-responsive self-disclosing display", *Proc. of CHI 90*, pp. 3 - 10.
- Starner T., Leibe B., Singletary B., Lyons K., Gandy M. & Pair J., 2000a, "Guided by Voices: an audio-only augmented reality", in " Towards Augmented Reality Gaming", *Imagina 2000*. 2nde Conférence.
- Starner, Leibe, Singletary, Lyons, Gandy & Pair, 2000b, "Towards Augmented Reality Gaming", *Imagina 2000*.
- Stoakley R., Conway M. J. & Pausch R., 1995, "Virtual reality on a WIM: interactive worlds in miniature ", *Proc. of CHI'95*, Pages 265 - 272 (1995)
- Strippgen S. & Peters K., 1999, "The other way round!: collaborative communication with agents", *Proc. of Agents 1999*, Pages 108 - 115.
- Szeliski R. & Tonnesen D., 1991, "Surface Modeling with Oriented Particles", *CRL 91/14*, Digital Equipment Corporation, Cambridge Research Lab, Cambridge, Massachusetts, Dec.,1991.
- Szeliski R. & Tonnesen D., 1992, "Surface modeling with oriented particle systems", *Proc. of SIGGRAPH'92*, pp. 185-194.
- Taillefer R., Scapin D.L., 2000, "Modeling Designer's Activities in the Garment Industry.", *ERCIM News*, volume 46, 2001.
- Taillefer R., 2001a, "Analyse du travail des concepteurs (stylistes-modélistes) dans les industries du textile.", *Rapport de contrat COMEDIA*, INRIA, Rocquencourt, 2001.
- Taillefer R., 2001b, "Ergonomie de la conception graphique dans les industries du textile. Revue de littérature. ", *Rapport de contrat COMEDIA*, INRIA, Rocquencourt.
- Tamura H., 2000, (Mixed Reality Systems Laboratory), "Real-time Interaction in Mixed Reality Space: Entertaining Real and Virtual Worlds", *IMAGINA 2000*, Session "3-D and interactive techniques. When Reality meets Virtuality".
- Tan D. S., Robertson G. G. & Czerwinski M., 2001, "Exploring 3D navigation: combining speed-coupled flying with orbiting", *Proc. of CHI 2001*, p. 418.
- Tanriverdi V. & Jacob R. J. K., 2000, "Interacting with eye movements in virtual environments", *proc. of CHI 2000*, pp. 265-272.

-
- Thórisson, K. R., 1995, "Computational Characteristics of Multimodal Dialogue.", *AAAI Fall Symposium on Embodied Language and Action*, Massachusetts Institute of Technology, Cambridge, Massachusetts, November 10-12, 102-108.
- Thórisson, K. R., 1997, "Layered Modular Action Control for Communicative Humanoids", *Proc. of Computer Animation '97*, Geneva, Switzerland, June 5-6, 134-143.
- Tisseau J., 2001, "Réalité Virtuelle — autonomie in virtuo —", *Habilitation à Diriger des Recherches, Spécialité : Informatique*, Université de Rennes 1, décembre 2001.
- Tonnesen D., 1998, "Dynamically Coupled Particle Systems for Geometric Modeling, Reconstruction, and Animation", *PhD Thesis*, University of Toronto.
- Tsingos N., Funkhouser T., Ngan A. & Carlbom I., 2001, "Modeling acoustics in virtual environments using the uniform theory of diffraction", *Proc. of the 2001 conference on Computer Graphics*, SIGGRAPH 01, pp. 545-552.
- Van Ballegooij A. & Eliéns A., 2001, "Navigation by query in virtual worlds". *Proc. of VRML2001*. Pages: 77–83.
- Vanderveken B., 1988, "Les actes de discours". Pierre Mardaga.
- Vlahakis V., Karigiannis J., Tsotros M., Gounaris M., Ioannidis N. & Stricker D., 2002, "Interactive Augmented Reality Touring of Archaeological sites", (projet Européen "ArcheoGuide"), *VRIC 2002*, Laval.
- Ware C., Arthur K. & Booth K. S., 1993, "Fish tank virtual reality", *Proc. of InterCHI 93*, pp. 37-42.
- Ware, C; Gobrecht, C; Paton, M A, 1998, "Dynamic Adjustment of Stereo Display Parameters", *IEEE Transactions on Systems Man and Cybernetics - Part A - Systems and Humans*, Vol. 28 (1998), No. 1, p. 56-65.
- Ware C. & Mikaelian H. H., 1987, "An evaluation of an eye tracker as a device for computer input2", *Proc. of CHI/GI 87*, Pages 183-188.
- Wexler, M., Kosslyn, S., Berthoz, A., 1998. "Motor processes in mental rotation", *Cognition*, 68 (1998) 77-94.
- Williams S.P. & Parrish R.V., 1990, "New Computational Control Techniques and increased understanding for stereo 3-D displays", in *Proc. of SPIE Stereoscopic Displays Applications*, Santa Clara, CA, 1990, pp. 300-304.
- Zhai S., Buxton B. & Milgram P., 1994, "The 'Silk Cursor': investigating transparency for 3D target acquisition", *Proc. of CHI94*, Pages 459 - 464.
- Zhai S., Milgram P. & Buxton B., 1996, " The FingerBall", in " The Influence of Muscle Groups on Performance of Multiple Degree-of-Freedom Input", *Proc. of CHI96*.

-
- Zhai S., Morimoto C. & Ihde S., 1999, "Manual and gaze input cascaded (MAGIC) pointing ", *Proc. of CHI 99*, pp. 246-253.
- Zimmerman T. G., Lanier J., Blanchard C., Bryson S. & Harvill Y., 1987, "A hand gesture interface device", *CHI/GI 1987 conference proceedings on Human factors in computing systems and graphics interface*, pp. 189-192.
- Zimmerman T. G., Smith J. R., Paradiso J. A., Allport D. & Gershenfeld N., 1995, "Applying Electric Field Sensing to Human-Computer Interfaces", *Proc. of CHI'95 Human Factors in Computing Systems*, Denver, May 9-11, 1995, ACM Press, New York



FACULTE DES SCIENCES & TECHNIQUES

**U.F.R. Sciences et Techniques Mathématiques, Informatique et Automatique
Ecole Doctorale IAEM LORRAINE
Département de Formation Doctorale en Informatique**

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Informatique

par **Charles H. A. TISON**

**Mise en œuvre et évaluation
d'interactions multimodales orientées communication
appliquées au déplacement d'objets graphiques 3D**

ANNEXES

Annexes

ANNEXES	1
1 CONSIGNE AUX PARTICIPANTS	3
2 QUESTIONNAIRE DE PARTICIPATION	15
3 PLATEFORME D'EXPERIMENTATION " SIO3D "	21
3.1 MACHINE GRAPHIQUE DE SORTIE.....	22
3.2 EQUIPEMENT DE L'UTILISATEUR EN ENTREE.	23
3.3 EQUIPEMENT DU " MAGICIEN "	25
3.4 RECEPTEUR DE COMMANDE.	26
3.5 MACHINE D'ENTREE.....	29
4 DETECTION DES "POSES" MANUELLES	31
4.1 PRINCIPE.....	32
4.2 AMV MULTIVIBRATEUR ASTABLE.....	34
5 TRANSCRIPTION DES PRODUCTIONS DES UTILISATEURS ERREUR ! SIGNET NON DEFINI.	
6 FOBS CONFIGURATION DES UNITÉS “FLOCK OF BIRDS” ...ERREUR ! SIGNET NON DEFINI.	
6.1 CONFIGURATION INITIALE DES « FLOCK OF BIRDS »	ERREUR ! SIGNET NON DEFINI.
6.2 CONFIGURATION D'ORIGINE DE LA PREMIERE UNITE	ERREUR ! SIGNET NON DEFINI.
6.3 CONFIGURATION D'ORIGINE DE LA SECONDE UNITE	ERREUR ! SIGNET NON DEFINI.
6.4 CONFIGURATION D'UNE UNITE « BIRD » EN « MASTER »	ERREUR ! SIGNET NON DEFINI.
6.5 CONFIGURATION D'UNE UNITE « BIRD SLAVE » EN FIN DE BUS.....	ERREUR ! SIGNET NON DEFINI.

1 Consigne aux participants

Consigne aux participants :

Introduction à l'expérimentation MERLIn sur SIO3D (Système d'Interaction avec des objets 3D):

Déplacement et positionnement d'objets 3D Virtuels.

Vous allez expérimenter un nouveau style d'interaction avec des objets 3D en participant à un test utilisateur.

Ce test est destiné à évaluer une application informatique, ce n'est pas vous qui êtes évalué.

Votre but est d'obtenir, sur une scène virtuelle, des mouvements sur des objets en exprimant une commande que le système devra interpréter.

Certaines modifications que vous devrez faire pourront vous paraître difficiles à mettre en oeuvre. Vous ne devrez pas dans ce cas vous sentir mis en cause à ce sujet, car il est possible que l'interaction soit mal conçue en ne permettant pas d'exprimer facilement la commande correspondante.

Nous recueillerons vos commentaires et suggestions d'améliorations de notre système en fin d'expérience.

Imaginez votre situation **comme si vous demandiez à quelqu'un** situé sur la scène d'y effectuer les modifications souhaitées, comme, par exemple, le déplacement d'un objet. La principale différence réside dans le fait que ce quelqu'un est le système informatique lui-même qui vous écoute et perçoit certains de vos mouvements.

Affichages

Vous disposez de deux écrans, la *scène virtuelle* et l'*écran d'instruction*.

La *scène virtuelle*: c'est le grand écran qui vous fait face et qui présente une scène virtuelle affichant *six objets colorés sélectionnables*.

Ces objets sont situés à diverses distances et dans des directions différentes. Ils sont constitués de l'empilement de plusieurs cubes de 1 mètre de côté, (les objets les plus hauts font donc 3m de hauteur).



Le **plancher** est constitué de dalles de 2m de côté, les cubes de pierre marbrés qui servent de repères font 1m de hauteur et sont inamovibles.

La bulle qui contient une main en bas à droite représente l'aspect de votre main droite, mais vous n'aurez pas besoin d'en tenir compte.

La position virtuelle que vous occupez par rapport à cette scène est fixe pour cette expérience, vous resterez donc assis sur la chaise que vous occupez.

L'**écran d'instruction** vous sera décrit plus loin.

Equipement d'interaction

Voyons maintenant les moyens dont vous disposez pour vous adresser au système.
- le système est en mesure de percevoir les mouvements de votre tête (1) pour en déduire une direction de visée sur la scène.

- il peut également percevoir les mouvements de votre main droite (2) ainsi que certaines formes particulières de la main pour indiquer par exemple une direction.
- enfin, le système perçoit les commandes vocales (3) que vous lui adresserez par l'intermédiaire du microphone que vous portez.

(1). Utilisation de la tête

Vous disposez sur votre tête d'un capteur directionnel qui permet au système de connaître l'orientation de votre tête.

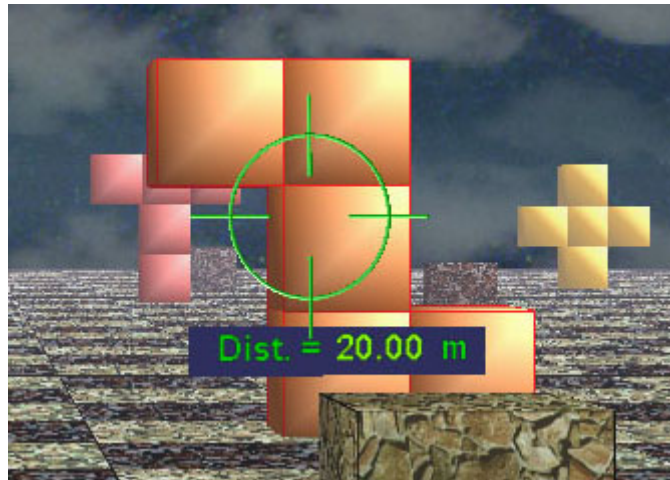
C'est à l'aide de cette information que le système vous présente un viseur dont vous pouvez piloter la position virtuelle en tournant la tête.

La première fonction du viseur est une **indication de direction** à partir de l'endroit où vous êtes par rapport à la scène.

La seconde fonction importante de ce viseur est la **présélection de l'un des objets** colorés de la scène.

Si vous superposez ce viseur sur l'un des objets sélectionnables de la scène, trois modifications se produisent:

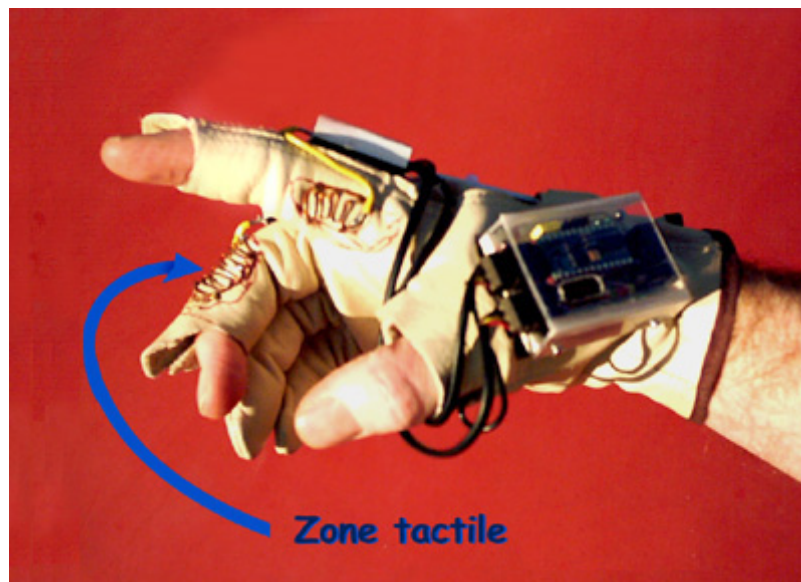
1. Le viseur change de couleur. Il devient vert pour signaler qu'une sélection d'objet est possible.
2. L'objet de la visée est représenté dans une boîte englobante rouge représentant les bordures de l'objet.
3. La distance vous séparant de l'objet de la visée est affichée en vert sur un bandeau noir sous le viseur.



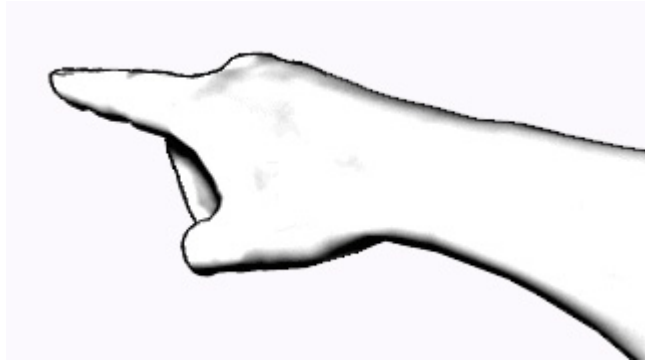
Test: A titre d'exercice, regardez en direction de la *scène virtuelle* en bougeant la tête à droite et à gauche, positionnez le viseur sur un objet coloré de la scène et observez ce qui se produit.

(2). Utilisation du Gant

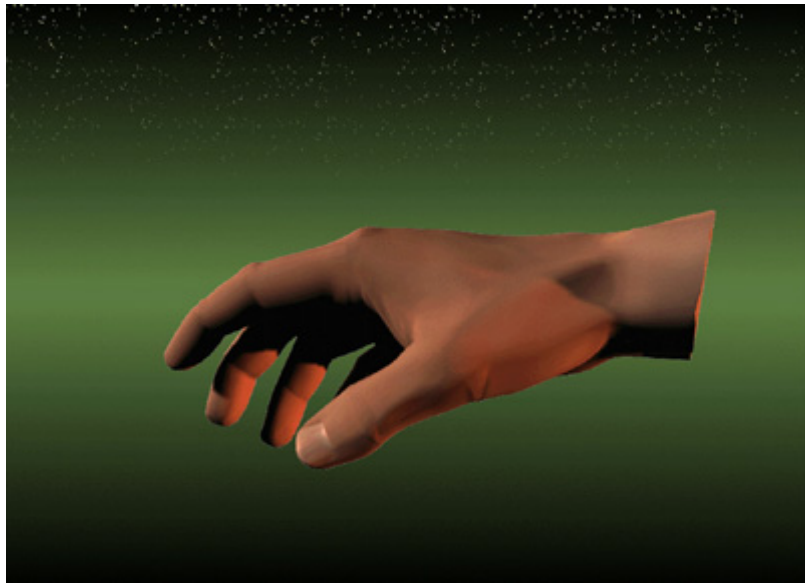
Examinons maintenant la fonction du gant que vous portez sur votre main droite. Vous disposez, sur le dos de ce gant, d'un capteur de position et d'orientation qui permet au système de connaître la position de votre main ainsi que son orientation dans l'espace. Ce gant possède par ailleurs deux zones tactiles dont une seule sera utilisée.



Cette zone tactile est située sur la gauche du majeur de votre main droite et vous permet, lorsque vous le touchez en appuyant avec le pouce, d'indiquer une direction avec l'index tendu tel que représenté ci-après.

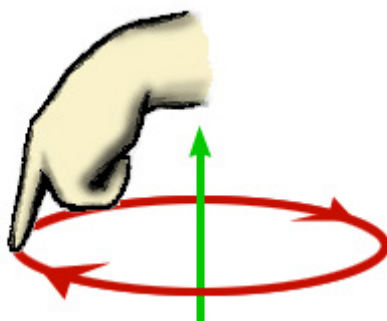


Lorsque vous ne touchez pas la zone tactile de ce capteur situé sur le côté de votre majeur, votre main droite est perçue par le système comme non active:



La première fonction du gant est l'*indication de direction manuelle*. En cela, il est complémentaire à l'indication d'une direction visée à l'aide de la tête avec le viseur. Il permet en particulier d'indiquer une direction pointant ailleurs que sur la scène.

La seconde fonction importante du gant est l'*indication d'une direction et d'un axe de rotation*. Tracer un cercle en pointant de l'index permet en effet d'indiquer à la fois un axe autour duquel effectuer une rotation ainsi que le sens de cette rotation.



Test: Afin de vérifier le bon fonctionnement du gant, montrez une direction à l'aide de la main en appuyant votre pouce comme précédemment indiqué.

Observez alors l'apparence de la main virtuelle se trouvant dans la bulle en bas à droite sur la *scène virtuelle*.

(3). Utilisation de la voix

Vous disposez d'un microphone de tête qui permet au système de capter vos énoncés vocaux. Le système pourra reconnaître certains termes du langage si vous les prononcez de manière claire et distincte (avec toutes les syllabes).

Ces termes permettent tout d'abord de spécifier au système une opération de mouvement à effectuer.

Ils permettent aussi de valider la sélection d'un objet particulier de la scène sur lequel effectuer cette opération, ainsi que celle d'une direction indiquée à l'aide d'un geste. On l'utilise enfin également pour préciser une grandeur telle qu'une distance ou un angle.

Enchaînement des interactions

Par analogie à ce que l'on demanderait à une personne présente dans la scène, l'enchaînement des actions consiste à demander:

1. L'opération qui doit être exécutée, sous forme de verbe. (Le "*Faire quoi*").
2. l'objet sur lequel l'action porte ("*Sur quoi*")
3. le détail sur l'action ("*Comment*")

exemple pour un changement d'échelle, "*Réduis cet objet* (l'un des objets colorés) *de 30 % !*". (Pas d'indication de direction dans une telle opération de réduction).

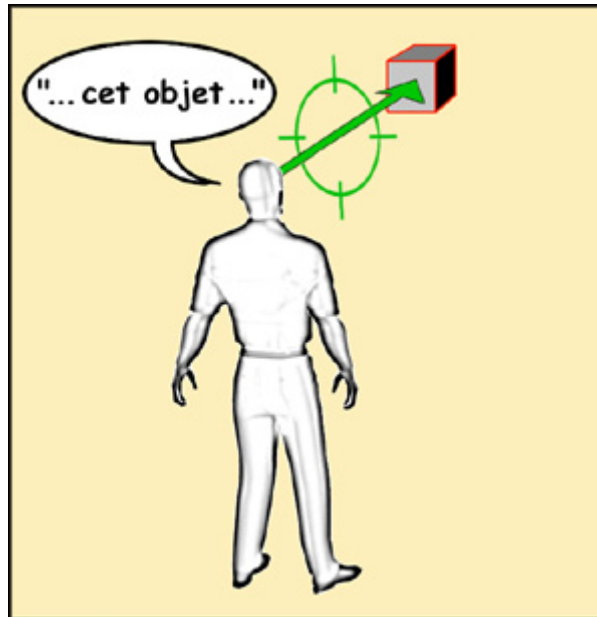
1. Quelle opération, ("Faire quoi").

Le mouvement à obtenir *s'énonce oralement* et distinctement comme un ordre *impératif* au *singulier*.

Ce qui consiste donc à utiliser un verbe traduisant le mouvement que vous voulez obtenir. Pour reprendre l'exemple de changement d'échelle, "*Réduis...*"

2. Sur quel objet, ("Sur quoi").

Pour obtenir la sélection d'un objet, il faut donc d'abord le présélectionner par visée (superposition du viseur sur l'objet pour obtenir une modification de son affichage en couleur verte et un affichage de la distance de l'objet), puis prononcer oralement, tout en conservant la présélection, "*...cet objet...*".



3. Le détail de l'action, ("Comment").

Il convient de préciser une *valeur* et une *direction*.

Exemple, "... *de 3 mètres* (valeur) *sur la gauche* !(direction)".

3.1 Exprimer une *valeur*.

Indication d'une distance de translation

Certaines opérations de mouvement impliquent de préciser une distance. Il pourra s'agir d'une distance à faire parcourir à l'objet dans une certaine direction, ou bien une distance depuis l'endroit où vous êtes, ou encore une distance à partir d'un autre objet. L'indication d'une distance s'effectue *oralement* en prononçant sa valeur et l'unité utilisée. Par exemple, "...*à 10 mètres...*", "...*de 15 mètres...*", etc.

Indication d'un angle

Deux catégories d'opérations de mouvement impliquent de préciser un angle. La première catégorie concerne les mouvements d'une certaine valeur d'angle par rapport à votre position.

Par exemple, "...*de 30 degrés...*", "...*à 10 degrés...*".

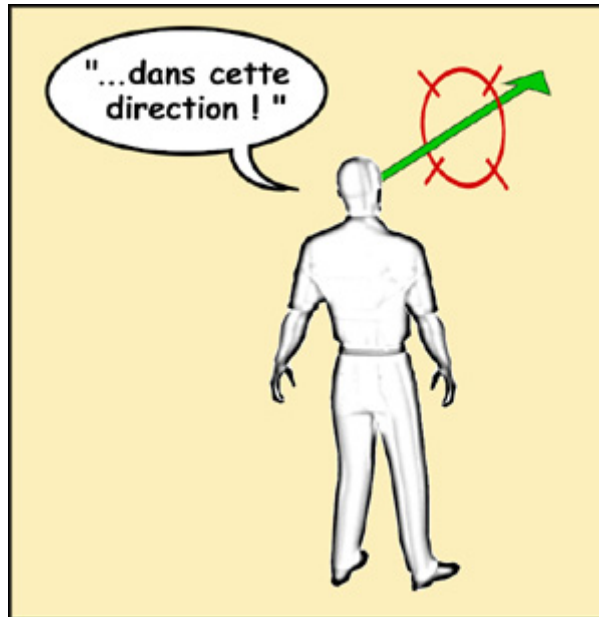
La seconde catégorie concerne les mouvements de rotation d'un objet sur place.

Par exemple, "...*de 45 degrés...*", "...*d'un demi-tour...*".

3.2 Exprimer une direction.

Indication d'une direction absolue à l'aide du viseur.

Indiquer une direction depuis l'endroit où vous êtes consiste à positionner le dispositif de visée avec la tête dans la direction de votre choix comme représenté sur la figure ci-après, puis à prononcer oralement "...*dans cette direction.*".

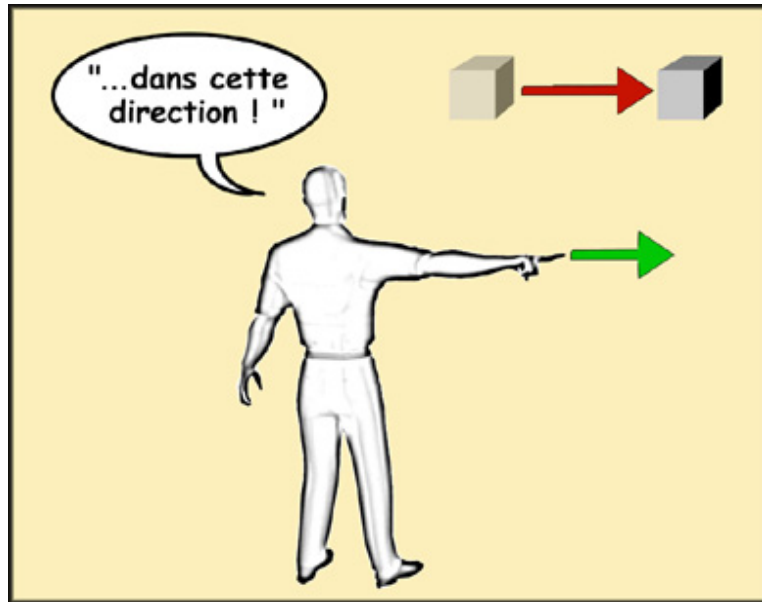


Indication manuelle d'une direction de translation

L'indication d'une direction à l'aide de la main s'effectue en pointant de l'index dans la direction voulue tout en appuyant du pouce sur la zone tactile du gant et en prononçant oralement "*...dans cette direction.*". (Voir la figure ci-dessous à gauche). Indiquer dans la même direction que celle du viseur comme figuré ci-dessous à droite en vert est équivalent à l'indication d'une direction par visée.



Toutefois, dans certains cas, indiquer une direction à l'aide du viseur ne permet pas d'exprimer une direction de translation telle que figurée ci-dessous en rouge.



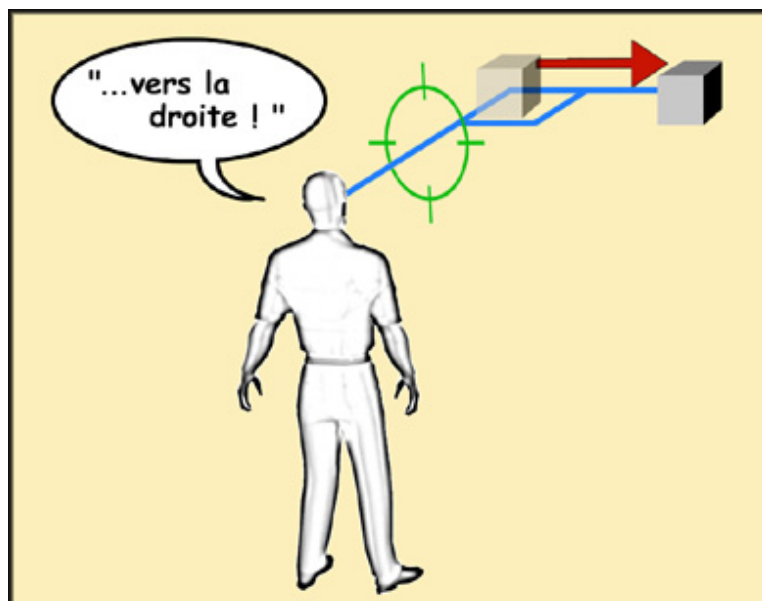
Dans ce cas, l'*indication d'une direction de translation à l'aide de la main est la seule façon de procéder*, comme ci-dessus en vert.

Dans chacun de ces cas, l'indication d'une direction voulue implique que vous prononciez oralement "*...dans cette direction.*".

Indication d'une direction latérale à angle droit

Dans le cas où vous voulez déplacer un objet latéralement à angle droit comme représenté sur la figure ci-dessous, il suffira d'*utiliser la voix seule* en prononçant les termes tels que "*...vers la droite.*", "*...vers le haut.*", etc.

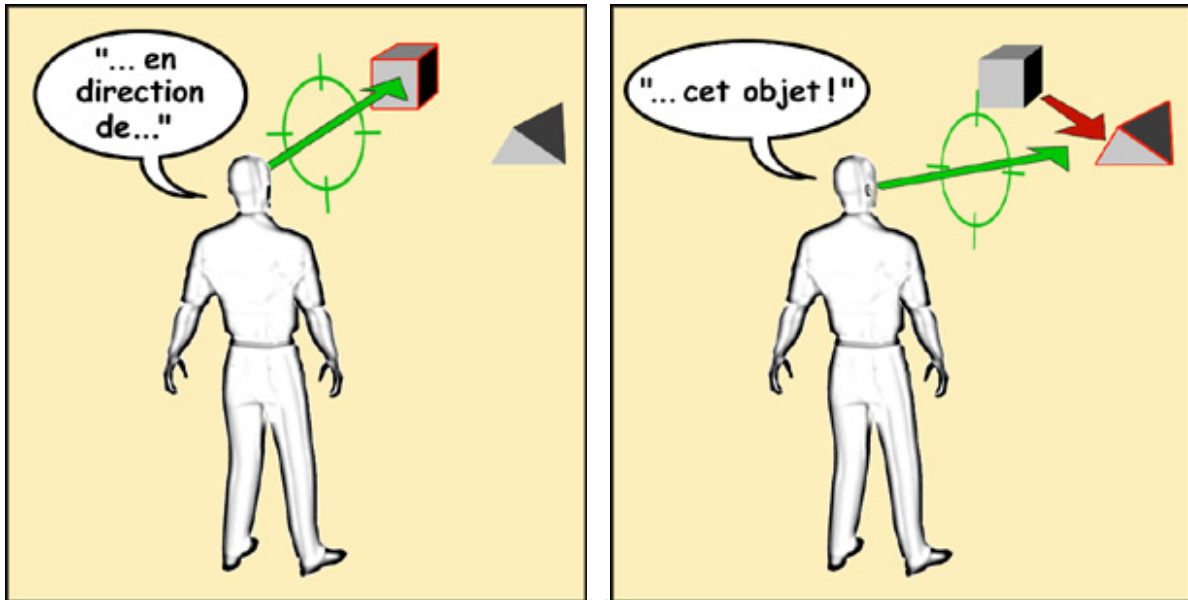
Dans ces cas particuliers, la direction du mouvement à effectuer forme un *angle droit* avec l'axe qui vous sépare de l'objet sélectionné pour le mouvement demandé. Un tel angle droit est figuré en **bleu** sur la figure suivante montrant un mouvement souhaité figuré en **rouge**.



Indication d'une direction par rapport à un second objet

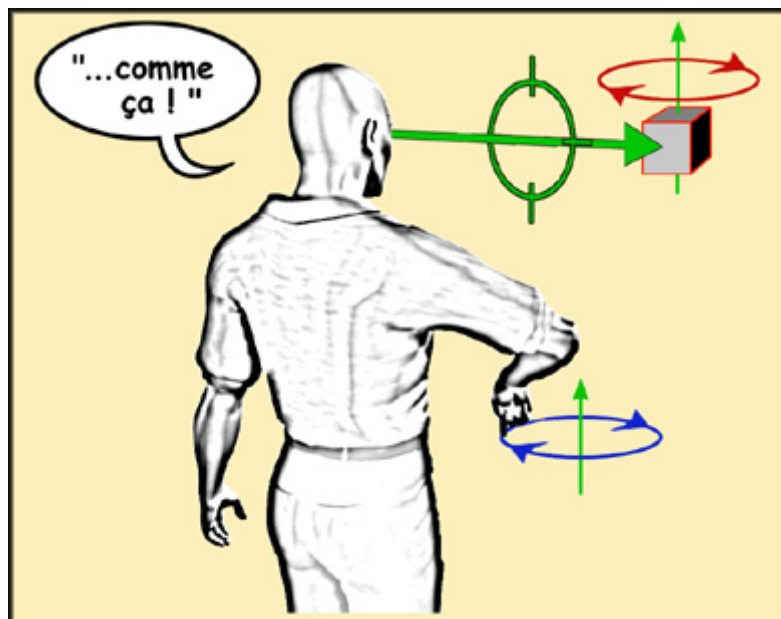
Le mouvement d'un objet sélectionné peut en effet impliquer la sélection d'un second objet en vue par exemple de rapprocher le premier du second.

L'indication d'une telle direction implique alors l'énonciation vocale du terme "*...en direction de...*", suivi d'une sélection de l'objet en direction duquel effectuer le mouvement, c'est-à-dire "*...en direction de cet objet !*" (l'un des objets colorés, comme précédemment décrit).



Indication d'une direction et d'un axe de rotation

L'indication de la façon dont sera effectué un mouvement rotatif s'indique en traçant un *cercle complet* et suffisamment ample à l'aide de l'index tendu (en appuyant du pouce sur la zone tactile du majeur). Ce faisant, et afin de signaler au système que vous lui fournissez ces informations d'axe et de sens de rotation à l'aide du gant, il est nécessaire de prononcer oralement "*...comme ça.*", ou encore "*...comme ceci.*".



En résumé

La syntaxe complète d'une expression est fixe.

Elle commence avec le *verbe* impératif traduisant le mouvement à obtenir, suivi d'une sélection de l'*objet* du mouvement, se poursuit avec une *valeur* (distance ou angle), et se termine avec une *direction*.

Opération + Objet + valeur + direction.

Dans quelques cas particuliers, la distance peut être omise si le mouvement conserve par exemple la même distance entre votre position et celle de l'objet du mouvement avant, et après le mouvement.

Dans d'autres cas, la direction indiquée peut comporter une référence à un second objet sélectionné.

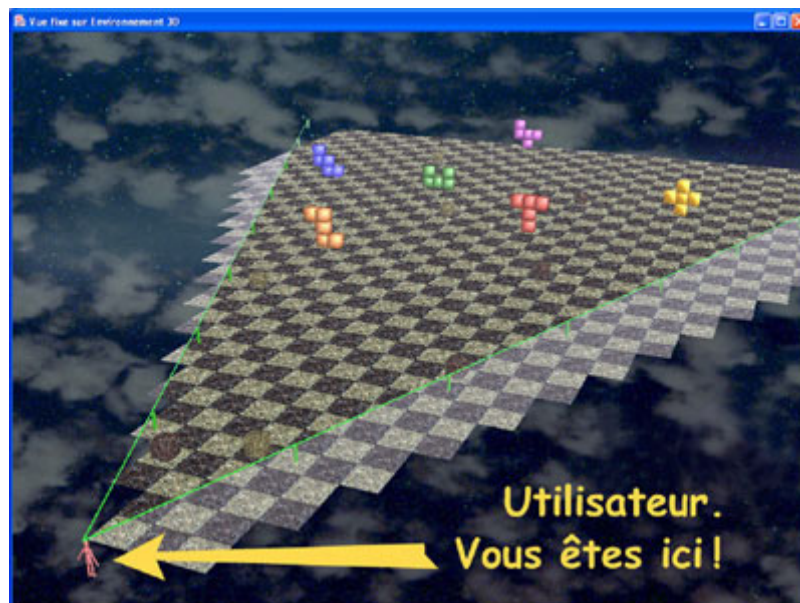
Dans ce cas, la construction de l'expression reste globalement similaire.

Quelles sont les tâches à faire accomplir au système

L'écran d'instruction

Sur votre droite se trouve un second écran. L'*écran d'instruction*.

Cet écran d'instruction vous montre initialement une vue aérienne plongeante de la scène ainsi que votre position par rapport à la scène virtuelle.



Comme le montre la figure ci-dessus, vous vous trouvez en bas à gauche face aux objets colorés.

Les deux lignes vertes y délimitent le champ de vision qui vous est présenté sur la vue principale de la scène virtuelle qui vous fait face.

C'est sur cet écran d'instruction que vous seront présentées les modifications de la scène à obtenir de la part du système.

Les figures qui seront présentées sur l'écran d'instruction comme tâches à faire accomplir au système montrent toutes la même vue aérienne plongeante agrémentée d'informations colorées.

Les informations en **rouge** figurent le **mouvement** de l'objet à obtenir.

Les informations en **bleu** figurent des **directions** qu'il vous appartiendra d'exprimer en conséquence.

Les informations en **jaune** représentent les **valeurs de dimensions** telles que **distances** ou **angles** ainsi que l'**unité utilisée**.

Chaque présentation d'une nouvelle tâche vous sera signalée par un message vocal du système. "*Affichage d'une tâche sur l'écran d'instruction!*".

Le système vous demandera alors de prendre le temps d'examiner la situation exposée. "*Veillez observer le mouvement à obtenir!*".

Il vous demandera également de signaler lorsque vous êtes prêt à formuler votre expression de commande. "*Lorsque vous êtes prêt, dites - Je suis prêt! - . Vous pourrez ensuite formuler votre expression après le bip sonore!*".

Le bip sonore dont il est question ci dessus vous signale que le système est prêt à recevoir une commande.

Après chaque commande reconnue donnant lieu à l'exécution de la tâche proposée, le système réinitialisera la scène en le signalant par un message vocal. ("*Réinitialisation de la scène!*").

Pour terminer cette présentation, si vous avez des questions à poser sur quoi que ce soit que vous n'avez pas bien compris, n'hésitez pas à le faire avant de commencer l'expérience.

2 Questionnaire de Participation

**Questionnaire de Participation
à l'expérimentation SIO3D.
Eté 2004.**

En vertu de la loi du 06 janvier 1978 portant sur la liberté individuelle, les informations portées sur le présent document sont strictement confidentielles. Les parties s'engagent à respecter cette confidentialité en acceptant leur exploitation excluant toute mention nominative permettant l'identification du participant à l'expérience dont relève ce document.

Le participant reconnaît avoir explicitement accepté librement et gracieusement sa participation à l'expérience à laquelle il a pris part.

Il reconnaît également n'avoir subi aucun traitement dérogeant aux règles déontologiques, éthiques ou contraire aux Droits de l'Homme et du Citoyen.

Le participant reconnaît par ailleurs avoir reçu, ce jour, au sujet de l'expérience à laquelle il a participé, tout complément d'information satisfaisant aux règles précitées.

Par ailleurs, le participant autorise, par le présent document, l'Institut National de Recherche en Informatique et en Automatique ainsi que l'expérimentateur à effectuer puis exploiter des enregistrements audio, vidéo et digitaux de sa personne dans le cadre de l'expérience.

Etabli le

Identité du participant:

Identifiant document:

" SIO3D-
04-.....

Nom:

Prénom:

Numéro d'ordre du participant dans la catégorie:

0.....

Sexe:

Homme/Femme.

Date de Participation : (*aa mm jj*, ex. 04 06 31)

Année de naissance:

activité:

(Initiales [Nom + Prénoms] + deux derniers chiffres de l'année de naissance)

Téléphone ou adresse e-mail:

Identité de l'expérimentateur:

Signature du participant: (*en fin d'expérience*)

Dominique SCAPIN, projet MerLin, INRIA.

(Précédé de la mention "*Lu, compris et approuvé*").

Charles TISON, projet MerLin, INRIA.

Signature du Responsable de l'Expérience:

Clause de confidentialité ?

Avez-vous des problèmes de vue et/ou d'audition, même après correction? Oui / Non
 Si oui, veuillez préciser:

Notions d'informatique

Utilisation régulière d'un système informatique.

Niveau de familiarité	1	2	3	4	5	6	7
	Jamais	Rarement	De temps en temps, A l'occasion	Régulièrement	Chaque jour moins d'une heure	Chaque jour plus d'une heure	Chaque jour à temps plein
Utilisation régulière d'un système informatique(*)							
Bureautique (traitement de texte, tableur, etc.)							
Internet (Navigateur, e-mail)							
Jeux video 3D (**)							
Logiciels 3D (***) (Modeleur, d'Animation, de Rendu graphique).							

Notions de dispositif d'affichage.

Niveau de familiarité	1	2	3	4	5	6	7
Dispositifs d'affichage	Jamais	Rarement	De temps en temps, A l'occasion	Régulièrement	Chaque jour moins d'une heure	Chaque jour plus d'une heure	Chaque jour à temps plein
Stéréoscopie (lunettes avec écran)							
Lunettes immersives de "Réalité Virtuelle" (pas d'écran)							
Projection sur grand écran							
Projection immersive "classique" (Géode, Futuroscope...)							
Projection interactive (Cave-Like, Reality Center...)							

Notions de dispositifs d'interaction.

Niveau de familiarité	1	2	3	4	5	6	7
Dispositifs d'interaction.	Jamais	Rarement	De temps en temps, A l'occasion	Régulièrement	Chaque jour moins d'une heure	Chaque jour plus d'une heure	Chaque jour à temps plein
Clavier-souris							
Touchpad, Tablette graphique.							
Instrument de Commande gestuelle 3D (Souris/Stylo 3D, Pantographe, tracker...).							
Commande vocale.							
Autre ? (****)							

(*) Systèmes d'exploitation déjà utilisés.

-
-

(**) Jeux vidéo 3D pratiqués.

-
-

(***) Logiciels 3D déjà pratiqués (Modeleur, d'Animation, de Rendu graphique).

-
-

(****) Dispositif d'interaction n'appartenant à aucune des catégories mentionnées.

-

3

3.1 Machine graphique de sortie

Examinons d'abord le synoptique de la " machine 3D " (Figure 3).

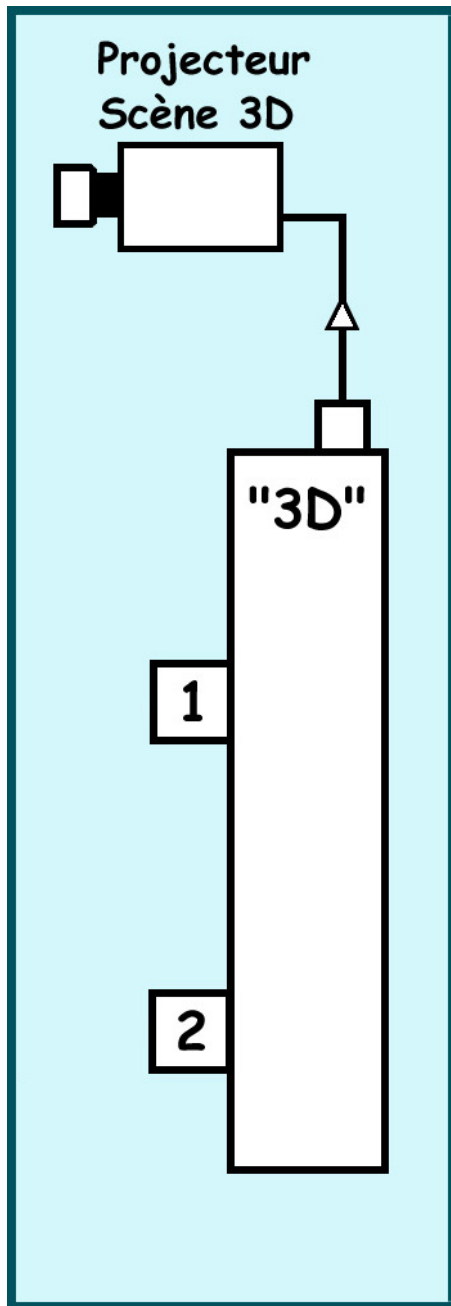


Figure 3. Synoptique de la machine graphique " 3D " en sortie du système.

Le rôle principal de ce sous-ensemble consiste à produire la présentation de l'environnement en sortie de l'application, destinée à l'utilisateur (Figure 2).



Figure 2. Présentation de la scène destinée à l'utilisateur.

Représentation schématique du modèle " Arch " (Figure 4).

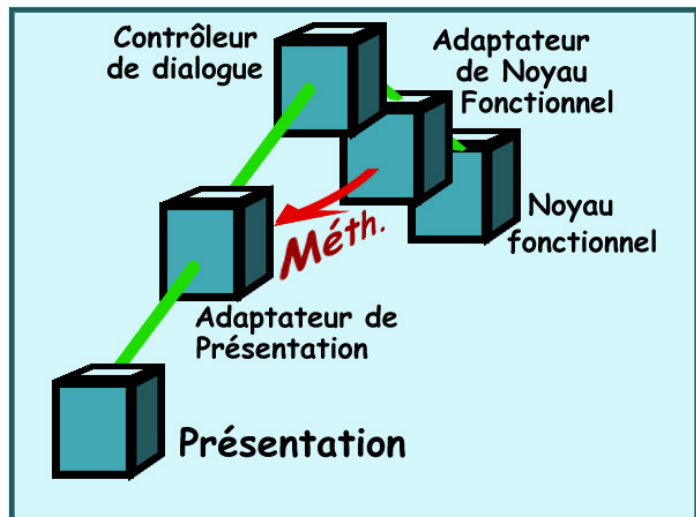


Figure 4. Représentation schématique du modèle " Arch ".

Ce modèle permet de rendre la programmation modulaire en dissociant le noyau fonctionnel de la présentation et du contrôle de dialogue. Le modèle reste souple par ailleurs en admettant des courts-circuits d'invocation de méthodes entre l'adaptateur

de noyau fonctionnel et l'adaptateur de présentation. Mais d'une manière générale, le noyau fonctionnel reste distinct de la présentation.

En dissociant, sur notre application, la présentation en sortie sur une machine distante, nous avons en quelque sorte délocalisé une partie du noyau fonctionnel et du contrôle de dialogue sur cette application séparée de présentation. Cette partie fonctionnelle était ici nécessaire pour constituer une réactivité observable de la scène aux sollicitations de l'utilisateur. Nous conviendrons donc que cette manière provisoire de procéder n'est pas totalement conforme au modèle qui reste donc " approximativement " adopté.

Voyons maintenant les entrées du système.

3.2 Equipement de l'utilisateur en entrée.

L'utilisateur est équipé de plusieurs dispositifs convoyant les informations des trois modalités retenues, lesquels sont collectés sur une captation de la parole, sur les mouvements de la tête de l'utilisateur, et sur les mouvements de la main droite.

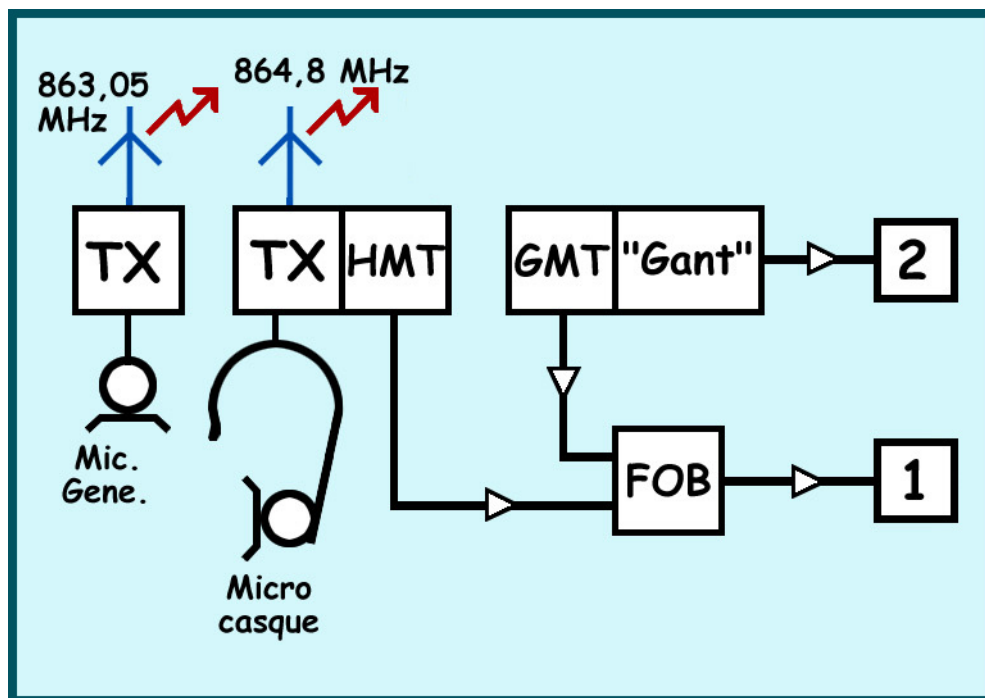


Figure 5. Equipement de mesure porté par l'utilisateur.

Le synoptique Figure 5 ci-dessus signale une capture des énoncés à l'aide d'un micro casque, retransmis par un émetteur audio spécialisé pour la voix. (" TX " 864,8MHz).

Outre ce micro casque, la tête de l'utilisateur est également équipée d'un capteur magnétique (" HMT " ¹ qui transmet une mesure de position et d'orientation de la tête de l'utilisateur.

¹ HMT : " Head Magnetic Tracker ".

La main de l'utilisateur est équipée d'un " gant de données " pourvu de deux dispositifs. Un second capteur magnétique (" **GMT** "²) fournit les informations de position et d'orientation de ce gant, cependant qu'une micro application embarquée produit des informations de " *pose* " de la main de l'utilisateur qui l'anime sous forme sérielle (sortie " 2 ").

Les mesures de capture magnétique (Tête et main de l'utilisateur), sont centralisées sur un analyseur " **FOB** "³ sur lequel ces informations sont rendues disponibles sous forme sérielle à haut débit (sortie " 1 ").

² GMT : " Glove Magnetic Tracker ".

³ FOB : " Flock of Birds ", Ascension Technology.

3.3 Equipement du " Magicien ".

Le sous-ensemble de commande dont dispose le " magicien " lui sert à orchestrer le déroulement des opérations (Figure 6).

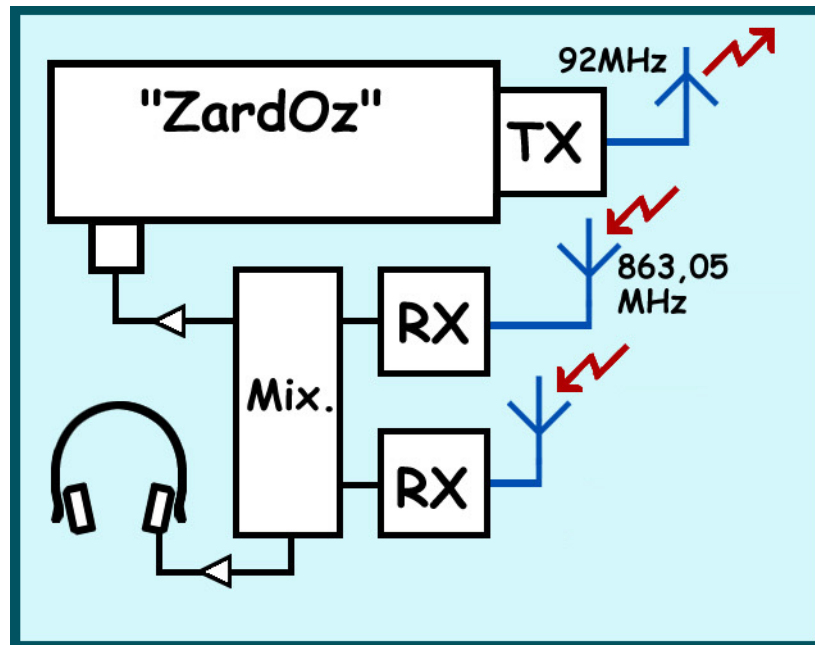


Figure 6. Le sous-ensemble de commande du " magicien ".

Ce sous-ensemble de commande comprend une machine baptisée " *ZardOz* " (contraction de " *Wizard of Oz* " qui rappellera le titre d'un film des années 70 à certains).

Cette machine à disposition du " magicien " dispose d'un émetteur de signaux digitaux de commande en sortie et d'une source audio en entrée.

Cette source audio en entrée transmet les énoncés de l'utilisateur provenant du micro casque qu'il porte (*cf.* Figure 5 p. 23) afin d'être enregistrés sous forme digitale sur disque dur.

Sur un second canal d'information sonore sur casque stéréophonique, le " magicien " perçoit l'ensemble des productions sonores de la salle d'expérimentation (en particulier les énoncés de l'utilisateur, les productions sonores synthétiques du système destinées à guider l'utilisateur, ainsi que les signaux sonores du " récepteur de commande " bornant les enregistrements).

Une table de mixage (" Mix. ") permet différents réglages de niveau et sélection de canal.

Revenons maintenant aux commandes transmises à l'aide de " *l'émetteur de commande* ".

Ces commandes peuvent concerner deux cibles distinctes.

La première cible est la " *machine 3D* " gérant la scène 3D avec les objets 3D qui s'y trouvent. Ces commandes qui lui sont spécifiquement destinées sont celles de repositionnement des objets des tâches sur la scène, ou encore de mise en rotation de ceux-ci autour de l'axe vertical.

La seconde cible est à la fois la " *machine d'entrée* " et le " *récepteur de commande* ". Le démarrage d'une présentation de tâche se traduit par celle de l'éclairage de l'utilisateur au niveau du " *récepteur de commande* ", et par le démarrage de la présentation sur la " *machine d'entrée* ". La mise en service de la fonction d'enregistrement se traduit par l'émission d'un signal sonore sur le " *récepteur de commande* " et l'enregistrement des informations magnétiques de positions et d'orientations sur la " *machine d'entrée* ".

3.4 Récepteur de commande.

Le " *récepteur de commande* " est l'organe par lequel transitent les commandes digitales envoyées par radio depuis la machine " *ZardOz* " du " *magicien* " (Figure 7).

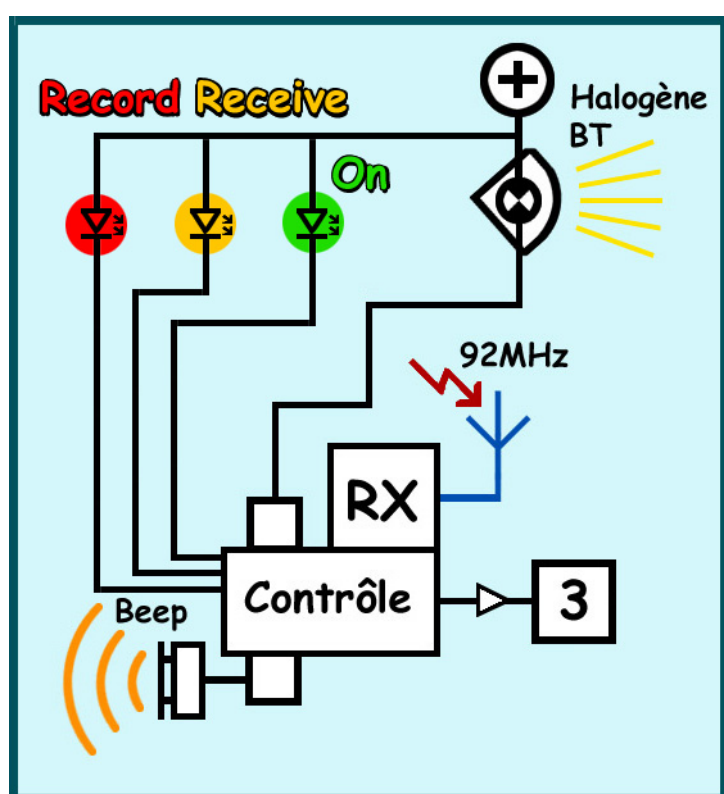


Figure 7. Récepteur de commande.

La fonction principale de ce montage consiste à retransmettre toutes les informations de commande provenant du " *magicien* " à la " *machine d'entrée* " sous forme sérielle.

Ce récepteur reprend toutefois certaines commandes à son compte pour signaler un état de la fonction d'enregistrement (allumage de la Led rouge en façade), signaler par un bip sonore à l'utilisateur qu'il peut formuler son expression de commande et mettre l'éclairage de l'utilisateur en service afin de fournir suffisamment de lumière à la caméra d'enregistrement vidéo.

Initialement, la mise en service d'un émetteur VHF de transmission des images d'une caméra était également prévue à l'aide d'une sortie supplémentaire à collecteur ouvert. Cette fonction a été remplacée par l'emploi d'une caméra d'enregistrement DV sur cassette mini-DV suite à un dysfonctionnement de la caméra mise en œuvre.

Le " *récepteur de commande* " fait appel à un microcontrôleur Basic Stamp II de Parallax. Le circuit imprimé utilisé est une simple face qui fut également utilisé pour construire " *l'émetteur de commande* " en sortie de la machine " *ZardOz* " du " *magicien* ".

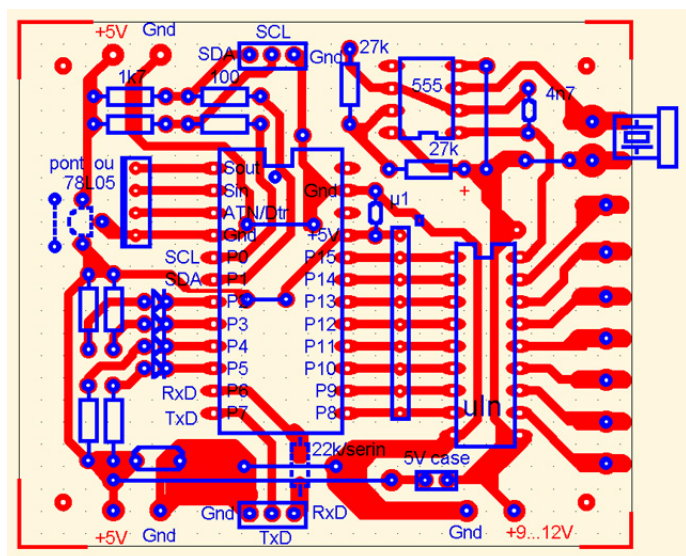


Figure 8. Circuit imprimé et câblage du récepteur de commande.

Le nombre de composants à mettre en place reste minime. Les trois circuits intégrés utilisés sont le BS2, un NE555 monté en multivibrateur astable délivrant une fréquence proche de 3KHz (fréquence correspondant à la plus grande sensibilité de perception sonore de l'oreille humaine), et un ULN 2800, réseau de transistors à collecteur ouvert sur drain commun.

Le récepteur VHF utilisé est un modèle courant Radiométrix recevant sur la bande 92MHz.

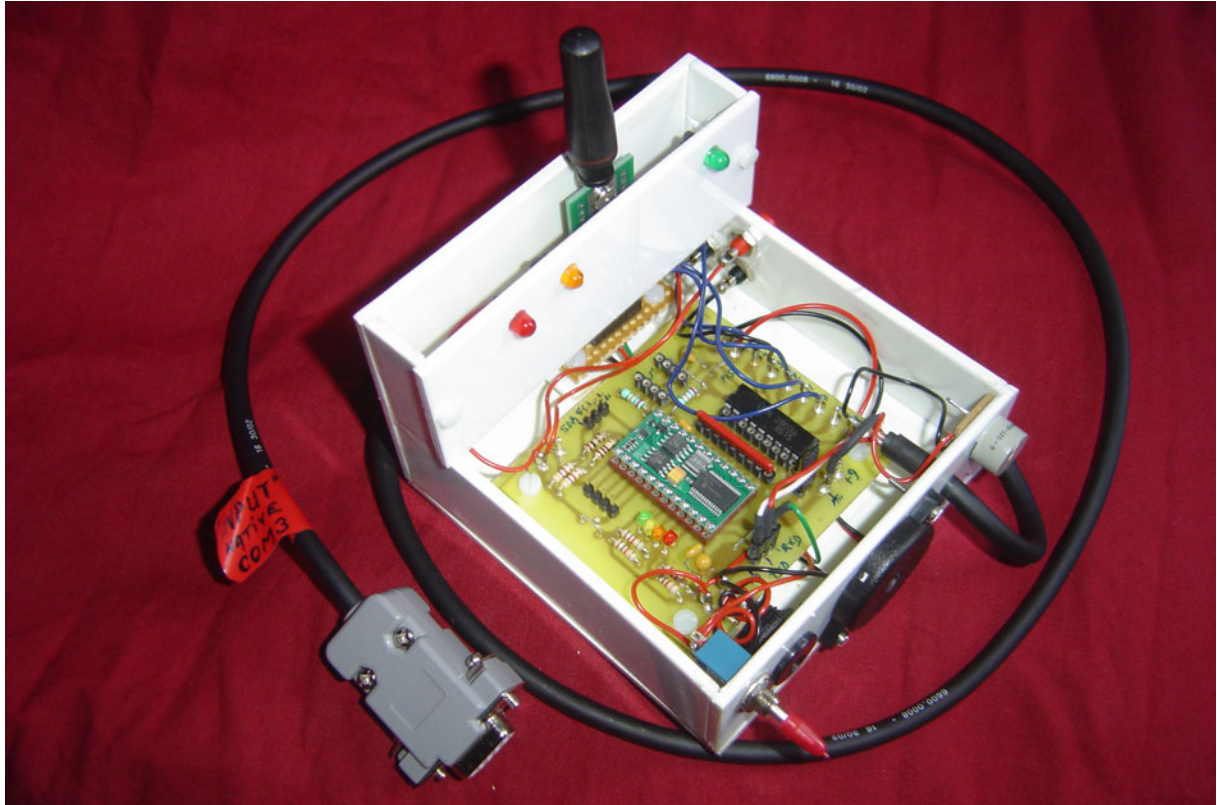


Figure 9. Aperçu du " récepteur de commande " dans son boîtier.

Le montage prend place dans boîtier visible Figure 9. On y remarque les LEDs positionnées en sorte d'être accessible au regard.

4 Détection des "poses" manuelles

L'acquisition de dispositifs d'entrée relatifs aux gestuelles utilisateurs (tracking magnétique de la tête et d'une main de l'utilisateur), ayant pris du retard au cours de l'année 2003, l'incertitude sur ces acquisitions conduisit au développement parallèle d'un dispositif artisanal de sémiotique manuelle indiquant les "poses" déictiques et "plan" de la main de l'utilisateur.

Une autre raison plus importante contribuant à ce développement est l'incapacité d'un gant de données financièrement "accessible" au projet à fournir des informations fiables quant aux mouvements du pouce.

Le gant de données le plus accessible, expérimenté en 1999⁴, contient en effet habituellement cinq fibres optiques permettant la mesure de flexion de chaque doigt sur un octet. Une telle mesure peut être considérée comme pertinente en ce qui concerne quatre des doigts d'une main, cependant, une mesure de flexion simple du pouce ne permet pas de rendre compte de la diversité de ses mouvements. Une réalisation simple fut donc envisagée afin de détecter les "poses" relevant des modalités gestuelles manuelles définies.

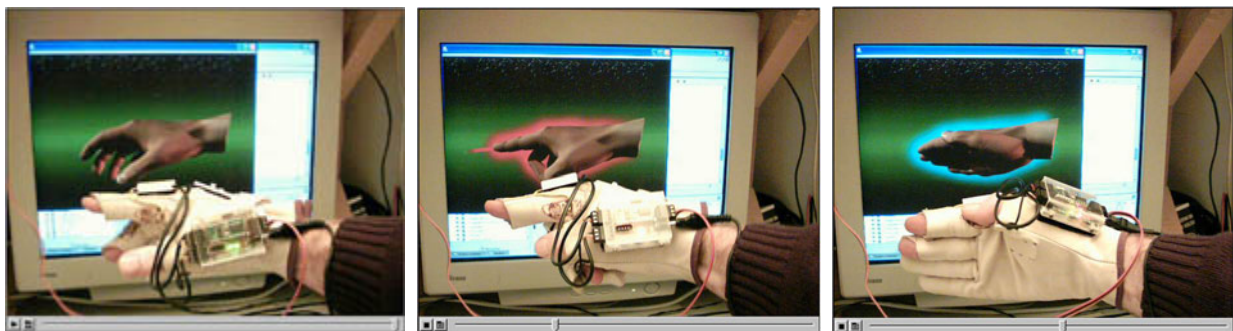


Figure 11. Trois extraits de la vidéo montrant le gant en fonctionnement. La seconde figure montre le signalement nécessaire du geste de désignation d'une direction.

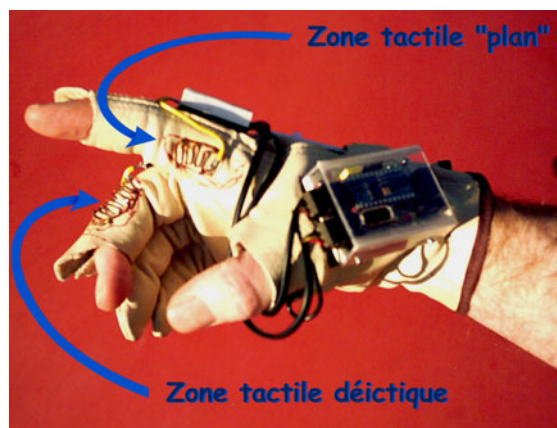


Figure 12. Disposition des capteurs sensitifs tactiles.

Ce dispositif est basé sur un gant de jardinage en cuir, modifié pour supporter un microcontrôleur et deux capteurs sensitifs tactiles à l'aide de Velcro.

Ces capteurs sensitifs figurent essentiellement la contrepartie de boutons poussoirs extra plats disposés en deux endroits supposés dénoter l'une des poses particulières lors d'un contact avec la peau d'un doigt, (le pouce en l'occurrence). La

⁴ Un tel gant de données fut mis en œuvre au cours d'un projet de master of computer science in HCI. C. Tison, « Immersive Modelling Aids for Building Designers in Java3D and VRML », Herriot-Watt University, Edinburgh, 1999.

première surface sensitive figure sur la deuxième phalange du majeur que vient toucher le pouce lors d'une pose déictique. La seconde surface sensible prend place sur la première articulation de l'index, là où le pouce entre en contact avec le métacarpe lors d'une pose "plan".

4.1 Principe

Le principe de fonctionnement de l'ensemble repose sur la comparaison d'une mesure de fréquence d'un multivibrateur astable (Figure 13). La fréquence initiale du multivibrateur au repos reste constante. Lorsqu'une sollicitation d'un contact digital dérive une partie des circulations de courant, le multivibrateur perd sa stabilité.

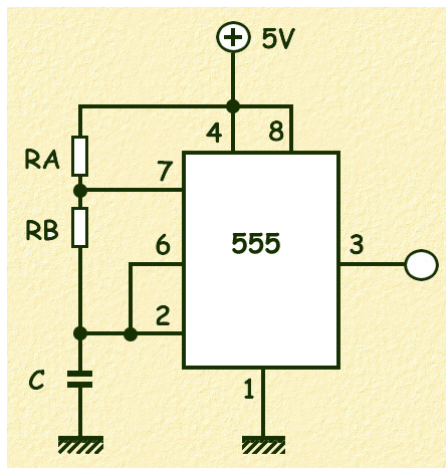


Figure 13. Principe du multivibrateur astable.

disponibilité en fond de tiroir, (10nF, 33K, 1460Hz). Sinon, ces capteurs fonctionnent sur la base d'un condensateur de 10nF chargé au travers de 2 x 330Ω (et par conséquent déchargé via 330Ω).

C'est précisément cette perte de stabilité qui est mesurée par le microcontrôleur pour signaler qu'un contact est effectivement en cours sur le canal de capture correspondant.

On trouvera, plus loin, le détail de fonctionnement du multivibrateur astable ainsi que sa mise en œuvre particulière dans le cadre de cette réalisation.

Dans le cas présent, c'est un NE555 qui est utilisé, l'assemblage ayant été réalisé dans l'urgence (le premier prototype ainsi que sa programmation fut réalisé en moins d'une semaine), les composants utilisés furent sélectionnés en fonction de leur

Le montage donne satisfaction lors d'un contact. Hors contact, on peut évaluer la remarquable stabilité du signal en sortie à l'aide d'un magnétoscope, ou encore « l'écouter » à l'aide d'un simple haut parleur (intercaler une résistance de quelques K Ω).

Le rôle du microcontrôleur est à la fois de relever les mesures sur les capteurs et d'en transmettre les informations au PC via une liaison série. La Figure 14 présente une copie du circuit réalisé. Initialement, Le contrôleur pressenti était le « Javelin » du fait que celui-ci se programme en Java. Le contrôleur utilisé fut finalement un Basic Stamp II. Nous ne l'avons équipé que de quelques

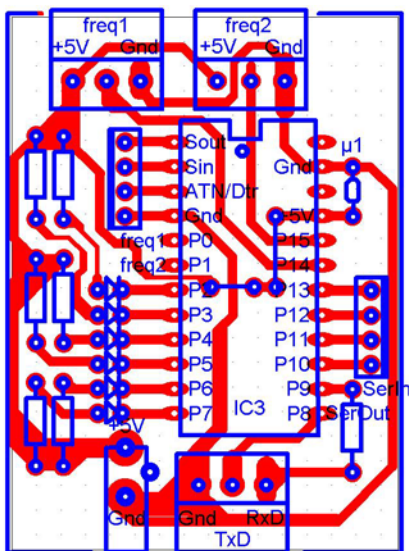


Figure 14. PCB du circuit à microcontrôleur.

LEDs permettant de signaler localement son état interne de bon fonctionnement, l'état de chacune des entrées de capture de manière visuelle (vert : hors contact, rouge : contact !).

Le programme installé sur le microcontrôleur est rédigé dans un langage basic propriétaire (*freq-in-BI-555-3.bs2*), et reste relativement simple malgré plusieurs branchements inconditionnels propres à ce type langage.

Deux classes java de récupération des signaux et d'affichage du résultat ont été écrites pour visualiser la réaction fonctionnelle de l'ensemble afin de pouvoir en filmer une démonstration de fonctionnement déjà mentionnée précédemment.

Ces deux classes se trouvent sont "*gant03bi555v03.java*" et "*montreMain.java*".

La classe de récupération des signaux "*gant03bi555v03.java*" fait appel à l'API "*comm*" de Java qui permet d'adresser le port série.

La classe "*montreMain.java*", qui gère l'affichage contient trois méthodes affichant trois image différentes reproduites ci-après en fonction des informations provenant des deux capteurs. Ces deux classes ne furent écrites que pour réaliser une vidéo placée sur un serveur à des fins de démonstration distante (*gant01.wmv*).



Figure 15. Copies des images utilisées pour réaliser la démonstration filmée du fonctionnement du capteur.

4.2 AMV Multivibrateur Astable

Un moyen de conversion de certaines grandeurs physiques utilisant très peu de composants, à partir d'éléments capteurs se comportant soit comme une **résistance variable**, soit comme une **capacité variable**, consiste à les utiliser parmi les éléments qui déterminent la fréquence d'oscillation d'un **multivibrateur astable** basé sur un composant intégré très courant, le **NE555**.

L'oscillation obtenue peut ensuite être transmise sans problème sur une ligne de plusieurs mètres.

Reprise du cours sur Multivibrateur Astable.

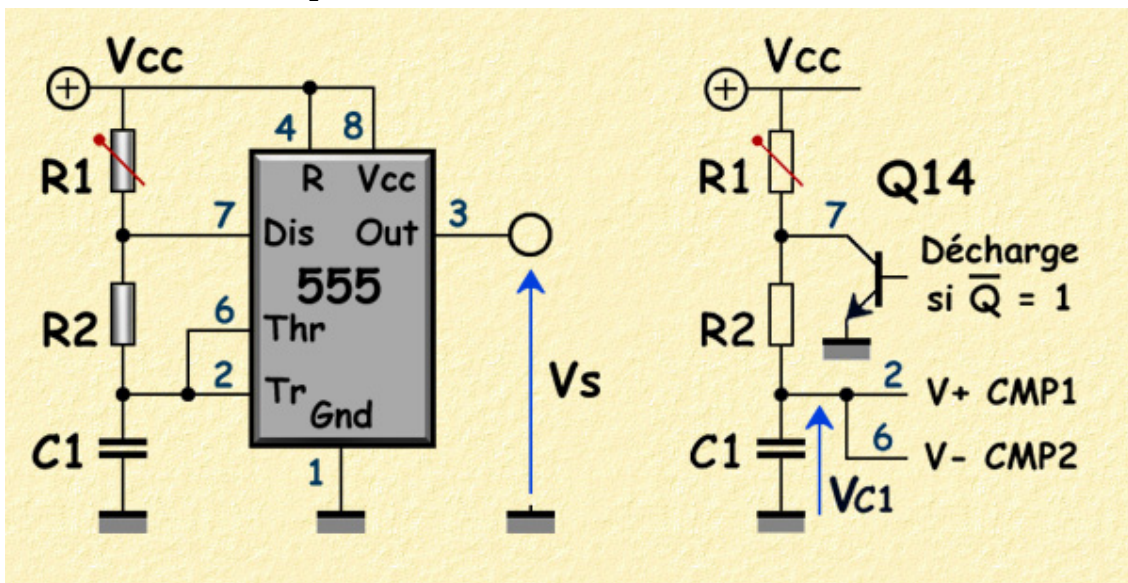


Figure 16. Schéma de principe de base et schéma équivalent.

Fonctionnement.

Partons du cas initial pour lequel la capacité est déchargée. L'alimentation du circuit implique que la charge du condensateur s'effectue au travers des résistances R1 et R2. La charge caractéristique du condensateur, en supposant que les entrées de comparateurs n'ont pas d'effet sur la branche

[Vcc + R1 + R2 + C + Gnd] (Résistances d'entrée parfaites = infini).

sera:

$$V_c = V_{cc} \left(1 - e^{-\frac{t}{C(R1+R2)}} \right)$$

Avec $\bar{Q} = 0$ jusqu'à $\frac{2}{3} V_{cc}$, valeur à laquelle **CMP1** bascule à "1", c. à d. jusqu'à

$V_{cc} \left(1 - e^{-\frac{t}{C(R1+R2)}} \right) = \frac{2}{3} V_{cc}$	\Leftrightarrow	$e^{-\frac{t}{C(R1+R2)}} = \frac{1}{3}$
--	-------------------	---

À partir de quoi, $\bar{Q} = 1 \Rightarrow Q14$ devient passant $\Leftrightarrow C$ se décharge via $R2$ à la broche 7.

On a C chargé aux $2/3 V_{cc}$

$$VC \text{ décharge} = \frac{2}{3} V_{cc} \left(e^{-\frac{t}{C \cdot R2}} \right)$$

$CMP1$ rebascule donc immédiatement à "0", on est toujours dans la phase de décharge tant que $\bar{Q} = 1$.

Arrivé à $VC_{déch.} = 1/3 V_{cc}$; on a

$CMP2 \rightarrow 1 \Rightarrow Q \rightarrow 1$

$CMP1 = 0$

Le flanc montant du comparateur de déclenchement repositionne la bascule, $/Q$ passe à "0", \Leftrightarrow le transistor $Q14$ bloque, le condo ne se décharge plus. En d'autres termes, C recommence à charger, mais cette fois à partir de $1/3 V_{cc}$ (voir chrono ci-après).

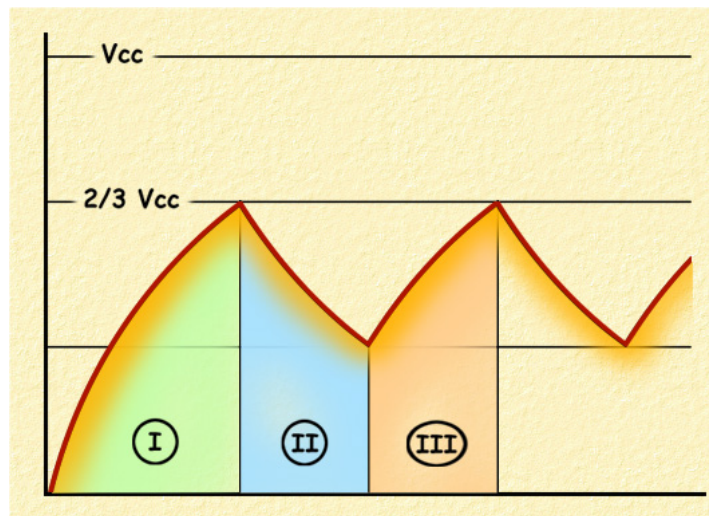


Figure 17. Chronogramme de démarrage d'un multivibrateur.

Bilan: L'équation de base du cas général d'un circuit à condensateur restant:

Charge:	$UC = U0 - (E-U0) \cdot \left(1 - e^{-\frac{t}{RC}} \right)$
Décharge:	$UC = U0 - (E-U0) \cdot e^{-\frac{t}{RC}}$

La tension aux bornes du condensateur va donc "osciller" entre $1/3V_{cc}$ et $2/3V_{cc}$ selon deux expressions:

Charge:	$\frac{1}{3} V_{cc} + V_{cc} (1 - e^{-\frac{t}{C(R1+R2)}})$
Décharge:	$\frac{2}{3} V_{cc} (e^{-\frac{t}{C.R2}})$

Enfin, les deux valeurs extrêmes qui déterminent le fonctionnement du circuit et entre lesquelles oscillera **VC** seront:

Charge maximale:	<p>Pour $\frac{1}{3} V_{cc} + \frac{2}{3} V_{cc} (1 - e^{-\frac{t}{C(R1+R2)}}) = \frac{2}{3} V_{cc}$</p> $\Leftrightarrow 1 - e^{-\frac{t}{\theta}} = \frac{1}{2}$ <p style="text-align: center;">□</p> $\Leftrightarrow t_{ch} = \ln 2 \cdot C(R1+R2) = \text{env. } 0,7 \cdot C(R1+R2)$
Charge minimale:	<p>pour $\frac{2}{3} V_{cc} (e^{-\frac{t}{C.R2}}) = \frac{1}{3} V_{cc}$</p> $\Leftrightarrow e^{-\frac{t}{\theta}} = \frac{1}{2}$ $\Leftrightarrow \frac{t}{\theta} = \ln 2$ $\Leftrightarrow t_{Dch} = \ln 2 \cdot C.R2 = \text{env. } 0,7 \cdot C.R2$

La fréquence d'oscillation du AMV s'exprime donc par

$$F = 1 / [\ln 2 \cdot C(R1+R2) + \ln 2 \cdot C.R2] = 1 / [\ln 2 \cdot C(R1+2.R2)]$$

Obtenir un rapport cyclique de 50% impliquerait ainsi pour **R1 + R2 = R2**. (... !!!)

En pratique, on admet **R1 == 5.R2** env.

$$F = 1 / [\ln 2 \cdot C(7.R2)]$$

Au-delà du seul principe de base de ce circuit, quelques ajouts lui sont éventuellement nécessaires selon l'environnement de fonctionnement.

En particulier, une alimentation distante impliquera une compensation d'impédance des lignes d'alimentation. Une telle précaution n'est pas critique, dans la mesure où des variations de la tension d'alimentation n'affectent pas le comportement du circuit, mais se répercutent par contre sur la valeur de tension en sortie.

Le tampon **C2** voisine la valeur usuelle de **10nF** et **C3** est une capacité de découplage habituelle (multicouche **100nF**). **C4** peut constituer un filtrage optionnel selon la place disponible vu les dimensions d'une capacité de quelques **10µF/16V**.

Pour la limitation de ligne de sortie, une résistance **R3** de **100 Ohms** conviendra.

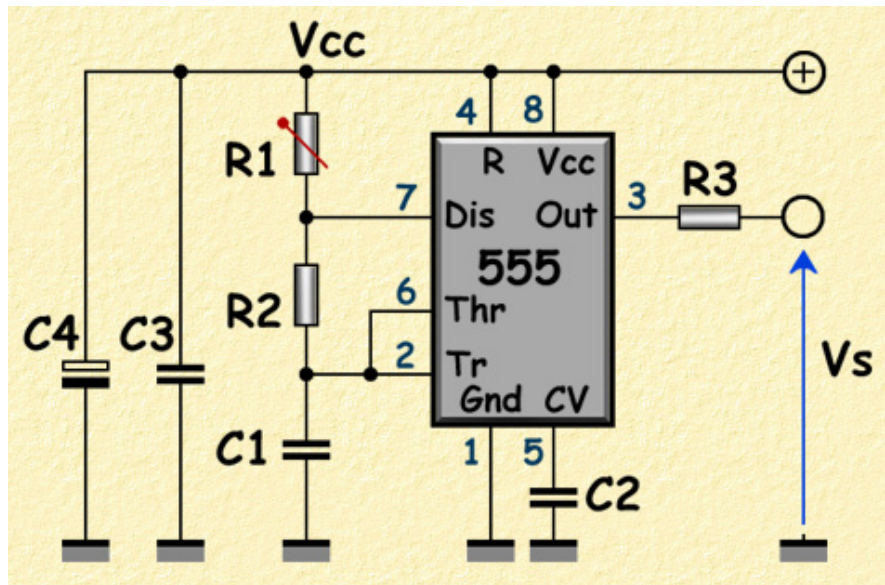


Figure 18. Montage avec filtrage d'alimentation.

Les valeurs **50K-10K-10nF** pour le trio **R1, R2** et **C3** produisent une oscillation de l'ordre de **200Hz** avec un rapport cyclique de 85%-15%.

Travailler sur de telles fréquences présente l'inconvénient de lignes de transmission du signal de courtes distances, et l'intérêt de se situer dans les fréquences audibles qui peuvent être testées rapidement sans nécessiter d'oscilloscope.

On pourrait être tenté, au vu le rapport cyclique de 85%, de ne mesurer que cette partie de la période du signal. La chose est possible, mais la linéarité s'en ressentirait du fait que ce rapport cyclique n'est pas constant. Néanmoins, pour des applications acceptant une erreur de linéarité ou disposant d'une correction logicielle, la chose est envisageable.

En ce qui nous concerne, on déterminera la fréquence en mesurant la période entière.

Avec une technique de mesure de longueur de période, on doit évidemment tenir compte de la durée du pas de comptage disponible.

Les limites de comptage d'un μC disposant d'un timer 16 bits cadencé à 11,0592MHz se rapportent à 1/24 de la fréquence d'horloge en mode compteur, soit env. 500KHz ($t=2\mu\text{s}$).

On obtient ainsi une valeur de comptage maximale qui ne doit pas dépasser 65535 unités de comptage, soit env. 65000 incrément de $2\mu\text{s}$ pour éviter de passer en overflow, c. à d. 130ms ($F_{\text{min}} = 7,7\text{Hz}$), et, pour une résolution minimale de 100 points-incréments pour la période la plus courte, $200\mu\text{s}$, qui nous donnent une F_{max}

de 5KHz, (on pourrait utiliser des fréquences plus élevées, et ainsi une meilleure transmission des signaux de capture sur de plus longues distances, au détriment de la précision de comptage qui approche déjà les 1% d'erreur).

5 Transcription des productions des utilisateurs

La transcription des expressions produites par les utilisateurs est effectuée à l'aide de quatre formes complémentaires de présentation.

Ces quatre formes sont reportées dans les pages qui suivent pour chacune des tâches qui ont été présentées aux participants à l'expérience.

Les enregistrements audio des énoncés des utilisateurs sont tout d'abord retranscrits sous forme dactylographiée afin de constituer une liste de productions verbales pour chaque tâche.

Pour chacune des tâches, ces productions verbales sont ensuite réparties dans un tableau distinguant les unités syntaxiques définies pour les expressions de commande.

La première unité syntaxique retenue est le terme d'opération (le verbe) reporté dans la première colonne du tableau.

La seconde est l'unité de sélection de l'objet de l'opération et figure en seconde colonne du tableau.

La troisième unité syntaxique est l'ensemble constitué par la préposition introduisant la quantité exprimée, la valeur numérique de la quantité, ainsi que l'unité de mesure géométrique linéaire ou angulaire. Ce sous-ensemble de l'énoncé est reporté en troisième colonne du tableau.

La dernière unité est la direction, ou, le cas échéant, un terme reportant à une information spatiale de direction produite à l'aide d'une autre modalité d'entrée, reportés en quatrième et dernière colonne du tableau.

Un tel tableau permet d'identifier rapidement si les informations d'opération, de sélection d'objet, de quantité et de direction ont effectivement été produites (indépendamment de l'ordre dans lequel elles ont été produites).

Une troisième forme de présentation consiste en un tableau analytique caractérisant les énoncés des participants tel que décrit à la section 6.4.2. Ce tableau signale de façon binaire les diverses erreurs d'usage ou encore les styles caractéristiques qui ont été identifiées au cours des premiers tests de mise au point de l'expérience. Les erreurs typiques sont par exemples un usage inadéquat de la préposition introduisant la quantité, les styles caractéristiques sont par exemple les

inversions d'ordre dans la production des unités syntaxiques. Un « 1 » dans le tableau signale la présence de la caractéristique correspondante en tête de colonne. Une telle présentation binaire des erreurs ou usages de la langue au sein des énoncés produits facilite le traitement ultérieur de ces informations à l'aide de calcul.

Nous rappelons page suivante la légende relative aux abréviations en tête de chaque colonne de ces tableaux.

La dernière forme de présentation consiste en un tableau lui aussi binaire qui rapporte la composition non verbale de l'expression. Ce tableau signale si l'objet de l'opération a été indiqué à l'aide du dispositif de visée, si un geste déictique manuel de désignation redondante de l'objet de l'opération a également été produit. Il signale par ailleurs si la destination a été indiquée par une visée, ou si une direction a été indiquée de façon manuelle. Il signale enfin si un second objet a été indiqué à l'aide d'une visée, et si une indication d'un second objet a été réalisée de façon manuelle.

Légende du tableau analytique caractérisant les énoncés des participants.

DEG: Direction manuelle indiquée de visu par l'utilisateur mais sans usage correct du dispositif de capture (défaut de conception du "gant" ou geste déictique sans contact du pouce sur la deuxième phalange du majeur). Le cas échéant, ce défaut du gant ne devrait apparaître que pour les opérations impliquant un acte déictique de la main droite, (c'est à dire F2 et F9).

EIC: Enoncé incomplet: Opération absente

EIP: Enoncé incomplet: Absence de **préposition**

EIQ: Enoncé incomplet: Quantité de mouvement absente

EID: Enoncé incomplet: Direction absente ou incomplète

ECA: Enoncé complet: Mais correspondant à une **autre tâche** (En cas d'incompréhension ou de confusion sur la tâche proposée, ou de confusion en cours d'énonciation suite à surcharge

ECI: Enoncé complet: Sans Identification possible d'une destination claire.

VSOO: Variation d'ordre syntaxique: Inversion de l'ordre syntaxique d'indication de l'**opération** et de la **sélection de l'objet** du mouvement.

VSQD: Variation d'ordre syntaxique: Inversion de l'ordre syntaxique d'indication de la **quantité** de mouvement et de la **direction**.

ALDP: Abus de langage: Usage de **verbes** dédiés au **déplacement** en lieu et place de verbes dédié au **positionnement** (stricto sensu).

Cet abus ne saurait apparaître que pour les tâches de positionnement non ambiguës (F1, F7). Toutefois, F4 peut, par son absence de spécification de quantité de mouvement, apparaître aussi bien comme positionnement que comme déplacement.

ALDI: Abus de langage: Usage d'une **partie** seulement de l'énoncé permettant cependant l'indication complète d'une **direction**.

UFPI: Confusion d'usage de la langue française: Usage d'une **préposition inappropriée**.

UFVI: idem: Usage de **verbes inappropriés** pour la tâche proposée, verbe signifiant une autre opération cependant possible.

UFLI: idem: Usage de **latéralité inversée** (usage de "à droite" pour "à gauche" et inversement).

TID: Addition de termes inutiles à l'énoncé: Usage d'un acte déictique manuel non requis pour l'expression. Les actes déictiques manuels d'indication d'une direction ne sont strictement nécessaires que pour les tâches F2 et F9.

NF: Erreurs d'usage de la langue française: (l'énoncé n'est pas conforme à la langue)

InS: Insuccès. Pas d'énoncé permettant de situer la destination du mouvement.

TIn: Tâche incomprise. L'observation de la tâche proposée n'a pas permis de comprendre le mouvement à effectuer.

Productions verbales des utilisateurs pour la tâche F1 de positionnement égocentré.

Exemple arbitraire d'énoncé pour la tâche : « *Place cet objet à 20m dans cette direction* ».

Participants informaticiens masculins :

- GP-F1** Bouger.. cet objet à 20m dans cette direction.
- LV-F1** Met c..Cet objet, à 20m de moi, dans cette direction.
- RF-F1** Déplacer cet objet, dans cette direction, de 20m.
- RD-F1** Déplace cet objet, à 20m de moi, selon cette direction.
- RR-F1** Déplace cet objet à 20m dans cette direction .

Participants informaticiens féminins :

- VA-F1** Déplacer cet objet, dans cette direction, à 20m de moi.
- CF-F1** Place cet objet en direction...Dans cette direction à 20m de moi.
- BE-F1** Déplace cet objet, à 20m de moi dans cette direction.
- MC-F1** Déplacer cet objet de.. pour l'amen...à 20m vers moi.
- CI-F1** Déplacer cet objet en diagonale vers cette direction jusqu'à 20m devant.

Participants non informaticiens masculins :

- MM-F1** Déplace cet objet dans cette direction à 20m de moi.
- CV-F1** Cet objet, bouger cet objet à 20m de moi, dans cette direction.
- BC-F1** Déplacer cet objet de 20m dans cette direction.
- BJ-F1** Cet objet, 20m, face à moi .
- LS-F1** Déplacer cet objet de 20m dans cette direction .

Participants non informaticiens féminins :

- GE-F1** Cet objet,.... ...Dans cette direction à 20m.
- RC2-F1** Cet objet, déplacer vers la droite de 20m.
- AC-F1** Déplace cet objet vers cet objet à 20m de moi .
- DC-F1** Déplace cet objet à 20m dans cette direction .
- SI-F1** Cet objet, cet objet, déplacer, à droite, à 20m de moi.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F1.**

Utilisateur / F1	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F1	Bouger	cet objet	à 20m	dans cette direction
LV-F1	Met	c..Cet objet,	à 20m de moi,	dans cette direction.
RF-F1	Déplacer	cet objet,	de 20m.	dans cette direction,
RD-F1	Déplace	cet objet,	à 20m de moi,	selon cette direction.
RR-F1	Déplace	cet objet	à 20m	dans cette direction
VA-F1	Déplacer	cet objet,	à 20m de moi.	dans cette direction,
CF-F1	Place	cet objet	à 20m de moi.	en direction...Dans cette direction
BE-F1	Déplace	cet objet, .	à 20m de moi	dans cette direction
MC-F1	Déplacer	cet objet	de.. pour l'amen...à 20m	vers moi.
CI-F1	Déplacer	cet objet	à 20m devant.	en diagonale vers cette direction jusqu'
MM-F1	Déplace	cet objet	à 20m de moi	dans cette direction
CV-F1	bouger	cet objet	à 20m de moi,	dans cette direction.
BC-F1	Déplacer	cet objet	de 20m	dans cette direction.
BJ-F1		Cet objet,	20m,	face à moi
LS-F1	Déplacer	cet objet	de 20m	dans cette direction
GE-F1		Cet objet,.... ...	à 20m.	Dans cette direction
RC2-F1	déplacer	Cet objet,	de 20m.	vers la droite
AC-F1	Déplace	cet objet	à 20m de moi	vers cet objet
DC-F1	Déplace	cet objet	à 20m	dans cette direction
SI-F1	déplacer,	Cet objet, cet objet,	à 20m de moi.	à droite,

Caractéristiques des énoncés des participants pour la tâche F1.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
LV-F1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F1	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	1	1
RD-F1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
RR-F1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
VA-F1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
CF-F1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
BE-F1	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
MC-F1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1
CI-F1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
MM-F1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0
CV-F1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
BC-F1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0
BJ-F1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
LS-F1	2	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0
GE-F1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
RC2-F1	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	1	1
AC-F1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1
DC-F1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
SI-F1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1	0

Composition non verbale des expressions des utilisateurs pour la tâche F1.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F1	1	0	0	1	0	0
LV-F1	1	0	1	1	0	0
RF-F1	1	0	0	1	0	0
RD-F1	1	0	1	1	0	0
RR-F1	1	0	1	0	0	0
VA-F1	1	0	1	1	0	0
CF-F1	1	1	1	1	0	0
BE-F1	1	0	1	1	0	0
MC-F1	1	0	0	0	0	0
CI-F1	1	0	1	1	0	0
MM-F1	1	0	1	0	0	0
CV-F1	1	1	1	1	0	0
BC-F1	1	0	1	0	0	0
BJ-F1	1	0	0	0	0	0
LS-F1	1	1	1	1	0	0
GE-F1	1	0	1	0	0	0
RC2-F1	1	0	0	0	0	0
AC-F1	1	0	0	0	1	0
DC-F1	1	1	1	1	0	0
SI-F1	1	0	0	0	0	0

Productions verbales des utilisateurs pour la tâche F2 de déplacement exocentré.

Exemple arbitraire d'énoncé pour la tâche: « *Déplace cet objet de 20m dans cette direction* ».

Participants informaticiens masculins :

- GP-F2** Bouger cet objet de 20m dans cette direction
- LV-F2** Sélectionne....Sélectionne cet objet, et déplace-le, de 20m, dans cette direction.
- RF-F2** Déplacer cet objet dans cette direction, de 20m.
- RD-F2** Déplace cet objet de 20...de 20m, dans cette direction.
- RR-F2** Déplace relativement cet objet à 20m dans cette direction.

Participants informaticiens féminins :

- VA-F2** Déplacer cet objet, de 20m, à 90° à gauche.
- CF-F2** Déplace, déplace cet objet de 20m dans cette direction.
- BE-F2** Déplace cet objet de 20m dans cette direction.
- MC-F2** Déplacer ce...cet objet de 20m à gauche.
- CI-F2** Rapprocher cet objet de cet objet de 20m.

Participants non informaticiens masculins :

- MM-F2** Déplace cet objet, dans cette direction, 20m.
- CV-F2** Cet objet, déplacer, de 20m, dans cette direction.
- BC-F2** Déplacer cet objet de 20m à gauche.
- BJ-F2** Déplacer. Cet objet, déplacer, de 20m, ici.
- LS F2** Déplacer cet objet de 20m dans cette direction.

Participants non informaticiens féminins :

- GE-F2** Cet objet,.... ...Dans cette direction, déplacer de 20m.
- RC2-F2** Cet objet, déplacer vers la gauche, d'une distance de 20m.
- AC-F2** Déplace cet objet de 20m sur la gauche dans cette direction.
- DC-F2** Déplace cet objet à 20m dans cette direction.
- SI-F1** Cet objet, déplacer vers la gauche, 20m.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F2.**

Utilisateur / F2	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F2	Bouger	cet objet	de 20m	dans cette direction
LV-F2	et déplace-	Sélectionne....Sélectionne cet objet, ... -le,	de 20m,	dans cette direction.
RF-F2	Déplacer	cet objet	de 20m.	dans cette direction,
RD-F2	Déplace	cet objet	de 20...de 20m,	dans cette direction.
RR-F2	Déplace relativement	cet objet	à 20m	dans cette direction
VA-F2	Déplacer	cet objet,	de 20m,	à 90° à gauche.
CF-F2	Déplace, Déplace	cet objet	de 20m	dans cette direction
BE-F2	Déplace	cet objet	de 20m	dans cette direction.
MC-F2	Déplacer	ce...cet objet	de 20m	à gauche.
CI-F2	Rapprocher	cet objet	de 20m.	de cet objet
MM-F2	Déplace	cet objet	20m	dans cette direction.
CV-F2	déplacer,	Cet objet,	de 20m,	dans cette direction.
BC-F2	Déplacer	cet objet	de 20m	à gauche.
BJ-F2	déplacer,	Cet objet,	de 20m,	ici
LS F2	Déplacer	cet objet	de 20m	dans cette direction
GE-F2	déplacer	Cet objet,.... ...	de 20m.	Dans cette direction,
RC2-F2	déplacer	Cet objet,	d'une distance de 20m.	vers la gauche,
AC-F2	Déplace	cet objet	de 20m	sur la gauche dans cette direction
DC-F2	Déplace	cet objet	à 20m	dans cette direction
SI-F2	déplacer	Cet objet,	20m....	vers la gauche,

Caractéristiques des énoncés des participants pour la tâche F2.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LV-F2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F2	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RD-F2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RR-F2	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
VA-F2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
CF-F2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BE-F2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MC-F2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
CI-F2	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	1	1
MM-F2	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
CV-F2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
BC-F2	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1
BJ-F2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LS-F2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GE-F2	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	1	0
RC2-F2	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1
AC-F2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
DC-F2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
SI-F2	1	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	1	1

Composition non verbale des expressions des utilisateurs pour la tâche F2.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F2	1	0	0	1	0	0
LV-F2	1	0	0	1	0	0
RF-F2	1	0	0	1	0	0
RD-F2	1	0	0	1	0	0
RR-F2	1	0	0	1	0	0
VA-F2	1	0	0	0	0	0
CF-F2	1	1	1	1	0	0
BE-F2	1	0	0	1	0	0
MC-F2	1	0	0	0	0	0
CI-F2	1	0	0	0	0	0
MM-F2	1	0	0	1	0	0
CV-F2	1	0	0	1	0	0
BC-F2	1	0	0	0	0	0
BJ-F2	1	1	0	1	0	0
LS-F2	1	0	0	1	0	0
GE-F2	1	0	1	0	0	0
RC2-F2	1	0	0	0	0	0
AC-F2	1	0	0	1	0	0
DC-F2	1	1	0	1	0	0
SI-F2	1	0	0	0	0	0

Productions verbales des utilisateurs pour la tâche F3 de déplacement relatif orthogonal.

Exemple arbitraire d'énoncé pour la tâche: « *Déplace cet objet de 50m vers la droite* ».

Participants informaticiens masculins :

- GP-F3** Donc, Euh, Bouger cet objet, de 50m vers la droite.
- LV-F3** Déplace cet objet à 50m, vers la droite.
- RF-F3** Déplacer, cet objet, vers la droite, de 50m.
- RD-F3** Déplace cet objet sur la droite, de 50m.
- RR-F3** Déplace cet objet de 50m vers la droite.

Participants informaticiens féminins :

- VA-F3** Déplacer cet objet de 50m vers la droite, à 90°.
- CF-F3** Déplace cet obj...Déplace cet objet à 50m vers la droite.
- BE-F3** Déplace cet objet de 50m vers la droite.
- MC-F3** Déplacer cet objet de 50m à droite.
- CI-F3** Bouger cet objet vers la droite d'une distance de 50m.

Participants non informaticiens masculins :

- MM-F3** Déplace cet objet à droite 50m.
- CV-F3** Cet objet, emmener cet objet à 50m dans cette direction.
- BC-F3** Déplacer cet objet de 35° à droite.
- BJ-F3** Cet objet, bouge, de 50m, à droite.
- LS-F3** Déplacer cet objet de 50m à droite.

Participants non informaticiens féminins :

- GE-F3** Cet objet,.... ...Vers la droite à 50m.
- RC2-F3** Cet objet, euh, alors, déplacer, sur la droite, heuu, 50m.
- AC-F3** Déplace cet objet de 50m sur la droite.
- DC-F3** Déplace cet objet de 50m sur la droite.
- SI-F1** Cet objet, cet...Cet objet, déplacer, à droite, 50m..De 50m.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F3.**

Utilisateur / F3	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F3	Bouger	cet objet	de 50m	vers la droite
LV-F3	Déplace	cet objet	à 50m,	vers la droite.
RF-F3	Déplacer,	cet objet,	de 50m.	vers la droite,
RD-F3	Déplace	cet objet	de 50m.	sur la droite,
RR-F3	Déplace	cet objet	de 50m	vers la droite
VA-F3	Déplacer	cet objet	de 50m	vers la droite, à 90°.
CF-F3	Déplace	cet objet	à 50m	vers la droite
BE-F3	Déplace	cet objet	de 50m	vers la droite.
MC-F3	Déplacer	cet objet	de 50m	à droite.
CI-F3	Bouger	cet objet	d'une distance de 50m.	vers la droite
MM-F3	Déplace	cet objet	50m	à droite
CV-F3	emmener	cet objet	à 50m	dans cette direction.
BC-F3	Déplacer	cet objet	de 35°	à droite.
BJ-F3	bouge,	Cet objet,	de 50m,	à droite
LS F3	Déplacer	cet objet	de 50m	à droite
GE-F3		Cet objet,.... ...	à 50m.	Vers la droite
RC2-F3	déplacer,	Cet objet, euh, alors,	50m.	sur la droite, heuu,
AC-F3	Déplace	cet objet	de 50m	sur la droite
DC-F3	Déplace	cet objet	de 50m	sur la droite
SI-F3	déplacer,	Cet objet, cet...Cet objet,	50m..De 50m.	à droite,

Caractéristiques des énoncés des participants pour la tâche F3.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
LV-F3	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
RF-F3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RD-F3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RR-F3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VA-F3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
CF-F3	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
BE-F3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MC-F3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CI-F3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
MM-F3	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
CV-F3	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1
BC-F3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
BJ-F3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
LS-F3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GE-F3	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0
RC2-F3	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
AC-F3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DC-F3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SI-F3	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0

Composition non verbale des expressions des utilisateurs pour la tâche F3.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F3	1	0	0	1	0	0
LV-F3	1	0	0	0	0	0
RF-F3	1	0	0	0	0	0
RD-F3	1	0	0	0	0	0
RR-F3	1	0	0	0	0	0
VA-F3	1	0	0	0	0	0
CF-F3	1	1	0	1	0	0
BE-F3	1	0	0	0	0	0
MC-F3	1	0	0	0	0	0
CI-F3	1	0	0	0	0	0
MM-F3	1	0	0	0	0	0
CV-F3	1	0	0	1	0	0
BC-F3	1	0	0	0	0	0
BJ-F3	1	0	0	0	0	0
LS-F3	1	0	0	0	0	0
GE-F3	1	0	0	0	0	0
RC2-F3	1	0	0	0	0	0
AC-F3	1	0	0	0	0	0
DC-F3	1	1	0	1	0	0
SI-F3	1	0	0	0	0	0

Productions verbales des utilisateurs pour la tâche F4 de repositionnement égo-centré approximatif.

Exemple arbitraire d'énoncé pour la tâche: « *Place cet objet dans cette direction* ».

Participants informaticiens masculins :

- GP-F4** Bouger cet objet, dans cette direction.
- LV-F4** Tourne cet objet, jusqu'à, cette direction.
- RF-F4** Déplace, cet objet, à 63m de moi, dans cette direction.
- RD-F4** Déplace cet objet, dans cette direction, comme ça.
- RR-F4** Déplace cet objet dans cette direction.

Participants informaticiens féminins :

- VA-F4** Déplacer cet objet d'un angle jusqu'au bord droit de la scène.
- CF-F4** Déplace cet objet...Cet objet à droite dans cette direction.
- BE-F4** Déplace cet objet dans cette direction.
- MC-F4** Déplacer cet objet à droite...de cet objet.
- CI-F4** Faire tourner cet objet dans ce sens de 30° à droite.

Participants non informaticiens masculins :

- MM-F4** Déplace cet objet **selon un arc** dans cette direction.
- CV-F4** Cet objet, amener cet objet derrière la croix jaune, dans cette direction.
- BC-F4** Déplacer cet objet à 2m à gauche de cet objet.
- BJ-F4** Déplacer, cet objet, là-bas.
- LS-F4** Déplacer cet objet dans cette direction.

Participants non informaticiens féminins :

- GE-F4** Cet objet,... ..Déplacer dans cette direction.
- RC2-F4** Cet objet, vers la droite, vers, celui-ci?.
- AC-F4** Déplace cet...Cet objet, dans cette direction.
- DC-F4** Déplace cet objet dans cette direction.
- SI-F4** Cet objet, déplacer, ici!

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F4.**

Utilisateur / F4	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F4	Bouger	cet objet,		dans cette direction.
LV-F4	Tourne	cet objet,		jusqu'à, cette direction.
RF-F4	Déplace,	cet objet,	à 63m de moi,	dans cette direction.
RD-F4	Déplace	cet objet,		dans cette direction, comme ça.
RR-F4	Déplace	cet objet		dans cette direction
VA-F4	Déplacer	cet objet		d'un angle jusqu'au bord droit de la scène.
CF-F4	Déplace	cet objet...Cet objet		à droite dans cette direction.
BE-F4	Déplace	cet objet		dans cette direction.
MC-F4	Déplacer	cet objet		à droite...de cet objet.
CI-F4	Faire tourner	cet objet	de 30°	dans ce sensà droite.
MM-F4	Déplace	cet objet		dans cette direction
CV-F4	amener	cet objet		derrière la croix jaune, dans cette direction.
BC-F4	Déplacer	cet objet	à 2m	à gauche de cet objet.
BJ-F4	Déplacer,	cet objet,		là-bas
LS F4	Déplacer	cet objet		dans cette direction
GE-F4	Déplacer	Cet objet,.... ...		dans cette direction.
RC2-F4		Cet objet,		vers la droite, vers, celui-ci?
AC-F4	Déplace	cet...Cet objet,		dans cette direction
DC-F4	Déplace	cet objet		dans cette direction
SI-F4	déplacer,	Cet objet,		ici!

Caractéristiques des énoncés des participants pour la tâche F4.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LV-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
RF-F4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
RD-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RR-F4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VA-F4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
CF-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
BE-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MC-F4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
CI-F4	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0
MM-F4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
CV-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
BC-F4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
BJ-F4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LS-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GE-F4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
RC2-F4	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
AC-F4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DC-F4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SI-F4	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Composition non verbale des expressions des utilisateurs pour la tâche F4.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F4	1	0	0	1	0	0
LV-F4	1	0	1	1	0	0
RF-F4	1	0	1	0	0	0
RD-F4	1	0	1	1	0	0
RR-F4	1	0	1	0	0	0
VA-F4	1	0	0	0	0	0
CF-F4	1	0	1	1	0	0
BE-F4	1	0	1	1	0	0
MC-F4	1	0	0	0	1	0
CI-F4	1	0	0	0	0	0
MM-F4	1	0	1	0	0	0
CV-F4	1	0	1	1	0	0
BC-F4	1	0	0	0	1	0
BJ-F4	1	1	1	1	0	0
LS-F4	1	0	1	1	0	0
GE-F4	1	0	1	0	0	0
RC2-F4	1	0	1	0	1	0
AC-F4	1	0	1	1	0	0
DC-F4	1	0	1	0	0	0
SI-F4	1	0	1	0	0	0

Productions verbales des utilisateurs pour la tâche F5 de giration relative.

Exemple arbitraire d'énoncé pour la tâche: « *Déplace cet objet de 30° vers la droite* ».

Participants informaticiens masculins :

- GP-F5** Bouger cet objet de 30° dans la direction de celui-ci.
- LV-F5** Déplace cet objet, de 30°, vers la droite.
- RF-F5** Déplacer, cet objet, vers la droite, de 30°.
- RD-F5** Déplace cet objet sur la droite, sur...sur la droite, de 30°.
- RR-F5** Déplace cet objet de 30° vers la droite.

Participants informaticiens féminins :

- VA-F5** Déplacer cet objet de 30° vers la droite.
- CF-F5** Déplace l'objet bleu vers la droite jusqu'à l'objet jaune.
- BE-F5** Déplace cet objet de 30° vers la droite.
- MC-F5** Déplacer cet objet d'un angle de 30° à droite.
- CI-F5** Faire tourner cet objet d'un angle de 30° comme ça, vers la droite.

Participants non informaticiens masculins :

- MM-F5** Déplace cet objet sur la droite *selon un arc* de 30°.
- CV-F5** Cet objet, déplacer cet objet de 30° par rapport à ma position, vers la droite.
- BC-F5** Déplacer cet objet de 30° dans cette direction.
- BJ-F5** Cet objet, Cet objet, bouger, de 30° à droite.
- LS-F5** Déplacer cet objet de 30° sur la droite.

Participants non informaticiens féminins :

- GE-F5** Cet objet,.... ..Faire une rotation de 30°, vers cet objet.
- RC2-F5** Cet objet, vers la droite, de 90°.
- AC-F5** Déplace cet objet, 30°, dans cette direction.
- DC-F5** Déplace cet objet de 30° sur la droite.
- SI-F5** Oui, sélect...Euh..Cet objet, déplacer vers la droite, avec un angle de 30°.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F5.**

Utilisateur / F5	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F5	Bouger	cet objet	de 30°	dans la direction de celui-ci.
LV-F5	Déplace	cet objet,	de 30°,	vers la droite.
RF-F5	Déplacer,	cet objet,	de 30°.	vers la droite,
RD-F5	Déplace	cet objet	de 30°.	sur la droite, sur...sur la droite,
RR-F5	Déplace	cet objet	de 30°	vers la droite
VA-F5	Déplacer	cet objet	de 30°	vers la droite.
CF-F5	Déplace	l'objet bleu		vers la droite jusqu'à l'objet jaune
BE-F5	Déplace	cet objet	de 30°	vers la droite.
MC-F5	Déplacer	cet objet	d'un angle de 30°	à droite.
CI-F5	Faire tourner	cet objet	d'un angle de 30° comme ça,	vers la droite.
MM-F5	Déplace	cet objet	de 30°	sur la droite
CV-F5	déplacer	cet objet	de 30° par rapport à ma position,	vers la droite.
BC-F5	Déplacer	cet objet	de 30°	dans cette direction.
BJ-F5	bouger,	Cet objet, Cet objet,	de 30°	à droite
LS F5	Déplacer	cet objet	de 30°	sur la droite
GE-F5	Faire une rotation	Cet objet,.... ...	de 30°,	vers cet objet.
RC2-F5		Cet objet,	de 90°.	vers la droite,
AC-F5	Déplace	cet objet,	30°,	dans cette direction
DC-F5	Déplace	cet objet	de 30°	sur la droite
SI-F5	déplacer	Oui, sélect...Euh..Cet objet,	avec un angle de 30°.	vers la droite,

Caractéristiques des énoncés des participants pour la tâche F5.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F5	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1
LV-F5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RD-F5	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RR-F5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VA-F5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CF-F5	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1
BE-F5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MC-F5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CI-F5	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
MM-F5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
CV-F5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
BC-F5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
BJ-F5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
LS-F5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GE-F5	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0
RC2-F5	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
AC-F5	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
DC-F5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SI-F5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0

Composition non verbale des expressions des utilisateurs pour la tâche F5.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F5	1	0	0	0	0	1
LV-F5	1	1	0	0	0	0
RF-F5	1	0	0	0	0	0
RD-F5	1	0	0	0	0	0
RR-F5	1	0	0	0	0	0
VA-F5	1	0	0	0	0	0
CF-F5	1	1	0	1	1	1
BE-F5	1	0	0	0	0	0
MC-F5	1	0	0	0	0	0
CI-F5	1	0	0	0	0	0
MM-F5	1	0	0	0	0	0
CV-F5	1	0	0	1	0	0
BC-F5	1	0	1	0	0	0
BJ-F5	1	0	0	0	0	0
LS-F5	1	0	0	0	0	0
GE-F5	1	0	0	0	1	0
RC2-F5	1	0	0	0	0	0
AC-F5	1	0	1	0	0	0
DC-F5	1	1	0	1	0	0
SI-F5	1	0	0	0	0	0

Productions verbales des utilisateurs pour la tâche F6 de déplacement ciblé.

Exemple arbitraire d'énoncé pour la tâche: « *Déplace cet objet de 20m en direction de cet objet* ».

Participants informaticiens masculins :

- GP-F6** Bouger cet objet de 20m dans la direction de celui-ci.
- LV-F6** Rapproche cet objet, de 20m, de celui-ci.
- RF-F6** Déplacer, cet objet, en direction de, cet ob..Cet objet, de 20m.
- RD-F6** Déplace cet objet, de 20m, en direction, de cet objet.
- RR-F6** Déplace moi cet objet à 20m dans la direction de cet objet.

Participants informaticiens féminins :

- VA-F6** Déplacer cet objet de 20m dans la direction de cet objet.
- CF-F6** Déplace l'obj...Cet objet de 20m en direction de cet...De cet objet.
- BE-F6** Déplace cet objet de 20m, vers cet objet.
- MC-F6** Déplacer cet objet de 20m à droite.
- CI-F6** Avancer cet objet vers cet objet d'une distance de 20m.

Participants non informaticiens masculins :

- MM-F6** Déplace cet objet dans cette direction 20m.
- CV-F6** Cet objet, bouger cet objet de 20m vers la croix jaune.
- BC-F6** Déplacer cet objet en direction de cet objet...
- BJ-F6** Cet objet, Cet objet, bouge, de 20m, en direction de cet objet.
- LS-F6** Déplacer cet objet de 20m à droite de l'objet rouge.

Participants non informaticiens féminins :

- GE-F6** Cet objet, déplacer vers cet objet, de 20m.
- RC2-F6** Cet objet, le déplacer à droite sur une distance de 20m.
- AC-F6** Déplace cet objet de 20m vers cet objet.
- DC-F6** Déplace cet objet à 20m de cet objet dans cette direction.
- SI-F1** Alors attend, Ha oui, ha ha ha, cet objet, déplacer vers la droite, de 20m.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F6.**

Utilisateur / F6	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F6	Bouger	cet objet	de 20m	dans la direction de celui-ci
LV-F6	Rapproche	cet objet,	de 20m,	de celui-ci.
RF-F6	Déplacer,	cet objet,	de 20m.	en direction de, cet ob..Cet objet,
RD-F6	Déplace	cet objet,	de 20m,	en direction, de cet objet.
RR-F6	Déplace moi	cet objet	à 20m	dans la direction de cet objet
VA-F6	Déplacer	cet objet	de 20m	dans la direction de cet objet.
CF-F6	Déplace	l'obj...Cet objet	de 20m	en direction de cet...De cet objet.
BE-F6	Déplace	cet objet	de 20m,	vers cet objet.
MC-F6	Déplacer	cet objet	de 20m	à droite.
CI-F6	Avancer	cet objet	d'une distance de 20m.	vers cet objet
MM-F6	Déplace	cet objet	20m	dans cette direction
CV-F6	bouger	cet objet	de 20m	vers la croix jaune.
BC-F6	Déplacer .	cet objet		en direction de cet objet....
BJ-F6	bouge,	Cet objet, Cet objet,	de 20m,	en direction de cet objet
LS F6	Déplacer	cet objet	de 20m	à droite de l'objet rouge
GE-F6	déplacer	Cet objet,	de 20m.	vers cet objet,
RC2-F6	le déplacer	Cet objet,	sur une distance de 20m.	à droite
AC-F6	Déplace	cet objet	de 20m	vers cet objet.
DC-F6	Déplace	cet objet	à 20m de cet objet	dans cette direction
SI-F6	déplacer	Cet objet,	de 20m.	vers la droite,

Caractéristiques des énoncés des participants pour la tâche F6.

Utilisa- -teur	Deict	DEG	EIC	EIP	EQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LV-F6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RD-F6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RR-F6	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
VA-F6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CF-F6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
BE-F6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MC-F6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
CI-F6	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
MM-F6	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1
CV-F6	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
BC-F6	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
BJ-F6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
LS-F6	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
GE-F6	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
RC2-F6	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1
AC-F6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DC-F6	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	1
SI-F6	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1

Composition non verbale des expressions des utilisateurs pour la tâche F6.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F6	1	0	0	0	1	1
LV-F6	1	0	0	0	1	0
RF-F6	1	0	0	0	1	0
RD-F6	1	0	0	0	1	0
RR-F6	1	0	0	0	1	0
VA-F6	1	0	0	0	1	0
CF-F6	1	1	0	0	1	1
BE-F6	1	0	0	0	1	0
MC-F6	1	0	0	0	0	0
CI-F6	1	0	0	0	1	0
MM-F6	1	0	0	0	0	0
CV-F6	1	0	0	0	0	0
BC-F6	1	0	0	0	1	0
BJ-F6	1	0	0	0	1	0
LS-F6	1	0	0	0	0	0
GE-F6	1	0	0	0	1	1
RC2-F6	1	0	0	0	0	0
AC-F6	1	0	0	0	1	0
DC-F6	1	1	0	0	1	1
SI-F6	1	0	0	0	0	0

**Productions verbales des utilisateurs pour la tâche F7
de positionnement orthogonal référencé à un autre objet.**

Exemple arbitraire d'énoncé pour la tâche: « *Place cet objet à 20m à droite de cet objet* ».

Participants informaticiens masculins :

- GP-F7** Bouger cet objet à 20 m sur la droite de cet objet.
- LV-F7** Met cet objet, à 20m, de celui-ci, en direction d' la droite.
- RF-F7** Déplacer, cet objet, à 20m à droite, de cet objet.
- RD-F7** Déplace cet objet, de 20m, sur la droite par rapport à cet objet.
- RR-F7** Déplace cet objet à 20m à droite de cet objet.

Participants informaticiens féminins :

- VA-F7** Déplacer cet objet, de 20m à 90° de cet objet.
- CF-F7** Déplace cet objet à 20m à droite de cet objet.
- BE-F7** Déplace cet objet, à 20m de cet objet, vers la droite.
- MC-F7** Déplacer cet objet pour l'amener à 20m à droite de c...de cet objet.
- CI-F7** Avancer cet objet vers la droite d'une distance de 20m....

Participants non informaticiens masculins :

- MM-F7** Déplace cet objet à droite de cet objet à 20m.
- CV-F7** Cet objet, déplacer dans cette direction, à 20m de l'objet orange.
- BC-F7** Déplacer cet objet de 20° à droite.
- BJ-F7** Cet objet, déplacer, 20m, à droite de cet objet.
- LS-F7** Déplacer cet objet à dr.. à 20m à droite de l'objet...De l'objet jaune.

Participants non informaticiens féminins :

- GE-F7** Deplac... Déplacer cet objet, à la place de celui-ci.
- RC2-F7** Cet objet, déplacer vers la droite, de 90°.
- AC-F7** Déplace cet objet à droite de cet objet, à 20m.
- DC-F7** Déplace cet objet à 20m sur la droite de cet objet.
- SI-F7** Ha oui, cet objet, déplacer de 20m, par ici.

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F7.**

Utilisateur / F7	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F7	Bouger	cet objet	à 20 m	sur la droite de cet objet
LV-F7	Met	cet objet,	à 20m,	de celui-ci, en direction d' la droite.
RF-F7	Déplacer,	cet objet,	à 20m	à droite, de cet objet.
RD-F7	Déplace	cet objet,	de 20m,	sur la droite par rapport à cet objet.
RR-F7	Déplace	cet objet	à 20m	à droite de cet objet
VA-F7	Déplacer	cet objet,	de 20m	à 90° de cet objet.
CF-F7	Déplace	cet objet	à 20m	à droite de cet objet.
BE-F7	Déplace	cet objet,	à 20m de cet objet,	vers la droite.
MC-F7	Déplacer	cet objet	pour l'amener à 20m	à droite de c...de cet objet.
CI-F7	Avancer	cet objet	d'une distance de 20m...	vers la droite
MM-F7	Déplace	cet objet	à 20m	à droite de cet objet
CV-F7	Déplacer	Cet objet,	à 20m de l'objet orange.	dans cette direction,
BC-F7	Déplacer	cet objet	de 20°	à droite.
BJ-F7	déplacer,	Cet objet,	20m,	à droite de cet objet
LS F7	Déplacer	cet objet	à 20m	à droite de l'objet jaune
GE-F7	Deplac... Déplacer	cet objet,		à la place de celui-ci.
RC2-F7	déplacer	Cet objet,	de 90°.	vers la droite,
AC-F7	Déplace	cet objet	à 20m.	à droite de cet objet,
DC-F7	Déplace	cet objet	à 20m	sur la droite de cet objet
SI-F7	déplacer	Ha oui, cet objet,	de 20m,	par ici.

Caractéristiques des énoncés des participants pour la tâche F7.

Utilisa - -teur	Deict	DEG	EIC	EIP	EIQ	EID	ECA	ECI	VSOO	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
LV-F7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
RD-F7	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
RR-F7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
VA-F7	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0
CF-F7	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
BE-F7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
MC-F7	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CI-F7	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	1	1
MM-F7	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0
CV-F7	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0
BC-F7	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1	1
BJ-F7	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
LS-F7	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
GE-F7	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1
RC2-F7	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	0	1	1
AC-F7	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
DC-F7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SI-F7	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	1

Composition non verbale des expressions des utilisateurs pour la tâche F7.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagnant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F7	1	0	0	0	0	1
LV-F7	1	0	0	0	1	0
RF-F7	1	0	0	0	1	0
RD-F7	1	0	0	0	1	0
RR-F7	1	0	0	0	1	0
VA-F7	1	0	0	0	1	0
CF-F7	1	1	0	0	1	1
BE-F7	1	0	0	0	1	0
MC-F7	1	0	0	0	1	1
CI-F7	1	0	0	0	1	0
MM-F7	1	0	0	0	1	0
CV-F7	1	0	1	1	0	0
BC-F7	1	0	0	0	0	0
BJ-F7	1	0	0	0	1	0
LS-F7	1	0	0	0	1	1
GE-F7	1	0	0	0	1	1
RC2-F7	1	0	0	0	0	0
AC-F7	1	0	0	0	1	0
DC-F7	1	1	0	0	1	1
SI-F7	1	0	1	1	0	0

Productions verbales des utilisateurs pour la tâche F9 de rotation.

Exemple arbitraire d'énoncé pour la tâche: « *Tourne cet objet d'un quart de tour comme ça* ».

Participants informaticiens masculins :

- GP-F9** Tourner cet objet de 90° dans cette direction.
- LV-F9** Tourne cet objet, de 90°, dans ce sens-là.
- RF-F9** Fais tourner cet objet, de 90°, comme ça.
- RD-F9** Pivote cet objet de 90°, comme ça.
- RR-F9** Tourne cet objet de 90° dans ce sens.

Participants informaticiens féminins :

- VA-F9** Faire tourner cet objet, sur lui-même, de 90°, à droite.
- CF-F9** Fais tourner cet objet de 90° vers la droite.
- BE-F9** Fais tourner cet objet de 90° vers la droite?
- MC-F9** Effectuer une rotation de cet objet de 90°, dans le sens contraire des aiguilles d'une montre.
- CI-F9** Faire tourner cette pièce d'un angle de 90° comme ça.

Participants non informaticiens masculins :

- MM-F9** Tourne cet objet Euh.. de 90° dans cette direction.
- CV-F9** Cet objet, cet objet, cet objet, faire tourner de 90°, comme ça.
- BC-F9** Tourner cet objet sur lui-même de 90°...
- BJ-F9** Cet objet, pivote de 90° sur la gauche.
- LS-F9** Faire tourner cet objet de 90° sur la gauche.

Participants non informaticiens féminins :

- GE-F9** Cet objet, faire une rotation, à 90°.
- RC2-F9** Cet objet, doit faire une rotation, de 90°....
- AC-F9** Déplace cet objet 90° en rotation sur cet axe.
- DC-F9** Déplace cet objet d'un quart de tour sur ta gauche.
- SI-F1** Heeu, alors, zut, cet...Cet objet, faire un mouvement de rotation de gauche à droite sur place..Sur son axe

**Répartition des productions verbales des utilisateurs en unité
syntaxiques pour la tâche F9.**

Utilisateur / F9	Opération	Sélection Objet d'opération	Préposition + Valeur	Direction
GP-F9	Tourner	cet objet	de 90°	dans cette direction.
LV-F9	Tourne	cet objet,	de 90°,	dans ce sens-là.
RF-F9	Fais tourner	cet objet,	de 90°,	comme ça.
RD-F9	Pivote	cet objet	de 90°,	comme ça.
RR-F9	Tourne	cet objet	de 90°	dans ce sens
VA-F9	Faire tourner	cet objet,	sur lui-même, de 90°,	à droite.
CF-F9	Fais tourner	cet objet	de 90°	vers la droite.
BE-F9	Fais tourner	cet objet	de 90°	vers la droite?
MC-F9	Effectuer une rotation	de cet objet	de 90°,	dans le sens contraire des aiguilles d'une montre.
CI-F9	Faire tourner	cette pièce	d'un angle de 90°	comme ça.
MM-F9	Tourne	cet objet	de 90°	dans cette direction
CV-F9	faire tourner	Cet objet, cet objet, cet objet,	de 90°,	comme ça.
BC-F9	Tourner	cet objet	sur lui-même de 90°...	
BJ-F9	pivote	Cet objet,	de 90°	sur la gauche
LS F9	Faire tourner	cet objet	de 90°	sur la gauche
GE-F9	faire une rotation,	Cet objet,	à 90°.	
RC2-F9	doit faire une rotation,	Cet objet,	de 90°...	
AC-F9	Déplace	cet objet	90°	en rotation sur cet axe
DC-F9	Déplace	cet objet	d'un quart de tour	sur ta gauche
SI-F9	faire un mouvement de rotation	Cet objet,		de gauche à droite sur place..Sur son axe.

Caractéristiques des énoncés des participants pour la tâche F9.

Utilisa - -teur	Deict	DEG	EIC	EIP	EIQ	EID	ECA	ECI	VS00	VSQD	ALDP	ALDI	UFPI	UFVI	UFLI	TID	NF	InS	TIn
GP-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LV-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RF-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RD-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RR-F9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VA-F9	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
CF-F9	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0
BE-F9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
MC-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CI-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MM-F9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
CV-F9	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
BC-F9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
BJ-F9	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1
LS-F9	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
GE-F9	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0
RC2-F9	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
AC-F9	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
DC-F9	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
SI-F9	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0

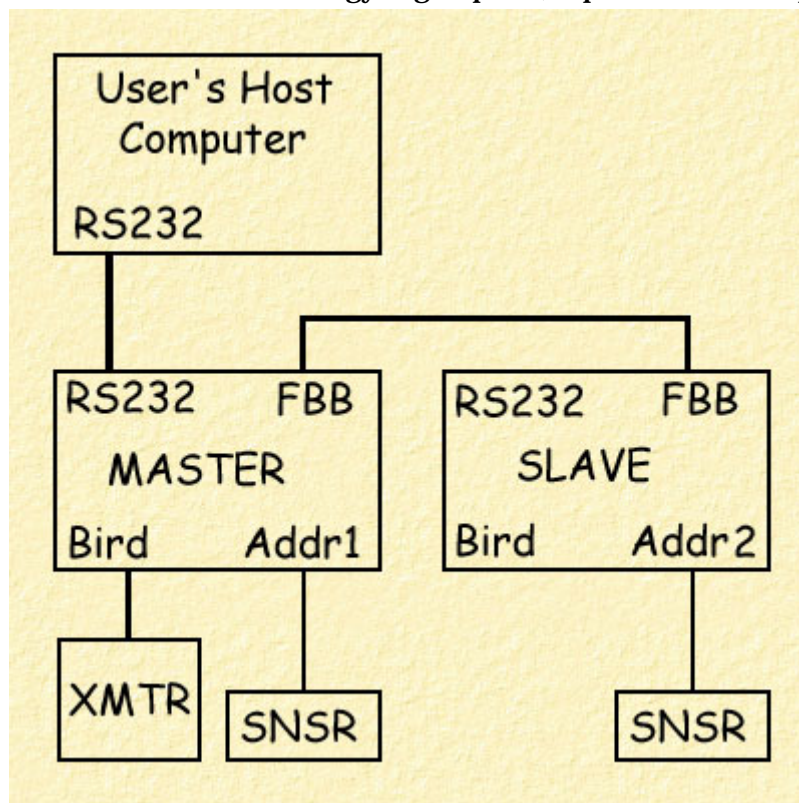
Composition non verbale des expressions des utilisateurs pour la tâche F9.

Utilisa- -teur	Indication d'objet par visée	Indication d'objet s'accompagn ant d'une Indication manuelle de direction	Indication de direction de destination par visée	Indication de direction de destination par Indication manuelle	Indication d'un second objet par visée	Indication d'un second objet par Indication manuelle
GP-F9	1	0	0	1	0	0
LV-F9	1	0	0	1	0	0
RF-F9	1	0	0	1	0	0
RD-F9	1	0	0	1	0	0
RR-F9	1	0	0	1	0	0
VA-F9	1	0	0	1	0	0
CF-F9	1	1	0	1	0	0
BE-F9	1	0	0	0	0	0
MC-F9	1	0	0	1	0	0
CI-F9	1	0	0	1	0	0
MM-F9	1	0	0	1	0	0
CV-F9	1	0	0	1	0	0
BC-F9	1	0	0	0	0	0
BJ-F9	1	0	0	1	0	0
LS-F9	1	0	0	0	0	0
GE-F9	1	0	0	1	0	0
RC2-F9	1	0	0	1	0	0
AC-F9	1	0	0	1	0	0
DC-F9	1	1	0	1	0	0
SI-F9	1	0	0	1	0	0

6 Configuration des unités "Flock of Birds" FOBs

Configuration Générale adoptée.

Documentation Ascension Technology Fig. 2 p. 9 (Reproduction adaptée).



Jumpers internes à chaque Bird. (Information générale).

Documentation Ascension Technology pp. 17

* = "Factory Default Setting"

= If the Circuit Board is Rev 6B or greater, three pin JPR 13 will not be present. in its place is a two pin JPR 17 which has a different function.

BIRD Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
9	1-2 2-3	TTL Sync Signal Enable CRT Sync Signal Enabled *
10	1-2 Aucun	FBB CTS Terminated FBB CTS not Terminated *
12	1-2 2-3	FBB Receive Data Enabled RS232 Receive Data Enabled *
13#	1-2 2-3	FBB CTS Enabled RS232 CTS Enabled *

14	1-2 Aucun	FBB Bird Data Terminated FBB Bird Data not Terminated *
15	1-2 2-3	TTL Single Sync Input Factory-Supplied CRT Sync Input Cable *
16	1-2 Aucun	FBB HOST Data Terminated FBB HOST Data not Terminated *
17#	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *

DipSwitches (Information générale).

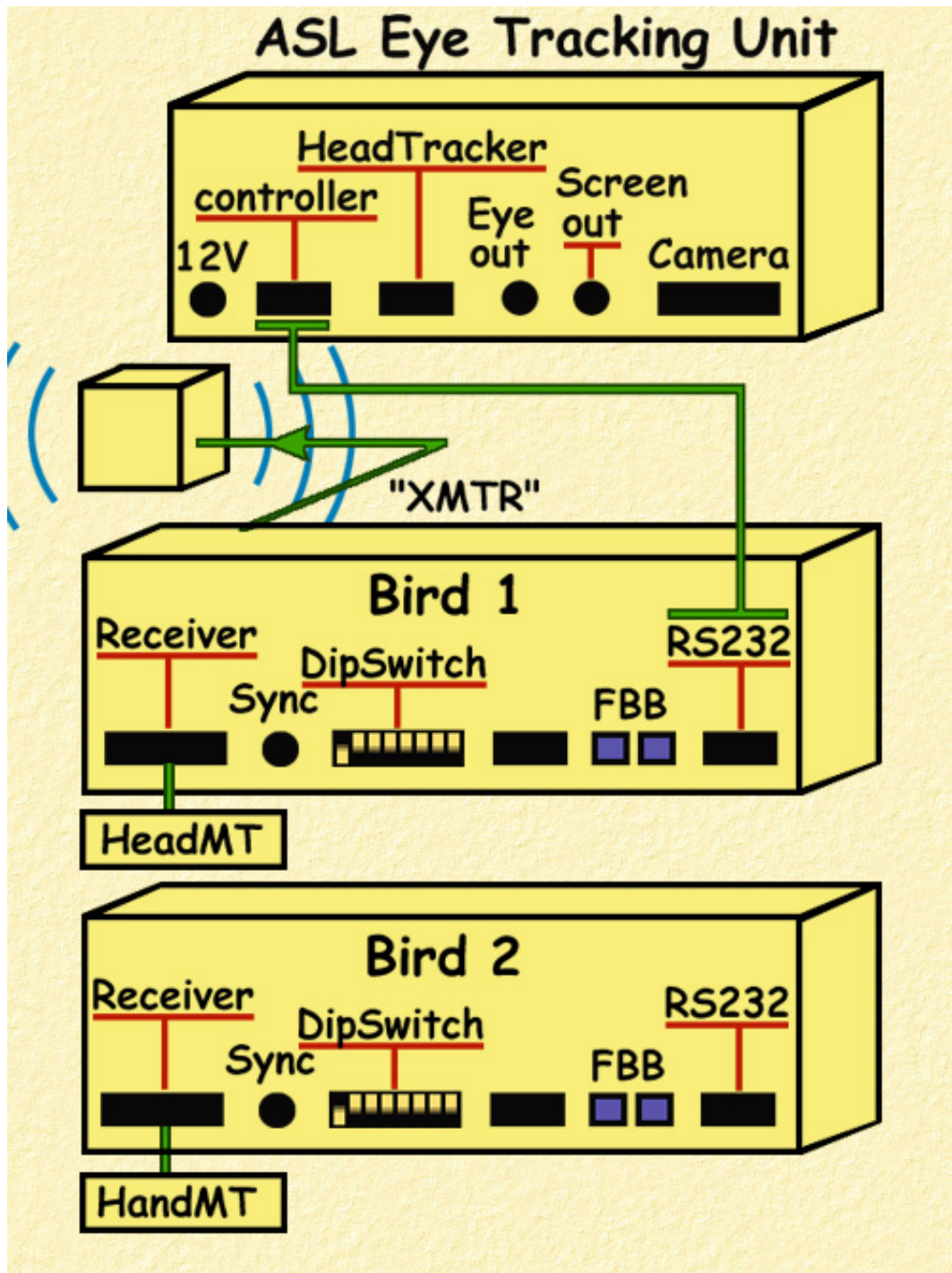
Switch	1	2	3	4	5	6	7	8
Rôle	Baud Rate			Device Address				Fly (en haut) / Test

Baud Rate	1	2	3
Not Used	Off (en haut)	Off (en haut)	Off (en haut)
2 400	Off (en haut)	Off (en haut)	On (en bas)
4 800	Off (en haut)	On (en bas)	Off (en haut)
9 600	Off (en haut)	On (en bas)	On (en bas)
19 200	On (en bas)	Off (en haut)	Off (en haut)
38 400	On (en bas)	Off (en haut)	On (en bas)
57 600	On (en bas)	On (en bas)	Off (en haut)
115 200	On (en bas)	On (en bas)	On (en bas)

Unit	4	5	6	7	Address
FOB1 Master	Off (en haut)	Off (en haut)	Off (en haut)	On (en bas)	01
FOB2 Slave Last	Off (en haut)	Off (en haut)	On (en bas)	Off (en haut)	02

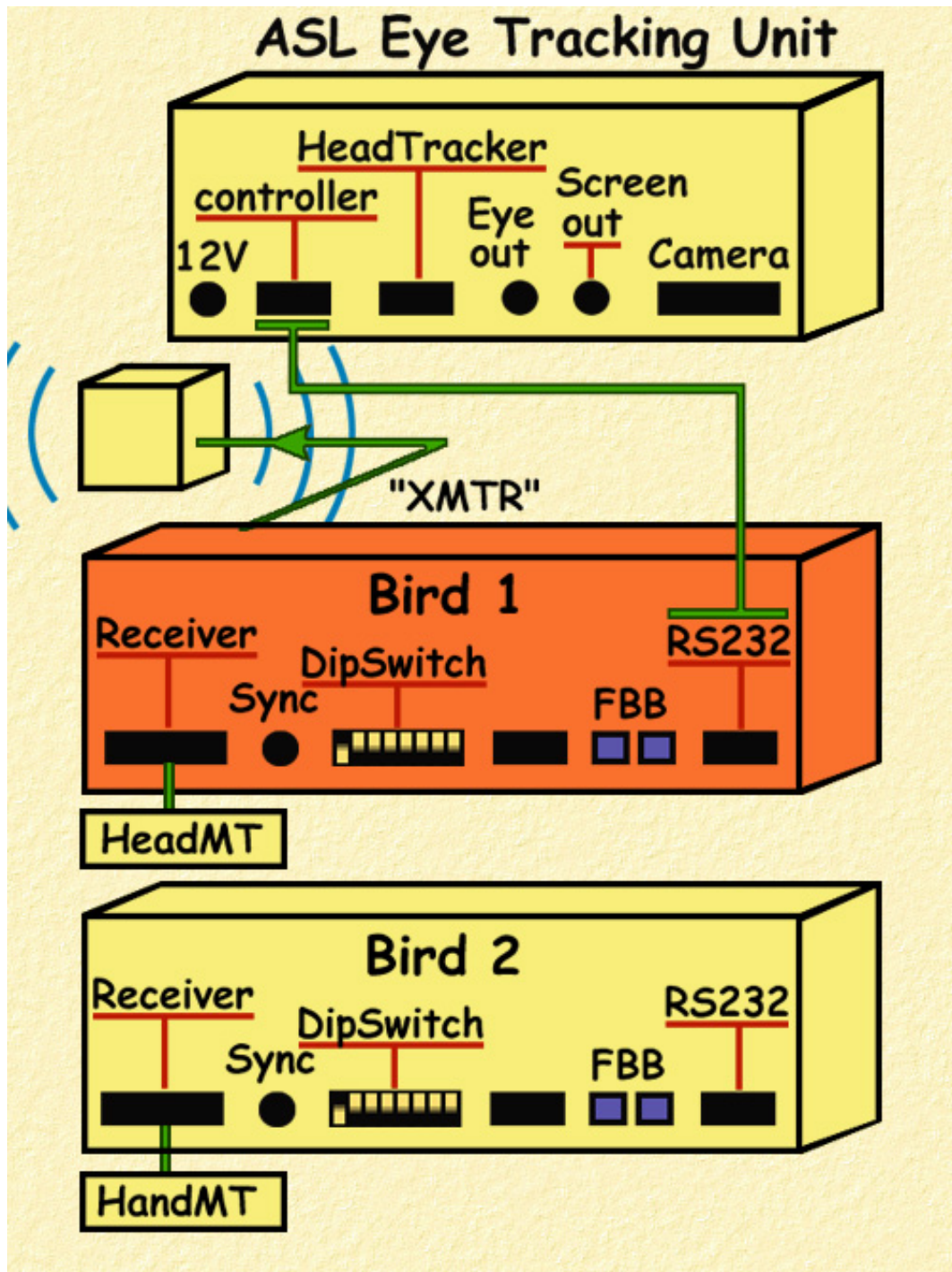
6.1 Configuration initiale des « Flock of Birds »

Configuration Générale d'origine pour le fonctionnement de l'Eye Tracking.



6.2 Configuration d'origine de la première unité

Configuration d'origine
du 1er Bird raccordé à l'unité ASL EyeTracker



BIRD Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
9	1-2 2-3 "COIL"	TTL Sync Signal Enable CRT Sync Signal Enabled *
10	1-2 Aucun	FBB CTS Terminated FBB CTS not Terminated *
12	1-2 2-3 "232"	FBB Receive Data Enabled RS232 Receive Data Enabled *
13#	1-2 2-3	FBB CTS Enabled RS232 CTS Enabled *
14	1-2 Aucun	FBB Bird Data Terminated FBB Bird Data not Terminated *
15	1-2 2-3 "COIL"	TTL Single Sync Input Factory-Supplied CRT Sync Input Cable *
16	1-2 Aucun	FBB HOST Data Terminated FBB HOST Data not Terminated *
17#	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *

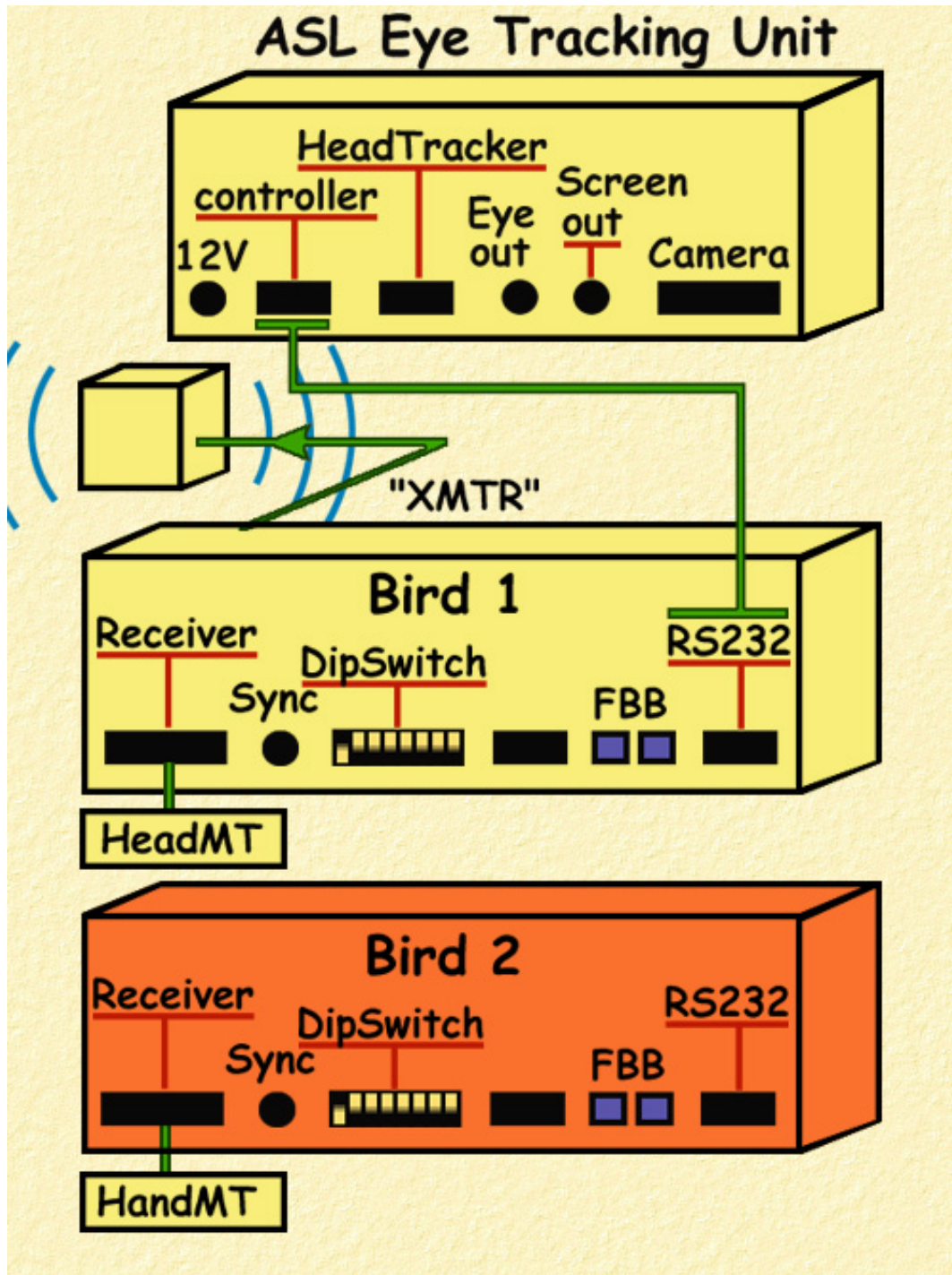
* = "Factory Default Setting"

= If the Circuit Board is Rev 6B or greater, three pin JPR 13 will not be present. in its place is a two pin JPR 17 which has a different function. **JPR17 Présent.**

DipSwiches	1	2	3	4	5	6	7	8
FOB1 Master	On (en bas)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)

6.3 Configuration d'origine de la seconde unité

Configuration d'origine
du 2nd Bird raccordé à l'unité ASL EyeTracker



Config equivalente sur un Bird en Master:

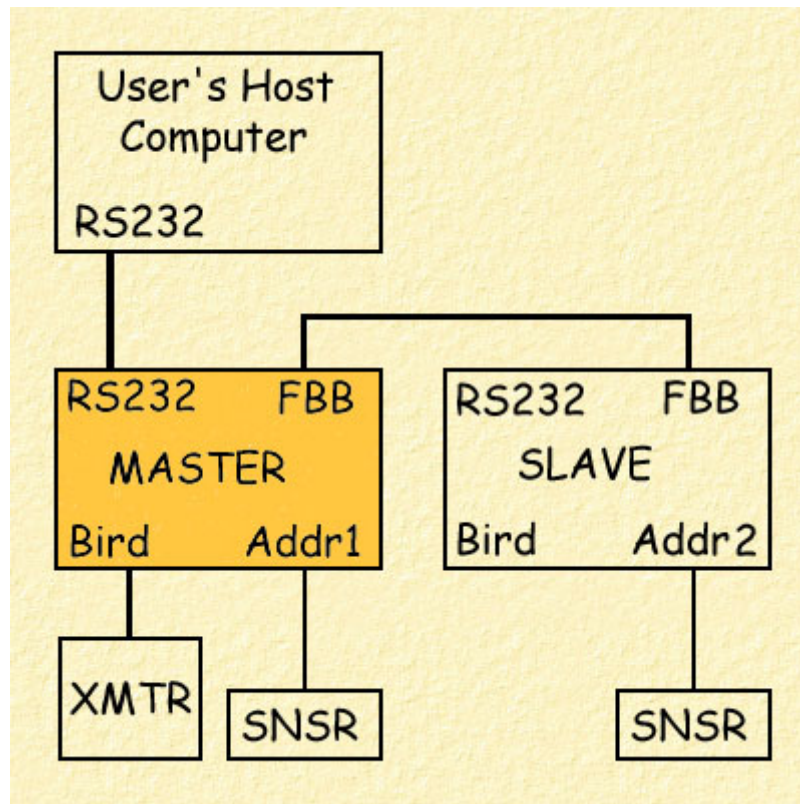
BIRD Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
9	1-2 2-3 "COIL"	TTL Sync Signal Enable CRT Sync Signal Enabled *
10	1-2 Aucun	FBB CTS Terminated FBB CTS not Terminated *
12	1-2 2-3 "232"	FBB Receive Data Enabled RS232 Receive Data Enabled *
13#	1-2 2-3	FBB CTS Enabled RS232 CTS Enabled *
14	1-2 Aucun	FBB Bird Data Terminated FBB Bird Data not Terminated *
15	1-2 2-3 "COIL"	TTL Single Sync Input Factory-Supplied CRT Sync Input Cable *
16	1-2 Aucun	FBB HOST Data Terminated FBB HOST Data not Terminated *
17#	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *

* = "Factory Default Setting"

= If the Circuit Board is Rev 6B or greater, three pin JPR 13 will not be present. in its place is a two pin JPR 17 which has a different function. **JPR17 Présent.**

DipSwiches	1	2	3	4	5	6	7	8
FOB1 Master	On (en bas)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)	Off (en haut)

6.4 Configuration d'une Unité « Bird » en « Master »



ERC Master Config:

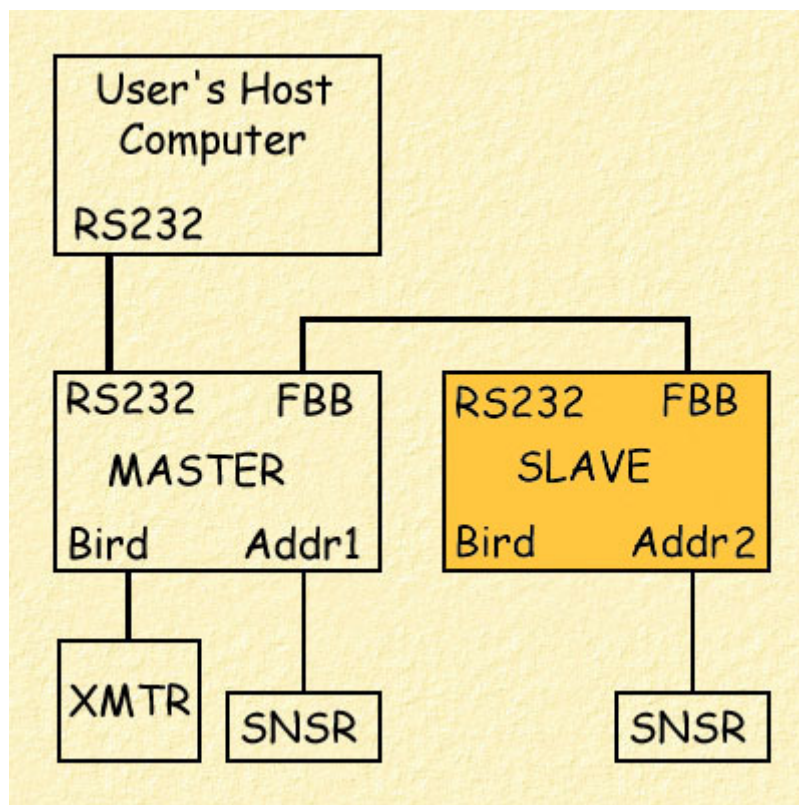
ERC Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
10	1-2 Aucun	FBB CTS Terminated * FBB CTS not Terminated
11	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *
12	1-2 2-3	RS485 Host Receive Data Enabled RS232 Host Receive Data Enabled *
14	1-2 Aucun	FBB Bird Data Terminated * FBB Bird Data not Terminated
16	1-2 Aucun	RS485 HOST Data Terminated RS485 HOST Data not Terminated*

Config equivalente sur un Bird en Master:

BIRD Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
9	1-2 2-3 "COIL"	TTL Sync Signal Enable CRT Sync Signal Enabled *
10	1-2 Aucun	FBB CTS Terminated FBB CTS not Terminated *
12	1-2 2-3 "232"	FBB Receive Data Enabled RS232 Receive Data Enabled *
13#	1-2 2-3 (pas de JPR13)	FBB CTS Enabled RS232 CTS Enabled *
14	1-2 Aucun	FBB Bird Data Terminated FBB Bird Data not Terminated *
15	1-2 2-3 "COIL"	TTL Single Sync Input Factory-Supplied CRT Sync Input Cable *
16	1-2 Aucun	FBB HOST Data Terminated FBB HOST Data not Terminated *
17#	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *

DipSwiches	1	2	3	4	5	6	7	8
FOB1 Master	On (en bas)	On (en bas)	On (en bas)	Off (en haut)	Off (en haut)	Off (en haut)	On (en bas)	Off (en haut)
		115 200 bps					Addr 1	Fly

6.5 Configuration d'une unité « Bird Slave » en fin de Bus



Second et dernier Bird en Slave sur le bus:

BIRD Jumper	Strap sur pins:	Signification
3	1-2 Aucun	FBB/RS485 Host Control RS232 Host Control *
9	1-2 2-3	TTL Sync Signal Enable CRT Sync Signal Enabled *
10	1-2 Aucun	FBB CTS Terminated FBB CTS not Terminated *
12	1-2 2-3	FBB Receive Data Enabled RS232 Receive Data Enabled *
13#	1-2 2-3 (pas de JPR13)	FBB CTS Enabled RS232 CTS Enabled *
14	1-2 Aucun	FBB Bird Data Terminated FBB Bird Data not Terminated *
15	1-2 2-3	TTL Single Sync Input Factory-Supplied CRT Sync Input Cable*
16	1-2 Aucun	FBB HOST Data Terminated FBB HOST Data not Terminated *
17#	1-2 Aucun	FBB Reset is Enabled FBB Reset is not Enabled *

DipSwiches	1	2	3	4	5	6	7	8
FOB2 Slave Last	On (en bas)	On (en bas)	On (en bas)	Off (en haut)	Off (en haut)	On (en bas)	Off (en haut)	Off (en haut)
		115 200 bps				Addr 2		Fly