



HAL
open science

Rendu multimodal en Réalité Virtuelle : Supervision des interactions au service de la tâche

Guillaume Bouyer

► **To cite this version:**

Guillaume Bouyer. Rendu multimodal en Réalité Virtuelle : Supervision des interactions au service de la tâche. Interface homme-machine [cs.HC]. Université Paris Sud - Paris XI, 2007. Français. NNT : . tel-00193064

HAL Id: tel-00193064

<https://theses.hal.science/tel-00193064>

Submitted on 30 Nov 2007

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

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

UNIVERSITÉ PARIS XI - U.F.R. DE SCIENCES

THÈSE

pour l'obtention du grade de

Docteur de l'Université Paris XI
Spécialité Informatique

Présentée par

Guillaume BOUYER

Rendu multimodal en Réalité Virtuelle :
Supervision des interactions au service de la tâche

Soutenue le 25 septembre 2007 devant le jury composé de

M.	Pierre	BOULANGER	Rapporteur
M.	Patrick	BOURDOT	Directeur de Thèse
M.	Jean-Marie	BURKHARDT	Examineur
M.	Philippe	FUCHS	Rapporteur
M.	Joseph	MARIANI	Examineur
Mme	Laurence	NIGAY	Examineur

Thèse préparée au sein du
Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
LIMSI - CNRS

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, M. Patrick Bourdot, Chargé de Recherches au CNRS et responsable de l'action transversale VENISE du LIMSI. Il m'a fait confiance pour aborder ce vaste sujet, et j'ai particulièrement apprécié les débats scientifiques animés et productifs que j'ai eu avec lui.

J'aimerais ensuite remercier les deux rapporteurs de cette thèse, M. Pierre Boulanger, Professeur à l'Université d'Alberta au Canada, et M. Philippe Fuchs, Professeur aux Mines de Paris, pour avoir accepté d'évaluer mon travail. Je remercie de même les autres membres de mon jury : M. Jean-Marie Burkhardt, Maître de Conférences à l'Université Paris 5, M. Joseph Mariani, Directeur de Recherches au LIMSI-CNRS, et Mme Laurence Nigay, Professeur à l'Université Joseph Fourier de Grenoble.

Je tiens ensuite à remercier chaleureusement mes collègues et amis des équipes VENISE, BioInfo et Geste (aka "La bande du 502 bis") pour leur aide, leurs conseils précieux mais aussi et surtout pour avoir créé autour d'eux une ambiance agréable et drôle, qui donne envie de se lever le matin pendant 4 ans. Merci donc à Nico, Antoine, Nico, Flavien, Joe, Thomas, Pilou, Bruno, Jérémie, Emilie, Rachid, Jean-Marc, Mehdi, Olivier, Christine, Annelies, Mickael, Cyril, Damien, Bob et Laurence. M. Blizzard a aussi contribué fortement à cette ambiance et à nos soirées (n'oublions pas que les jeux vidéo aident à la détente, donc à la productivité, et sont en plus un grand domaine d'application des technologies de RV !)

Je remercie également tous mes autres amis que je ne peux pas citer ici, qui n'ont pas toujours compris ce que je faisais, ni l'intérêt que cela pouvait avoir, mais qui ont quand même pris régulièrement de mes nouvelles.

Je pense aussi à ma famille, qui m'a soutenu depuis toujours, et auprès de qui j'ai pu me ressourcer quand j'en avais besoin.

Enfin, je pense fort à Aurélie qui par son amour et sa présence de tous les instants m'a permis d'aller au bout de cette étape.

Résumé

Cette thèse porte sur l'étude du rendu multimodal en Réalité Virtuelle (RV) et a pour objectif d'améliorer l'activité des utilisateurs dans les environnements virtuels. Plus particulièrement, nous cherchons à exploiter au mieux la bande passante sensori-motrice humaine et les capacités sémantiques des modalités (présentations possibles des informations sur les interfaces de communication) afin d'améliorer le confort, les performances ou la compréhension de l'opérateur dans les diverses tâches qu'il est amené à effectuer. L'ambition de nos recherches réside dans une approche générique et contextuelle du rendu multimodal en RV, c'est-à-dire qui s'éloigne des solutions adhoc et liées à des applications spécifiques. Nos travaux se veulent également orientés vers la tâche, et non vers le réalisme. Pour assister l'utilisateur dans ses interactions, nous avons voulu fonder la détermination du rendu le plus adéquat sur trois piliers : la sémantique des tâches, les caractéristiques des canaux et des modalités et le contexte des interactions.

Notre état de l'art montre tout d'abord que la combinaison intelligente de modalités selon différents canaux sensori-moteurs (visuel, audio et haptique) peut procurer des bénéfices aux utilisateurs et aux applications de RV. Les études se multiplient aussi bien dans les domaines applicatifs (apprentissage, exploration de données et assistance en environnement virtuel et augmenté) que dans les champs de la modélisation et des expériences psychophysiques. Cependant, ces études se montrent limitées par le caractère statique de la gestion de la multimodalité. C'est pourquoi nous avons eu pour objectif de concevoir un moteur de rendu multimodal générique permettant : 1) de réaliser un rendu adapté à diverses informations et tâches en environnement virtuel et 2) de tenir compte du contexte dynamique de l'exécution, lié en particulier à la succession et la simultanéité des interactions. Nous pensons également que la principale caractéristique de la RV est d'être le siège d'interactions entre un ou des humains et un environnement numérique. Cette immersion de l'humain dans son activité nous a amené à considérer l'aspect psychophysique et ergonomique de la multimodalité. C'est pourquoi nous présentons aussi une analyse ergonomique des besoins et des recommandations en terme de multimodalité dans trois applications de RV : l'exploration d'écoulements en Mécanique des Fluides, l'appariement de protéines en Bio-Informatique et la conception de formes assistée

par la RV.

Sur la base de cette analyse de fond, notre contribution majeure est l'élaboration d'un processus de supervision de la distribution multimodale dans les applications de RV. Pour cela, nous proposons un modèle générique des interactions, qui permet de représenter à l'aide d'un nombre limité d'éléments simples la sémantique des tâches et les rendus multimodaux possibles du système. Le coeur de la supervision est un moteur à base de règles logiques, capable de décider, à partir d'une base de connaissances, pour une interaction et dans un contexte donné, quel est le rendu multimodal le plus adéquat. Le contexte des interactions dont doit dépendre le choix du rendu est en partie statique, connu a priori (caractéristiques des modalités disponibles, capacités de rendu de l'application, capacités sensori-motrices du ou des utilisateurs, domaine de l'application, etc.) mais également dynamique, dépendant du moment du rendu (état courant des médias, objectifs de l'utilisateur, actions précédentes, etc.). La décision du moteur peut consister à distribuer les informations sur différents canaux sensori-moteurs, afin d'éviter la surcharge cognitive d'un canal unique ou profiter des caractéristiques perceptives de chacun. Elle peut également amener au remplacement d'une modalité par une autre, à une recherche de complémentarité ou de redondance des informations.

Notre superviseur multimodal est implémenté en langage Prolog. Puis, une méthode de construction de la base de connaissances est présentée. Enfin, l'évaluation du processus de supervision est détaillée sur des scénarii concrets d'interactions, à l'aide d'un simulateur d'application de RV. Les résultats sont satisfaisants à plusieurs titres. D'une part, le processus que nous avons élaboré se révèle fiable et les rendus décidés conformes aux recommandations ergonomiques. D'autre part, le superviseur est parfaitement adaptable à des contextes différents et dynamiques. Enfin, les décisions sont prises en temps réel, ce qui montre que le superviseur est compatible avec des interactions de RV.

Mots clefs : Réalité Virtuelle, Interaction multimodale, Superviseur de rendu, Modèle des interactions, Contexte d'interaction.

Abstract

Our research deals with multimodal rendering in Virtual Reality (VR), in order to improve the interactions between users and virtual environments. In particular, we try to make the most of human sensori-motor bandwidth and semantic capacities of visual, audio and haptic modalities, to increase comfort, performance or understanding of the operators in their different tasks. The ambition of our work lies in a generic and contextual approach of multimodal rendering in VR, in contrast with solutions tied to specific applications. Our research is also task-oriented rather than realism-oriented. Our major contribution is the design of a multimodal distribution supervision process dedicated to VR applications. It is based on a generic model of interactions, which allows representing the semantics of the tasks and the rendering possibilities of the system, with the help of a limited number of simple elements. The core of the supervision process is a rule-based engine. It is in charge of deciding what is the most appropriate multimodal rendering for a given interaction, depending on a knowledge base and on the interaction context. Evaluations are conducted on practical scenarios in the field of Fluid Mechanics exploration, Bio-Informatics protein docking and CAD. Results are satisfactory, in terms of the reliability of the decision process, its adaptability to various contexts as well as the speed of the answer, which is compatible with real-time VR interactions.

Keywords : Virtual Reality, Multimodal interaction, Rendering supervisor, Model of interactions, Interaction context.

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Table des matières	xii
Introduction générale	1
1 Problématique et objet de cette thèse	5
1.1 RV pour "Interaction Humain(s)-Environnement"	5
1.2 Améliorer le réalisme ou la tâche ?	8
1.3 Des modalités pour transmettre les informations	13
1.4 Objet de cette thèse	17
1.5 Conclusion	19
2 Etat de l'art	21
2.1 Introduction	21
2.2 Techniques de rendus en RV	21
2.2.1 Définitions	22
2.2.2 Rendu visuel	22
2.2.3 Rendu audio	24
2.2.4 Rendu et interaction haptique	26
2.2.5 Utilisation des rendus	29
2.2.5.1 Sonification de données	31
2.2.5.2 Haptisation de données	32
2.3 Applications multimodales en RV	33
2.3.1 Apprentissage, formation	33
2.3.2 Exploration de données	36
2.3.2.1 Sonification	37
2.3.2.2 Haptisation	41

2.3.3	Assistance en environnement virtuel et augmenté	43
2.3.4	Discussion	46
2.4	Modèles du rendu multimodal	48
2.4.1	Espaces de conception	48
2.4.2	Gestionnaires de rendu multimodal	55
2.4.3	Adaptativité au contexte	56
2.4.4	Discussion	58
2.5	Etudes psychophysiques	59
2.5.1	Illusions sensori-motrices	60
2.5.2	Illusions et effets multi-sensoriels	63
2.5.2.1	Visuel-Audio	63
2.5.2.2	Visuel-Haptique	64
2.5.2.3	Audio-Haptique	65
2.5.3	Charge cognitive	65
2.5.4	Discussion	66
2.6	Conclusion	67
3	Analyse d'applications de référence	69
3.1	Introduction	69
3.2	Potentialités de la multimodalité dans les applications visées	70
3.2.1	Appariement de protéines en Bio-Informatique	70
3.2.1.1	Description de l'application	70
3.2.1.2	Interactions multimodales pressenties	71
3.2.2	Exploration de cavité en Mécanique des Fluides	72
3.2.2.1	Description de l'application	72
3.2.2.2	Interactions multimodales pressenties	73
3.2.3	Conception Assistée par Ordinateur	77
3.2.3.1	Description de l'application	77
3.2.3.2	Interactions multimodales pressenties	79
3.3	Analyses ergonomiques	81
3.3.1	Méthodes d'analyse	81
3.3.2	Appariement de protéines en Bio-Informatique	83
3.3.2.1	Analyse de l'activité actuelle	83
3.3.2.2	Recommandations de conception et discussion	84
3.3.3	Exploration de cavité en Mécanique des Fluides	87
3.3.3.1	Analyse de l'activité actuelle	87
3.3.3.2	Recommandations de conception et discussion	89
3.4	Conclusion	92

4	Supervision multimodale en RV	93
4.1	Introduction	93
4.2	Modélisation des interactions de RV	95
4.2.1	Méthode et objectifs	95
4.2.2	Modèle de la tâche	97
4.2.2.1	Types de tâches	97
4.2.2.2	Paramètres des tâches	99
4.2.2.3	Données manipulées	101
4.2.3	Modèle du rendu multimodal	102
4.2.3.1	Modalités visuelles	102
4.2.3.2	Modalités auditives	104
4.2.3.3	Modalités haptiques	105
4.2.3.4	Instanciation des modalités	105
4.2.3.5	Combinaisons des modalités	106
4.2.4	Discussion et exemples	108
4.3	Processus de supervision	112
4.3.1	Principe	112
4.3.2	Choix d'implémentation	117
4.3.3	Implémentation du modèle des interactions	118
4.3.3.1	Utilisateur	118
4.3.3.2	Tâche	119
4.3.3.3	Canaux, médias et modalités	121
4.3.4	Processus de décision	121
4.3.4.1	Requête	121
4.3.4.2	Base de connaissances	123
4.3.4.3	Contexte	124
4.3.4.4	Elaboration du score	124
4.3.4.5	Résultat	126
4.3.5	Validation sur un exemple	127
4.4	Extension de la supervision	129
4.4.1	Gestion de l'espace	129
4.4.2	Gestion du temps	132
4.4.3	Extension du contexte	133
4.4.4	Gestion des retraitements	134
4.4.5	Traitement de cas particuliers	135
4.4.6	Enregistrement pour l'évaluation	136
4.5	Conclusion	136

5	Evaluation, résultats et discussion	139
5.1	Introduction	139
5.2	Evaluation de l'architecture	139
5.2.1	Editeur Prolog	140
5.2.2	Simulateur d'application	140
5.3	Construction de la base de connaissances	142
5.3.1	Corpus d'interactions	142
5.3.2	Contexte élémentaire	144
5.3.3	Constitution des listes de modalités	144
5.3.4	Affinage des pondérations	145
5.4	Evaluation de la supervision	147
5.4.1	Interactions élémentaires	147
5.4.2	Scénario de Bio-Informatique	149
5.4.2.1	Protocole	149
5.4.2.2	Résultats : contexte idéal	152
5.4.2.3	Résultats : sans interface haptique	152
5.5	Synthèse et perspectives	154
5.5.1	Principaux résultats	155
5.5.2	Choix de réalisation	155
5.5.3	Dépendance entre entrée et sortie	156
5.5.4	Respect des choix de l'utilisateur	157
5.5.5	Latences	157
5.5.6	Transitions entre rendus	158
5.5.7	Modules de rendu externes	159
5.5.8	Interactions multi-utilisateurs	159
5.6	Conclusion	160
	Conclusion et perspectives	161
	Annexe A : Glossaire des applications scientifiques analysées	167
	Annexe B : Programmation Prolog	169
	Table des figures	177
	Bibliographie	189

Introduction générale

Dans la vie de tous les jours, l'être humain utilise de manière naturelle tous les sens qu'il a à sa disposition : pour se déplacer, pour se nourrir, pour communiquer, etc. La conduite automobile est un exemple d'activité multisensorielle parmi d'autres. Lorsqu'il est au volant, l'être humain exploite sa vision pour se repérer dans l'espace, analyser son environnement, prévoir sa trajectoire et lire les informations du tableau de bord. Son audition lui sert également à percevoir son environnement, notamment en dehors de son champ de vision, à contrôler le régime de son moteur pour changer de vitesse, à détecter des ennuis mécaniques parmi les autres bruits normaux de fonctionnement (comme les clignotants), ou même à se faire une opinion de la qualité de fabrication du véhicule (e.g. bruits de portières). Son système kinesthésique lui permet d'activer les différentes commandes (pédales, volant, levier, etc.) tandis que son sens du toucher augmente la précision de ses gestes. Son sens vestibulaire l'aide à estimer ses accélérations et ses freinages. Même son odorat peut lui donner des indices sur l'état de marche de son véhicule (odeurs de brûlé).

Les technologies actuelles seraient capables de supprimer partiellement ou totalement ces différents stimuli sensoriels. Il existe par exemple des isolants suffisamment efficaces pour couper totalement l'utilisateur du bruit environnant. Pourtant, aucun constructeur automobile ne s'y risque. Certains bruits sont même complètement récréés dans l'habitacle, comme le son du clignotant. La raison en est que ces informations multisensorielles sont utiles à l'humain, et peuvent rendre plus confortables ou plus efficaces ses tâches dans l'environnement.

Les nouvelles technologies ont intégré ce besoin de multi-sensorialité pour aider les utilisateurs à manipuler les interfaces des objets ou pour rendre leurs actions plus naturelles. Par exemple, les terminaux de paiement par carte bancaire émettent souvent des bips sonores lorsque l'utilisateur rentre son code, et l'on peut remarquer que l'absence de son entraîne souvent des erreurs dans la saisie, ou du moins des doutes quant à sa validité. Dans un autre domaine, l'ajout de son voire de retour d'effort dans les jeux vidéo tend à améliorer les performances des joueurs (simulateurs de conduite, jeux de tir, etc.). Dans tous ces exemples, la combinaison réfléchie de rendus sensori-moteurs (ou multimodalité) est employée à dessein pour rendre la tâche plus facile, plus efficace, plus immersive, plus naturelle.

Les applications de Réalité Virtuelle (RV) offrent aux utilisateurs humains d'interagir comme dans la vie de tous les jours, mais avec des environnements numériques artificiels. Ils peuvent y effectuer des tâches variées, dans des domaines allant du loisir à la production industrielle, en passant par la recherche scientifique. Avec les technologies courantes, les communications d'informations du système vers l'humain peuvent être réalisées selon trois canaux principaux : la vision, l'audition et l'haptique. Les applications exploitent ces canaux à leur guise, en fonction de leurs objectifs, des utilisateurs, des moyens matériels, etc.

Dans ce contexte, notre travail porte sur l'étude du rendu multimodal en RV, afin d'améliorer l'activité des utilisateurs. Plus particulièrement, nous cherchons à perfectionner la transmission des informations issues des environnements virtuels (EV) en les distribuant sur les différents canaux sensori-moteurs des utilisateurs et en exploitant les multiples présentations possibles de ces informations (appelées "modalités"). Pour cela, nous voulons concevoir un "superviseur multimodal", chargé de la gestion intelligente du rendu, qui prenne en compte la sémantique des informations, les capacités de rendu et le contexte de l'interaction. L'aspect novateur de nos travaux est également que ce superviseur sera générique, et adaptable à différentes applications de RV.

Organisation du mémoire

Dans le **premier chapitre**, nous mettons en place la problématique de nos travaux au travers d'une définition détaillée de la RV et d'une comparaison avec le domaine de l'Interaction Homme-Machine (IHM). Nous étudions en particulier la place de l'utilisateur humain au sein du système informatique. Puis nous discutons du concept de multimodalité et de son apport potentiel pour aider l'être humain dans ses activités. Pour finir, nous précisons le cadre et l'objet de cette thèse.

Dans le **second chapitre**, nous présentons tout d'abord les dispositifs de RV ainsi que les techniques d'interactions associées aux canaux visuel, audio et haptique. Puis nous illustrons à l'aide d'exemples trois catégories d'applications de RV faisant appel à la multimodalité en sortie. Ensuite nous examinons des espaces de conception et des questionnaires multimodaux, principalement en IHM. Enfin nous nous intéressons aux expériences psychophysiques portant notamment sur la combinaison des modalités. Cet état de l'art nous permet de mettre en évidence les différents éléments capitaux à prendre en compte et à modéliser dans notre superviseur de rendu multimodal.

Le **troisième chapitre** présente en détail trois applications qui pourraient potentiellement bénéficier de rendu multimodal : l'exploration d'écoulements en Mécanique des Fluides, l'appariement de protéines en Bio-Informatique et la conception de formes assistée par la RV. Dans un premier temps, nous décrivons les problématiques

multimodales des trois domaines et nous imaginons des interactions multimodales utiles. Puis, dans le cadre du projet CoRSAIRE, les deux premières applications font l'objet d'analyses ergonomiques : les tâches et les besoins des utilisateurs sont identifiés et des recommandations multimodales sont formulées. Ces deux analyses détaillées, complétées d'une expertise plus pratique en CAO, doivent nous servir à bâtir un modèle des interactions qui soit générique et représentatif des activités réelles des utilisateurs, et donc un superviseur multimodal utile et concrètement utilisé.

Puis, nous présentons notre contribution au problème de la gestion des rendus multimodaux en RV dans le **quatrième chapitre**. Il s'agit d'un moteur de supervision chargé de coopérer avec l'application et l'utilisateur pour contrôler les interactions et leurs rendus multimodaux. Nous rappelons tout d'abord notre approche du problème. Puis nous présentons notre modèle des interactions, qui va servir de format de communication entre tous les acteurs du rendu. Nous décrivons ensuite en détail l'architecture de la supervision, le fonctionnement du superviseur et son implémentation. Nous évaluons enfin ce superviseur sur des interactions analysées dans les applications du chapitre 3.

Enfin, dans le **cinquième chapitre**, nous implémentons notre superviseur et nous réalisons un simulateur d'application de RV afin de tester le processus de supervision. Puis, nous élaborons une méthode de construction de la base de connaissances. Cela nous permet alors d'évaluer notre supervision multimodale sur des tâches concrètes issues des applications du chapitre 3, et notamment sur un scénario interactif complet de Bio-Informatique. Les résultats sont analysés, puis nous discutons de problèmes multimodaux spécifiques soulevés au cours des recherches.

Chapitre 1

Problématique et objet de cette thèse

Résumé. *Dans ce chapitre, nous mettons en place la problématique de nos travaux au travers d'une définition détaillée de la RV et d'une comparaison avec le domaine de l'Interaction Homme-Machine (IHM). Nous étudions en particulier la place de l'utilisateur humain au sein du système informatique. Puis nous discutons du concept de multimodalité et de son apport potentiel pour aider l'être humain dans ses activités. Pour finir, nous précisons le cadre et l'objet de cette thèse.*

1.1 RV pour "Interaction Humain(s)-Environnement"

La Réalité Virtuelle (RV) est une discipline qui permet à un ou plusieurs utilisateurs humains de percevoir et d'agir de manière immersive, pseudo-naturelle et temps réel sur un environnement numérique, appelé environnement virtuel (EV). Cet EV peut être une copie du réel, mais également une simulation de certains de ses aspects, une représentation symbolique d'un concept ou d'un phénomène, ou encore un monde totalement imaginaire [Fuchs et al., 2006a].

Cette définition montre que l'interaction est un aspect essentiel de la RV, et qu'il est nécessaire de détailler chacun de ses aspects. En pratique, elle est réalisée par l'intermédiaire d'interfaces, qui sont les outils de communications entre l'homme et le système. On distingue schématiquement les interfaces d'entrée (commande de l'homme au système) et les interfaces de sortie (réponse du système vers l'homme).

Tout d'abord, l'interaction doit être "immersive". Le concept d'immersion est encore largement discuté [Burkhardt et al., 2003]. D'un point de vue sensoriel, concret, il s'agit d'occulter partiellement ou totalement la perception de l'humain des objets réels qui l'entourent pour qu'elle soit uniquement ou principalement exposée aux stimuli du monde virtuel. Dans le cas le plus abouti de l'immersion virtuelle, l'humain est totalement découplé du monde réel : seul l'environnement virtuel l'occupe, c'est la notion de "présence". Mais l'immersion peut également être mentale, l'utilisateur

pouvant être littéralement "plongé", absorbé par son activité, occultant de lui-même le monde réel (comme il pourrait l'être par exemple pendant la vision d'un film ou la lecture d'un livre). L'immersion de l'utilisateur dans l'EV est donc créée à la fois par des moyens techniques, sensoriels, comme l'obscurité, de grands écrans et des enceintes diffusant des images en relief et du son 3d, des interfaces de commande non intrusives et naturelles, etc.; et des procédés logiciels, basés sur des connaissances psychologiques : les concepteurs jouent sur le rendu de la scène, les habitudes de l'utilisateur, ses émotions, etc.

Ensuite, l'interaction doit être pseudo-naturelle, c'est-à-dire proche de l'interaction de l'humain dans son environnement réel. En effet, dans le monde réel, l'être humain a acquis, assimilé un certain nombre de "schèmes", de comportements à la fois sensori-moteurs voire cognitifs. Il possède ses propres raisonnements assortis de représentations mentales, a construit une organisation du monde selon un ensemble de règles spatio-temporelles et causales, a acquis des réflexes, a appris des gestes, etc. Dans un EV, il va être tenté de répéter ces schèmes ou de les adapter au nouveau contenu qui s'offre à lui. L'une des caractéristiques des travaux sur l'interaction en RV est de faciliter cette transposition, pour que l'interaction se fasse inconsciemment et sans grand effort mental : "les interfaces exploitent la motricité ou les perceptions de l'homme issues de son comportement dans le monde réel" [Fuchs et al., 2006a].

Le schème transposé dans l'EV est cependant biaisé par rapport au schème réel, notamment en raison des interfaces de communication. Dans les cas où les schèmes ne peuvent pas être exploités (difficultés techniques, économiques ou théoriques) ou bien lorsque l'on ne veut pas les utiliser car ils ne se révèlent pas pertinents, on utilise des *métaphores*, des interactions symboliques : "il s'agit de communiquer à l'utilisateur les règles pour interagir avec l'EV au moyen d'un concept ou d'un objet connu qui permet de l'expliquer" [Fuchs et al., 2006c]. Une métaphore peut être par exemple une substitution sensorielle ou motrice : une information habituellement transmise par l'intermédiaire d'un canal sensoriel est cette fois communiquée selon un autre canal. Un exemple de substitution dans la vie de tous les jours serait un téléphone portable en mode "vibreur". Les métaphores demandent en général plus d'efforts cognitifs. L'emploi de métaphore modifie également le rapport de l'utilisateur à la scène, et change donc le domaine d'immersion qui passe du sensoriel au cognitif, des sens à l'activité, à la tâche. Ceci est particulièrement vrai pour une application où l'on interagit avec des données abstraites. Tout d'abord, les données étant abstraites, il n'existe pas de schème sensori-moteur réel. L'humain peut par contre avoir l'habitude d'utiliser certaines représentations symboliques, mathématiques, statistiques, etc. qui sont autant de métaphores, la plupart du temps visuelles. Elles peuvent également être substituées par des métaphores acoustiques ou haptiques, par exemple pour que des aveugles puissent explorer des graphes.

Enfin, l'interaction en RV se doit d'être en "temps réel" ou du moins en "temps interactif", c'est-à-dire respectivement que l'humain ne doit pas percevoir de latence

entre ses actions et la perception du résultat de ses actions dans l'EV, ou que ces latences ne perturbent pas son activité. En effet, l'utilisateur d'une application de RV peut accepter un certain délai entre une commande et le rendu visuel de cette commande, mais ce délai doit être faible et constant, il ne doit pas varier dans le temps de manière aléatoire. Il est courant dans la vie de tous les jours d'être confronté à des ralentissements, des attentes face à un système informatique. Il arrive par exemple que l'ordinateur familial ralentisse et qu'il réponde moins vite aux clics de souris, ou bien que le réseau soit surchargé pendant un jeu en ligne et que les déplacements de son personnage soient saccadés et nous nous en accommodons. Cependant, en RV, du fait de l'immersion, les conséquences de latences non contrôlées peuvent provoquer un inconfort plus gênant, des erreurs de manipulation, et même des malaises appelés "mal des EV" ("cybersickness").

On le voit, les recherches dans le domaine de l'interaction entre l'homme et les EV sont donc au coeur des recherches en RV. On peut d'ailleurs dire que la RV n'est qu'un contexte et une application particuliers (et même dérivés) des recherches dans le domaine plus global de l'Interaction Homme-Machine (IHM), qui vont des études des caractéristiques physiologiques et cognitives humaines à l'élaboration de nouvelles technologies (électronique, robotique, imagerie, acoustique, etc.) en passant par des recherches en informatique (architecture, systèmes et algorithmes de calcul, synthèse et traitement d'image, etc.) et des recherches théoriques (modélisation des interactions). En dehors de la RV, les recherches en IHM portent également sur des disciplines aussi variées que la téléopération, que ce soit pour la télémanipulation de dispositifs dans des environnements à risque (par exemple nucléaire) ou le travail à distance collaboratif [Khézami, 2005]; les dispositifs mobiles, par exemple pour la prise de notes efficace sur PDA [Lumsden and Gammell, 2004]; la domotique qui est à son commencement, l'informatique embarquée dans des véhicules, ce qui pose des problèmes particuliers de criticité [Kamp et al., 2004], etc. Mais l'IHM porte surtout, depuis l'origine, sur les systèmes informatiques de type PC.

Ce qui fait la spécificité de la RV par rapport aux autres dimensions de l'IHM est essentiellement son caractère anthropo-centré. Il ne s'agit pas seulement de permettre la communication d'informations entre un utilisateur et un système informatique évolué, mais d'immerger un humain dans un environnement, dans lequel il a des activités sensori-motrices. L'utilisateur est au coeur de l'interaction, en interaction directe et temps réel avec l'EV. Il vaut dès lors mieux parler d'"Interaction Humain-Environnement" que d'"Interaction Homme-Machine".

L'humain doit donc être au coeur de toutes les préoccupations des concepteurs de systèmes de RV. Ceux-ci doivent le prendre en compte au niveau physiologique, psychophysique et cognitif. Les interfaces doivent respecter l'ensemble des caractéristiques humaines pour éviter toute perturbation, au risque de soumettre les utilisateurs

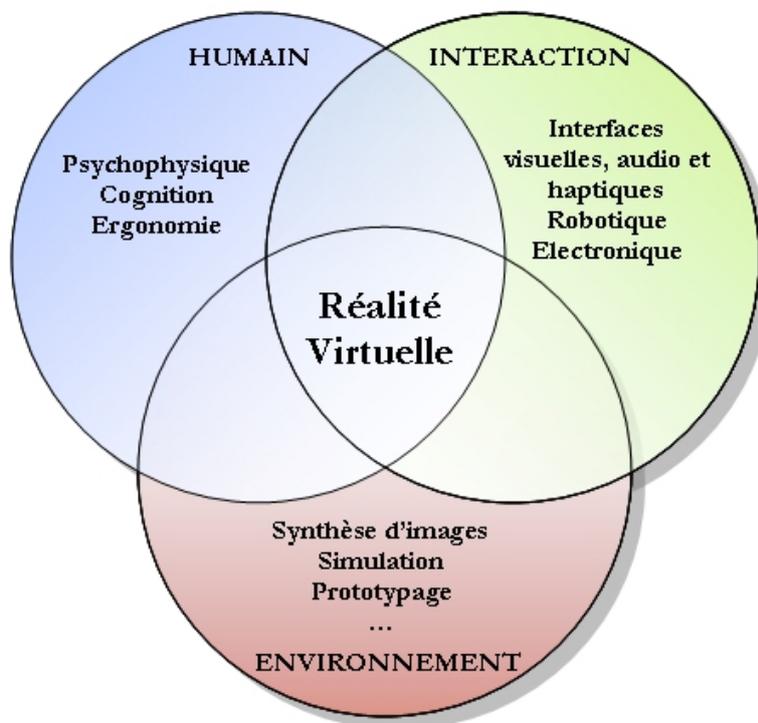


FIG. 1.1 – La Réalité Virtuelle : des Interactions Humain - Environnement

teurs au fameux "mal des EV". Ils doivent également en tenir compte au niveau de la gestion informatique de l'EV : l'utilisateur doit avoir sa représentation informatique (prise en compte de sa position, de ses actions, etc.). Enfin, il faut que les systèmes de RV tiennent compte de l'activité de l'humain, de sa tâche, de ses objectifs.

Mais justement, quels sont les activités humaines dans une application de RV ? Il nous faut identifier les différents domaines d'applications, et dans ceux-ci les usages, les objectifs, les fonctionnalités, etc. qui y sont implémentés ou nécessaires.

1.2 Améliorer le réalisme ou la tâche ?

Une vision un peu caricaturale de la RV nous permet de distinguer deux approches, deux cadres généraux qui pilotent les recherches.

D'un côté, certaines recherches défendent l'idée selon laquelle la RV doit être la plus proche possible de la réalité : c'est *l'approche orientée perception*. Elles tentent de rendre chaque retour sensoriel le plus réaliste possible, afin de faire ressentir à

l'utilisateur l'environnement virtuel "comme" le monde réel, de lui fournir une interaction naturelle avec lui, de lui faire retrouver ses schèmes sensoriels habituels. Or, comme les humains perçoivent le monde réel à l'aide de tous leurs canaux perceptifs, un des objectifs des systèmes de RV réalistes est d'une part de tous les stimuler à l'aide de dispositifs sensori-moteurs, d'autre part que ces dispositifs soient les plus proches possibles des caractéristiques humaines.

Certains domaines d'application sont très demandeurs d'environnement virtuels réalistes, où les rendus audio et haptiques complètent le rendu visuel. Cela peut être un paramètre critique, comme par exemple les systèmes virtuels d'entraînement médical pour les apprentis chirurgiens, ou les simulateurs de conduite ou de vol. Cela peut être également un besoin économique, par exemple pour le prototypage virtuel, où le produit virtuel doit être "ressenti" comme le produit final (figure 1.2). L'industrie des jeux vidéo est également friande de réalisme et d'immersion, mais cette fois dans un objectif ludique (figure 1.3).

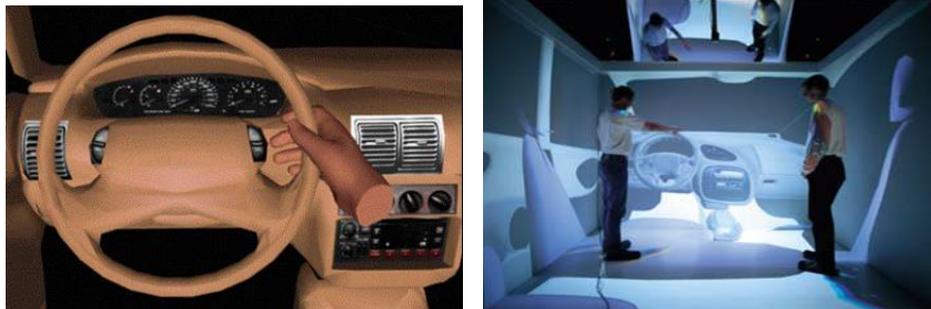


FIG. 1.2 – Approche Orientée Perception : design intérieur d'une automobile.

Dans le domaine de la RV, cette approche conduit à l'amélioration continue de tous les moyens technologiques utilisés dans les installations de RV : dispositifs visuels, périphériques d'interaction, ressources calculatoires, algorithmes, etc.

Cependant, cette recherche de réalisme en RV semble utopique. D'une part, les technologies sont limitées ou inadaptées par rapport aux caractéristiques humaines. De plus, la perception humaine a elle-même des limites par rapport à ce qu'est capable de créer un système numérique (on ne peut percevoir qu'un certain nombre de couleurs et de sons, on ne peut traiter qu'un certain nombre d'informations en même temps, etc.) ou bien peut être altérée (handicaps, vieillesse). Ces deux aspects compromettent la transposition des schèmes du monde réel.

Mais cette transposition peut être également impossible ou inadaptée. En effet, les EV peuvent présenter des informations denses ou complexes, parfois non réalistes (données abstraites, etc.), et si l'on veut améliorer la compréhension de

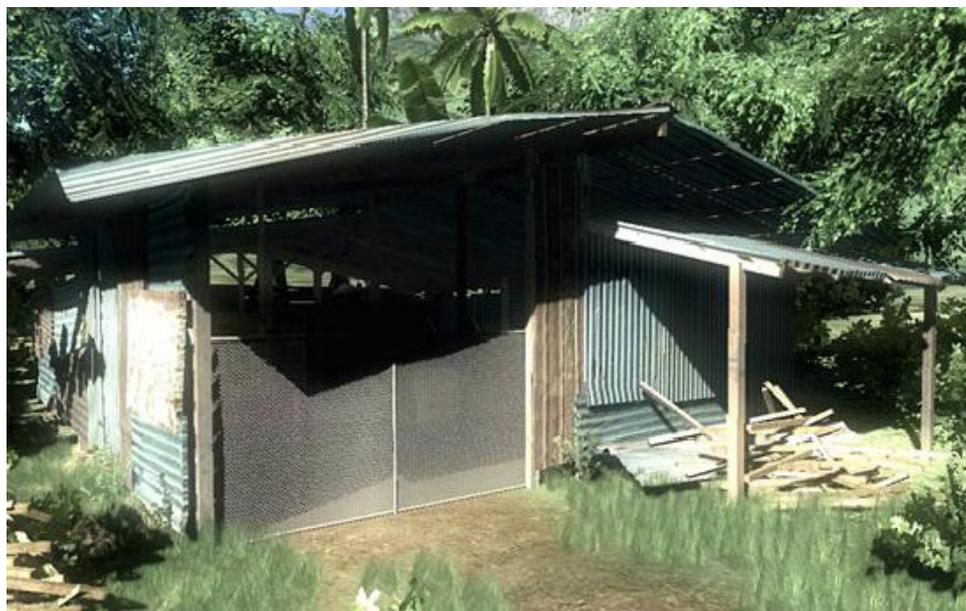


FIG. 1.3 – Approche Orientée Perception : rendu visuel extrêmement réaliste du jeu video Crysis (Crytek).

l'humain qui interagit, ou son efficacité, son confort durant la tâche, il peut être intéressant de changer ses schèmes. Plus globalement, certaines recherches estiment que nous devrions tirer partie de l'extraordinaire potentiel de la RV pour essayer de diversifier l'utilisation des applications et d'améliorer la performance des utilisateurs dans les EV : c'est *l'approche orientée tâche*. Les possibilités quasi infinies de la RV peuvent être exploitées pour créer des mondes, des applications et des interactions qui précisément n'existent pas dans la réalité. Les chercheurs tentent d'exploiter au mieux les qualités et les limitations des rendus disponibles pour créer des interactions nouvelles, efficaces et pertinentes, afin d'améliorer le sens de la communication homme-machine.

L'exploration de données est l'une des applications les plus favorables à cette approche (figures 1.4 et 1.5). La visualisation graphique est largement utilisée pour explorer des données multidimensionnelles, mais elle est limitée pour l'analyse et la compréhension de certains types de données scientifiques. C'est pourquoi d'autres canaux sensoriels sont testés pour substituer ou compléter la vision. On peut par exemple essayer de représenter les données dans le domaine acoustique : au lieu de présenter la variation d'un indice boursier sous forme de graphe, on peut utiliser un son dont on fait varier la fréquence en fonction de l'indice. Un autre domaine d'application important est la Réalité Augmentée (RA). Par définition, le but de la RA n'est pas la poursuite du réalisme puisque l'on cherche à "augmenter" le réel en y ajoutant des informations virtuelles.

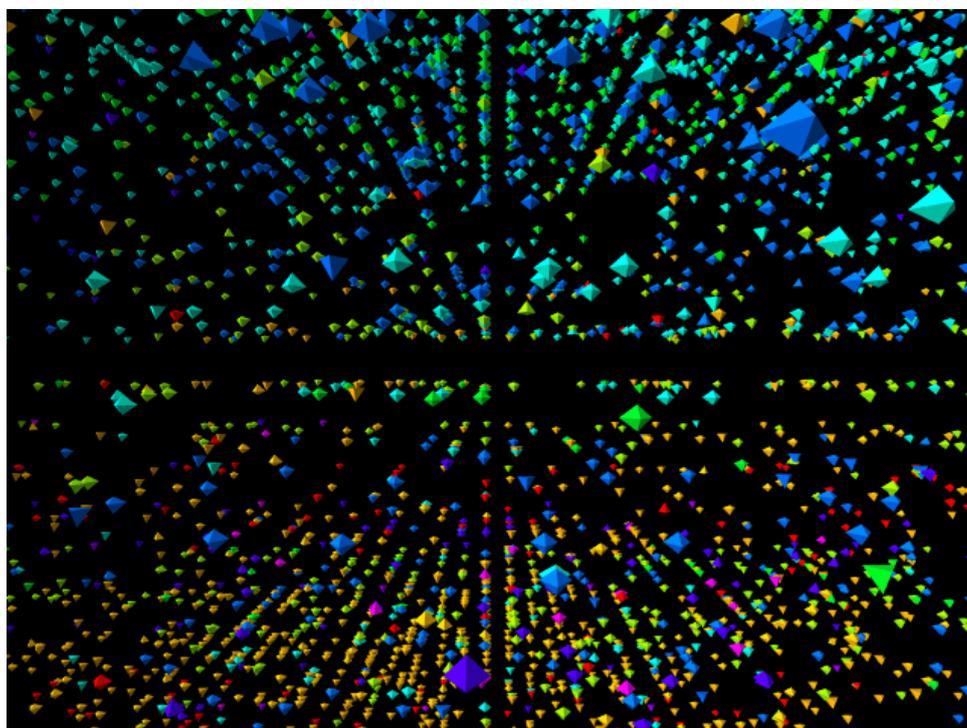


FIG. 1.4 – Approche Orientée Tâche : exploration de données pour le projet 3D Visual Data Mining (Aalborg University).

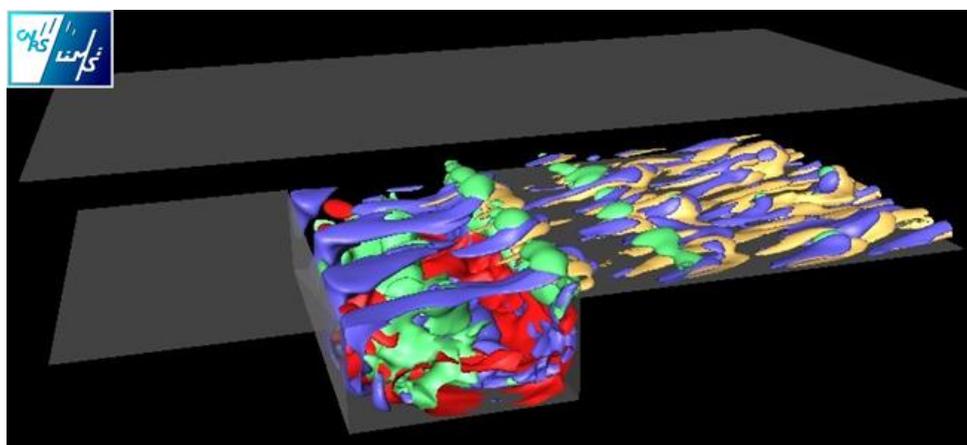


FIG. 1.5 – Approche Orientée Tâche : exploration d'un écoulement en Mécanique des Fluides (LIMSI-CNRS).

Cette approche orientée tâche est plus propice aux connaissances psychophysiques et aux progrès logiciels et théoriques et qu'aux avancées technologiques. On cherche à utiliser au mieux les interfaces existantes dans l'intérêt de l'utilisateur

face à sa tâche : pour pouvoir contourner les limites technologiques, il faut d'abord connaître en profondeur le fonctionnement psychophysique et cognitif humain, évaluer l'influence réciproque des canaux sensoriels, puis proposer des modèles d'interaction qui respectent et exploitent ce fonctionnement humain, et enfin les intégrer dans des applications et les évaluer.

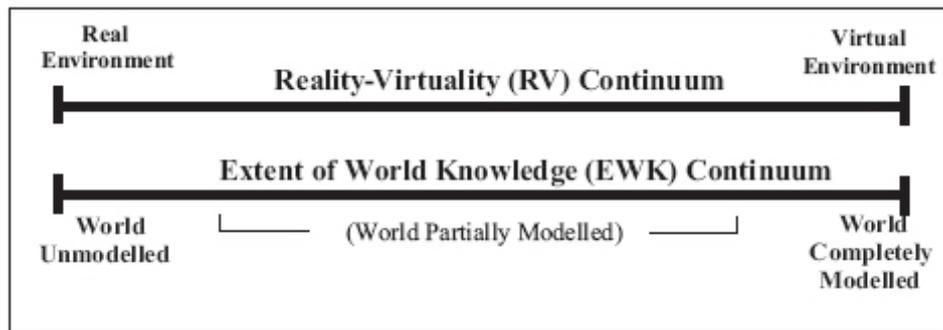


FIG. 1.6 – Continuum Réel-Virtuel de Milgram.

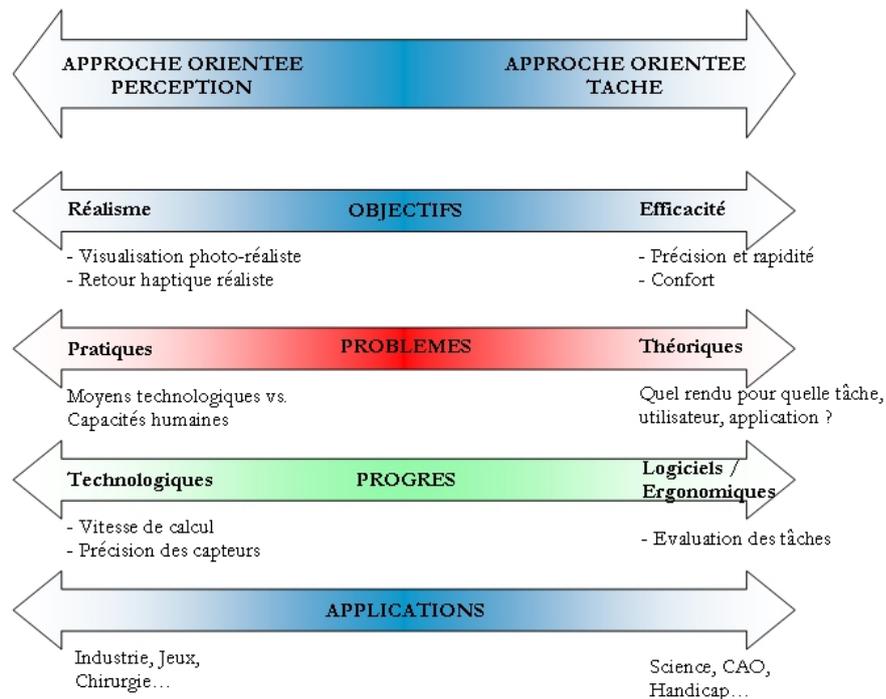


FIG. 1.7 – Deux approches de la RV : un Continuum Perception-Tâche inspiré de Milgram

La distinction entre les deux approches présentée n'est cependant pas aussi claire. Souvent, les applications de RV combinent les deux approches : elles peuvent simuler certains aspects du réel, tout en acceptant des rendus non réalistes (symboliques ou imaginaires) pour d'autres. En référence au continuum réel-virtuel de Milgram (figure 1.6) [Milgram et al., 1995], on peut établir un "continuum perception-tâche" (figure 1.7) qui relie et résume les deux approches que nous avons décrites. Ce continuum est composé de quatre curseurs (objectifs principaux, problèmes rencontrés, progrès attendus et domaine d'application concerné) indiquant le degré de réalisme ou d'efficacité recherché. Ce continuum a également une explication historique. En effet, dans le domaine de l'interaction pour les ordinateurs de bureau, au début des années 90, les experts se sont rendus compte que les ordinateurs étaient dotés de périphériques "classiques" (clavier, souris, écran) qui d'une part n'exploitaient pas toutes les capacités des différents canaux sensori-moteurs humains et l'usage simultané de ces canaux, d'autre part ne tenaient pas compte des compétences humaines de traitement des informations. En conséquence, les systèmes IHM ont commencé à intégrer de nouvelles interfaces (parole, geste, audio, etc.), dans une approche orientée perception. Ces systèmes ont été qualifiés de "multimédia" ou de "multisensoriels" dans le cas de la RV. Ensuite, ces nouveaux moyens d'interaction ont été utilisés dans une approche orientée tâche pour améliorer la rapidité, l'efficacité et le confort de l'utilisateur (mais parfois au détriment de la complexité et des coûts). On parle désormais de systèmes "multimodaux". La partie suivante explicite ces termes dans le domaine de la RV.

1.3 Des modalités pour transmettre les informations

Parmi toutes les études dans le domaine de l'IHM, de nombreuses définitions et usages des mots "modalité" et "multimodalité" sont employés. Ces emplois sont souvent déterminés par les objectifs des auteurs (ergonomes, informaticiens, etc.). Nigay [Nigay and Coutaz, 1996] définit une modalité comme un couple composé d'un dispositif physique et d'un système représentationnel, c'est-à-dire un système conventionnel structuré de signes assurant une fonction de communication. Une modalité associe un type de canal de communication utilisé pour transmettre ou acquérir de l'information, et un type de données échangées. La multimodalité est alors la capacité d'un système à communiquer avec un utilisateur par le biais de différents types de canaux et à extraire et transmettre automatiquement du sens. Un système multimodal est capable de modéliser automatiquement le contenu de l'information à un haut niveau d'abstraction. Pour Buxton, cité dans [MacDonald and Vince, 1994], cette capacité à transmettre du sens est la différence entre la multimodalité et le multimédia, qui "se concentre sur le média ou la technologie plutôt que sur l'application ou l'utilisateur". Schomaker [Schomaker et al., 1995] est du même avis et ajoute que

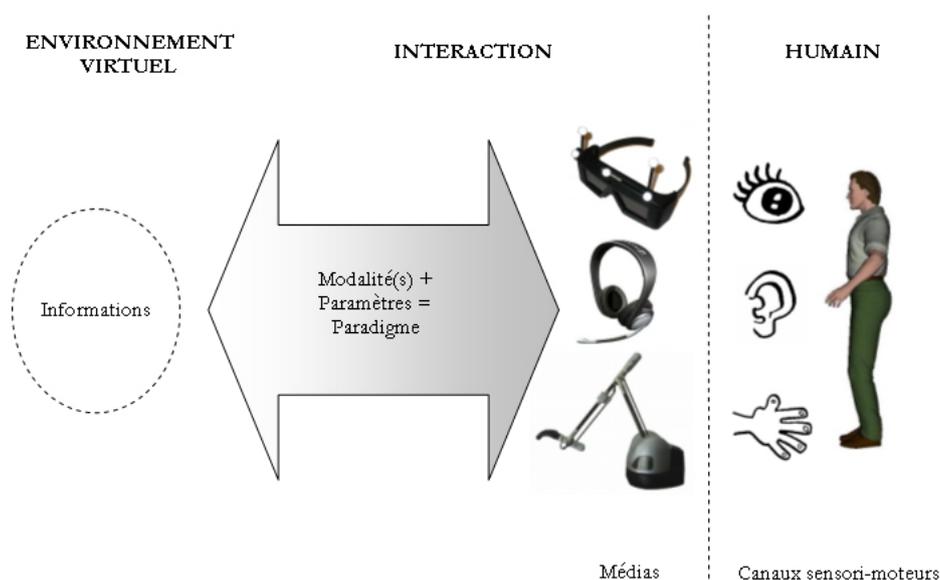


FIG. 1.8 – Taxonomie des éléments de l'interaction en RV

la multimodalité essaye "d'améliorer le débit de la communication homme-machine". Enfin, Bellik [Bellik, 1995] définit une modalité comme la structure de l'information perçue par l'opérateur humain (et non la structure utilisée par la machine pour représenter l'information). Nos travaux se fondent sur des définitions assez proches des auteurs évoqués ci-dessus, et mettent l'accent sur certains points qui nous semblent essentiels.

Comme nous l'avons vu dans les sections précédentes, une application de RV est le siège d'interactions entre des humains et un environnement, qui sont autant d'échanges d'informations (figure 1.8). La communication a lieu dans les deux sens. On distingue donc les moyens utilisés par l'humain pour agir ou commander l'environnement, que nous qualifierons d'entrées (vis à vis du système informatique), et ceux utilisés par l'EV pour transmettre les informations à l'humain, que nous qualifierons de sorties. Chacun des éléments que nous définissons ci-dessous peuvent être concernés par cette distinction.

Tout d'abord, un être humain possède des *canaux sensori-moteurs* qui lui permettent de ressentir et d'agir sur le monde extérieur : vision, audition, récepteurs kinesthésiques et tactiles, etc. Ces canaux ont des caractéristiques physiologiques précises qui peuvent cependant varier d'un individu à un autre. Remarquons que ces canaux sont appelés "modes" chez Bellik [Bellik, 1995] et Rousseau [Rousseau, 2006]

Les *médias* sont les dispositifs technologiques correspondants à ces canaux :

écrans, projecteurs, enceintes, bras à retour d'effort, micro, caméras, etc. Un système de RV peut intégrer un certain nombre de ces médias, également appelés périphériques ou interfaces, et s'en servir comme supports techniques de communication entre l'utilisateur et l'EV.

Les *modalités* sont les différents moyens de transmettre ou acquérir une information à travers les médias, ce qui correspond au deuxième membre du couple défini par Nigay. Des modalités en sortie sont, par exemple, l'affichage de texte, de formes, de couleurs, etc. sur un écran, ou la diffusion de musique, de voix, de sons iconiques, etc. à travers un casque audio. En entrée, l'utilisateur pourra utiliser la parole, le geste ou un pointeur laser (voir figure 1.9). Une modalité possède intrinsèquement une dimension sémantique en plus de sa dimension sensorielle : elle est perçue par l'humain, mais elle est également porteuse de sens, elle est traitée et comprise. C'est pourquoi toutes les modalités ne sont pas adaptées à la transmission d'une information donnée. Par exemple, les modalités visuelles permettent de présenter une information étendue ou plusieurs informations dans l'espace, tandis que les modalités haptique sont beaucoup plus localisées et les modalités sonores propices au rendu d'informations temporelles (voir section suivante). De plus la compréhension de l'information transmise par une modalité, et donc la validité sémantique de la modalité, varie en fonction de l'utilisateur : elle dépend de son expérience, de sa sensibilité, du caractère naturel de la modalité (i.e. qu'elle fait appel aux schèmes d'interaction du monde réel). Enfin, le choix d'une modalité dépend du contexte de la communication : nature de la tâche, importance de l'information, présence d'autres informations, nombre d'utilisateurs, etc.

La distinction entre les termes "modalité" et "média" nous permet alors de séparer les systèmes appelés *multimédia* ou *multi-sensoriels* et ceux qualifiés de *multimodaux*. Les premiers font appel à plusieurs canaux humains en utilisant plusieurs médias et plusieurs modalités, mais sans que le choix de celles-ci soit guidé par des règles informationnelles, c'est-à-dire la recherche de la meilleure adéquation entre le message à transmettre et le moyen de transmission. Ce sont par exemple les récents baladeurs audio-video permettant d'écouter de la musique avec des écouteurs et de regarder des films sur un petit écran. Dans ces systèmes, la modalité et l'information sont confondues.

Les systèmes multimodaux sont au contraire tournés vers la meilleure transmission de l'information, et cherchent à exploiter au mieux les modalités en fonction de leurs capacités sémantiques et du contexte de la tâche. "Multimodalité" est alors synonyme de "multimédia intelligent" dans le cadre de l'IHM¹, et donc de "multi-

¹Dans la suite de ce manuscrit, nous emploierons le terme "IHM" pour désigner le domaine de l'Interaction Homme-Machine non immersive, par opposition à la RV, et non pas pour désigner le domaine général incluant toutes les applications dans lequel un utilisateur humain interagit avec

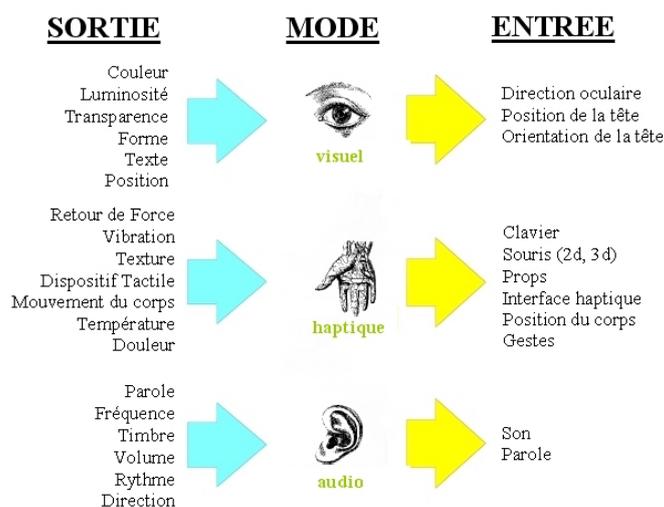


FIG. 1.9 – Exemples de modalités d'entrée et de sortie en RV [Nesbitt, 2003]

sensoriel intelligent" dans le cadre de la RV. Cette "intelligence" suppose alors des choix de rendu, ce qui peut amener au remplacement d'une modalité par une autre, à une recherche de complémentarité ou de redondance, en fonction de l'utilisateur et du contexte. Nous appellerons ce choix de rendu un *paradigme*. En IHM, Dragicevic [Dragicevic, 2004] définit un paradigme d'interaction comme étant "un ensemble cohérent de techniques d'interaction qui coopèrent de façon étroite, ou qui reposent sur les mêmes principes techniques ou conceptuels". Il donne l'exemple du WIMP² qui regroupe des techniques comme le pointage et les menus. En linguistique, Le Petit Larousse définit un paradigme comme "l'ensemble des formes d'un mot pris comme modèle (déclinaison ou conjugaison)" et "l'ensemble des unités qui peuvent être substituées les unes aux autres dans un contexte donné". En philosophie, un paradigme est également une représentation du monde, une manière de voir les choses, une modélisation. Notre définition d'un paradigme est un peu plus proche de l'acceptation philosophique de de celles généralement utilisées en IHM. En effet, dans ce manuscrit, un paradigme sera l'usage particulier (i.e. fonction de certains paramètres) d'une ou plusieurs modalités pour communiquer une information donnée et dans un contexte précis.

Voici deux exemples de paradigmes simples et bien connus. Pour représenter une information de chaleur, on peut choisir la modalité "couleur" sur le média de projection, et il est alors pertinent d'utiliser le rouge et le bleu pour symboliser respectivement le chaud et le froid. Pour une information de volume sonore, un

une système informatique.

²WIMP : Windows, Icons, Mouse, Pull-down menus (fenêtres, icônes, souris, menus déroulants).

paradigme possible est l'utilisation d'une barre graduée : on assiste ici à un rendu partiellement redondant et complémentaire entre le volume lui-même qui est perçu par l'ouïe, et la quantification de ce volume sur une échelle qui est intégrée à partir de la vue.

1.4 Objet de cette thèse

Les rappels théoriques des sections précédentes nous permettent de définir le cadre général de cette thèse : la multimodalité en sortie, ou rendu multimodal, dans une application de RV orientée tâche. Nous étudierons donc la communication d'informations par l'environnement à l'utilisateur, et la combinaison des modalités dans l'objectif d'améliorer l'activité des utilisateurs d'applications de RV.

Comme nous l'avons vu précédemment, dans une approche non réaliste, la multimodalité en sortie est affaire de choix : pour une ou plusieurs informations à transmettre à l'utilisateur, quelle modalité ou combinaison de modalités est la plus adéquate ? On parle de décision de distribution. Cette décision dépend bien évidemment de l'information en elle-même. Ensuite, elle doit être issue de connaissances a priori, sur les caractéristiques des modalités disponibles, les capacités de rendu de l'application, les capacités sensori-motrices du ou des utilisateurs, le domaine de l'application, etc. Enfin, tous les travaux dans ce domaine montrent que la décision doit dépendre du contexte courant au moment du rendu : état des médias, autres informations en cours de rendu, actions et objectifs de l'utilisateur, etc. Or, l'état du contexte ne peut pas être connu a priori, avant l'exécution de l'application. Ils ne peuvent donc pas être pris en compte par le concepteur humain à moins de prévoir tous les cas possibles. A partir de ces observations, on pourrait imaginer que la situation idéale serait un composant artificiel et intelligent, positionné au-dessus de l'application, et capable de choisir à tout moment la combinaison la plus adéquate de modalités pour n'importe quelle tâche, n'importe quel utilisateur, n'importe quelle donnée et ce en toutes circonstances. On comprend cependant assez vite que cette vision de la multimodalité est utopique, et il faut même se demander s'il est souhaitable.

Face à ce graal, il est d'abord nécessaire d'observer où en est aujourd'hui le rendu multimodal en RV (et en IHM), comment il peut bénéficier ou nuire aux utilisateurs, et quelles sont les applications envisagées et leurs besoins. Puis il faut étudier comment il doit être géré et quelles sont les problèmes soulevés par les contraintes particulières du domaine (interaction immersive et temps réel, mouvements de l'utilisateur, médias particuliers, etc.).

Cette thèse va tout d'abord présenter l'état de l'art de la multimodalité en sortie en RV : interactions possibles, applications multimodales, modèles et études psychophysiques.

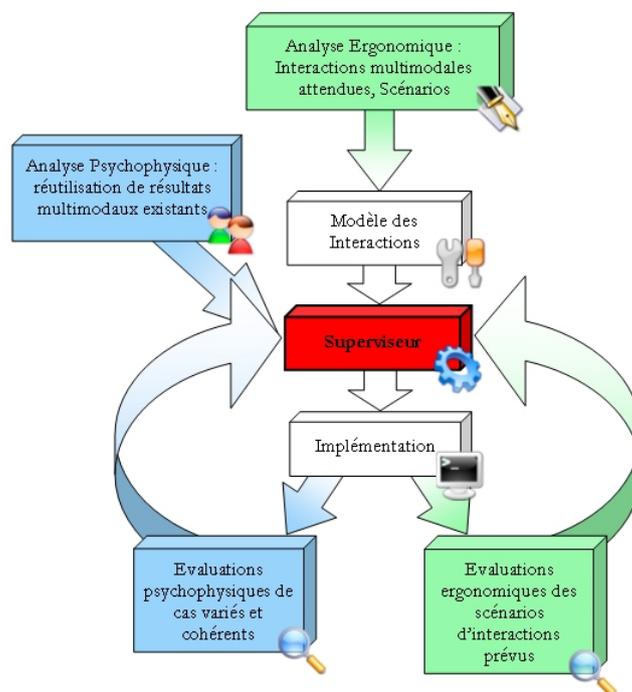


FIG. 1.10 – Double boucle de conception.

Puis, nous allons exposer notre modèle de gestion du rendu multimodal. Notre point de vue est que le rendu multimodal doit respecter l'essence de la RV, c'est-à-dire la coopération entre un utilisateur humain et un environnement numérique. Le rendu doit donc faire l'objet d'une décision "relativement intelligente" du système, sur la base de connaissances a priori données par le concepteur, et d'une observation continue de tous les éléments du contexte. Cependant cette décision doit être encadrée, et le système doit notamment respecter absolument les choix de l'utilisateur, qui veut rester maître de sa tâche. Le rendu doit donc être géré par le système, lui-même adapté à l'utilisateur. Nous détaillerons le concept de "superviseur multimodal". Ce modèle est indépendant de la réalisation technique du rendu (architecture des médias, gestion de la scène, synchronisation des différentes boucles de rendu, etc.).

Ce modèle doit respecter les deux contraintes majeures de la RV que nous avons évoquées ci-dessus : le respect de la psychophysique humaine, ainsi que l'utilité, l'utilisabilité et l'utilisation finale des applications. C'est pourquoi notre méthodologie de conception suit à la fois une boucle expérience-résultat psychophysique et une boucle analyse-évaluation applicative (figure 1.10).

Enfin, les résultats apportés par une première implémentation et les problèmes soulevés au cours des recherches seront discutés.

1.5 Conclusion

Dans ce chapitre, nous avons présenté la Réalité Virtuelle, dont la principale caractéristique est d'être le siège d'interactions entre un ou des humains et un environnement numérique. Nous avons vu que l'immersion de l'humain dans son activité pose des problèmes particuliers, notamment la nécessité d'études psychophysiques et d'évaluations ergonomiques. Nous avons montré en quoi les recherches sur la multimodalité en sortie nous semblent nécessaires, et nous avons détaillé le cadre et l'objectif de cette thèse : concevoir un système de gestion du rendu multimodal prenant en compte la sémantique des informations et le contexte de la tâche.

Dans la suite de ce manuscrit, nous présenterons en détail les applications, les recherches et les modèles qui existent dans le domaine de la multimodalité en sortie en RV. Puis nous expliquerons notre contribution au travers d'un modèle des interactions de RV, d'un processus de supervision et d'un moteur de décision du rendu multimodal. Ces contributions feront l'objet d'évaluations.

Chapitre 2

Etat de l'art

Résumé. *Dans ce chapitre, nous présentons tout d'abord les dispositifs de RV ainsi que les techniques d'interactions associées aux canaux visuel, audio et haptique. Puis nous illustrons à l'aide d'exemples trois catégories d'applications de RV faisant appel à la multimodalité en sortie. Ensuite nous examinons des espaces de conception et des questionnaires multimodaux, principalement en IHM. Enfin nous nous intéressons aux expériences psychophysiques portant notamment sur la combinaison des modalités. Cet état de l'art nous permet de mettre en évidence les différents éléments capitaux à prendre en compte et à modéliser dans notre superviseur de rendu multimodal.*

2.1 Introduction

Une première étape vers un superviseur multimodal est de savoir exactement quelles sont les interactions possibles dans des applications de RV. Après une étude des possibilités techniques de rendus en RV, nous observons des applications faisant appel à la multimodalité en sortie ainsi que des modèles et des espaces de conception multimodaux, notamment en IHM. Puis nous nous intéressons aux expériences psychophysiques mettant en évidence des illusions modales et des effets intermodaux.

2.2 Techniques de rendus en RV

Nous l'avons vu dans le chapitre précédent, les interactions entre l'humain et l'environnement virtuel sont au centre des préoccupations de la RV. Foley a défini une "technique d'interaction" comme la manière de se servir d'un dispositif d'entrée pour accomplir une tâche sur l'ordinateur [Foley et al., 1990]. Nous allons présenter dans cette section les trois principales catégories d'interactions en sortie en RV correspondant aux trois canaux que sont le visuel, l'audition et le sens haptique.

Certaines sont plus utilisées que d'autres. Nous verrons pour chacune d'entre-elles quelles sont les interfaces disponibles ainsi que leurs caractéristiques techniques et logicielles, leurs possibilités de rendus, que nous mettons en rapport avec les capacités sensori-motrices et psychophysiques humaines.

2.2.1 Définitions

Avant de détailler les trois principales catégories d'interactions, nous devons expliciter certains termes que nous avons utilisés dans le chapitre précédent. En miroir de notre distinction entre les interfaces d'entrée et de sortie (section 1.1) du système, nous qualifions différemment les interfaces d'"entrée" et de "sortie" de l'utilisateur humain. Tout d'abord, les interfaces sensorielles sont celles qui informent l'utilisateur de l'état de l'environnement par l'intermédiaire de ses sens. Ensuite, les interfaces motrices informent le système informatique des actions de l'humain. Enfin, certaines interfaces permettent à l'humain à la fois de ressentir et d'agir sur un environnement : ce sont les interfaces sensori-motrices.

Dans les trois sections suivantes, les interfaces visuelles et sonores sont des interfaces sensorielles (pour l'humain) et de sortie (pour le système informatique), tandis que les interfaces haptiques sont sensori-motrices, à la fois en entrée et en sortie du système.

2.2.2 Rendu visuel

La vue est le sens prédominant de l'être humain. Il est principalement utilisé pour se situer dans l'espace, pour percevoir les objets qui nous entourent ainsi que leurs caractéristiques, telles la taille, la forme, la couleur, la disposition, etc. notamment grâce à un système complexe de perception de la profondeur.

Les interfaces visuelles sont les premières à avoir été utilisées et sont donc les plus exploitées et les plus variées. Il existe deux grandes familles d'interfaces visuelles. Les dispositifs à support fixe utilisent un ou plusieurs écrans de grande taille, disposés sous forme de bancs ("workbench") ou de faces d'un cube (CAVE) et accompagnés d'un système de projection stéréoscopique. La configuration extrême serait composée de six écrans formant un cube dans lequel l'utilisateur serait totalement immergé. Les interfaces portables sont au contraire des dispositifs de type casque : des écrans de petite taille proposent une image stéréoscopique, proches des yeux de l'utilisateur et occultent plus ou moins l'environnement réel. On les appelle visiocasques (ou HMD pour Head Mounted Displays) ou visiolunettes. Les figures 2.1, 2.2 et 2.3 présentent quelques-unes des interfaces existantes.

Les interfaces doivent dans tous les cas respecter du mieux possible les caractéristiques psychophysiques humaines. Tout d'abord, le champ de vision humain avec les deux yeux et la tête mobiles est de 200° horizontalement, de 140° vers le haut

et de 170° vers le bas (170° horizontalement et 145° verticalement avec la tête fixe) [Fuchs et al., 2006b]. Dès lors, les visiocasques actuels semblent limités pour respecter ce champ. A l'intérieur de ce champ, l'acuité visuelle n'est vraiment précise que sur un cône de 2° d'angle. La limite théorique pour que l'humain puisse percevoir fluidement les mouvements dans une scène monoscopique est de 25 images par seconde. La fréquence de rafraîchissement est donc de 50 Hz pour une scène stéréoscopique (cependant, une scène très dynamique devra être rafraîchie plus rapidement). Il existe également des limites basses pour que la résolution et la luminosité ne gênent pas l'utilisateur. De plus, les couleurs sont également à prendre en considération. Ainsi, s'il est estimé à 2 millions le nombre de couleurs perceptibles par un humain, toutes n'ont pas le même effet. Par exemple, les couleurs chaudes font apparaître les objets plus grands que les couleurs froides. Une autre caractéristique de la perception visuelle intéressante est l'indépendance au champ : c'est la capacité pour un humain de reconnaître rapidement des objets dans une scène complexe sans effectuer une exploration approfondie, ce qui a des conséquences sur le sentiment d'immersion. En effet, ceux qui n'ont pas cette capacité, qui sont dépendant au champ, doivent explorer l'ensemble des objets d'une scène pour en reconnaître certains. Ils ont besoin de champs visuels plus importants pour se sentir immergés [Fuchs et al., 2006b]. La caractéristique principale de la vision est la perception de la profondeur. Cette perception est issue en partie de la fusion par le cerveau des deux images issues des yeux droit et gauche, mais également par un ensemble d'indices visuels monoculaires. Parmi ceux-ci, notons : les ombres et lumières, les dimensions relatives des objets, les occultations, les effets atmosphériques pour les objets lointains, la perspective, etc. Générer des images 3d en RV consiste à réutiliser les indices monoculaires dans les scènes, et à recréer artificiellement les images droite et gauche de l'utilisateur humain, à l'aide de différentes techniques. Pour les dispositifs portables, chaque oeil reçoit ses images. Pour les dispositifs à support fixe, la séparation entre les images peut être faite à l'écran ou à l'aide de lunettes, de manière active (synchronisation entre la génération d'une image pour un oeil et le masquage de l'oeil opposé) ou passive (anaglyphe, polarisation). Pour des raisons de confort, la stéréoscopie active demande une fréquence de balayage (et donc d'occultation des lunettes) de l'ordre de 110 Hz, afin d'éviter les effets de scintillation.

Les technologies actuelles de projection donnent des résultats satisfaisants en terme de qualité et de taille d'image. Certains visiocasques commencent à rivaliser avec ces performances, mais au prix d'un poids et d'un coût élevés. En terme de fréquence de rafraîchissement, les capacités de calcul actuelles sont acceptables pour la majorité des scènes. La complexité de la scène augmente évidemment les temps de calculs, et donc diminue la fluidité, à moins d'utiliser des techniques de parallélisation délicates à mettre en oeuvre.

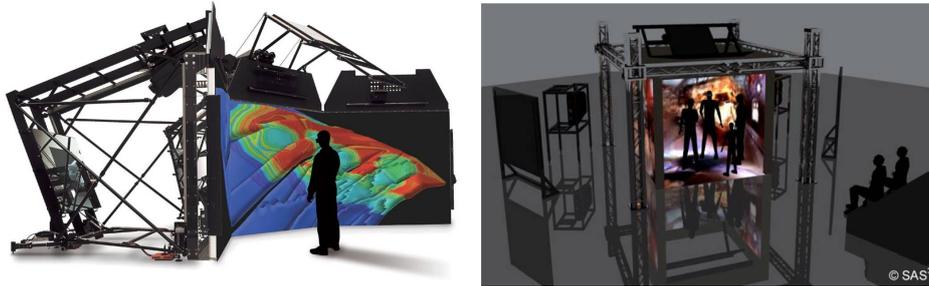


FIG. 2.1 – Dispositifs immersifs de type CAVE ou RAVE : reconfigurable (MoVE de BARCO) ou fixe (SAS de TAN).



FIG. 2.2 – Dispositifs immersifs de type Workbench de BARCO.

2.2.3 Rendu audio

L'audition est le deuxième canal le plus utilisé par l'être humain pour s'informer. Contrairement à la vision, il est omnidirectionnel. Ceci apporte à la fois des avantages, puisque tous les sons de l'environnement sont captés et que l'on n'a pas besoin d'être tourné vers quelque chose pour l'entendre ; mais aussi des inconvénients, car aucun son n'est filtré. L'audition est particulièrement performante pour la perception des phénomènes d'alerte et des phénomènes qui évoluent dans le temps. De plus, s'il est possible de ne pas prendre conscience d'une donnée visuelle (surtout si elle est brève ou mal située), il est beaucoup plus difficile de ne pas percevoir un changement dans un son.

Les fréquences audibles sont situées entre 20 Hz et 20 kHz. Dans la zone entre 1 kHz et 3 kHz, qui correspond à la voix humaine, nous sommes sensibles à des différences de l'ordre de 3 Hz, contre 10 Hz dans les basses fréquences, et 30 Hz dans les hautes fréquences. Cette discrimination est à comparer avec notre incapacité à



FIG. 2.3 – Visiocasque piSight de Sensics (résolution de 3200x1800 pixels par oeil, champ de vision de 150°) et visiocasque à support fixe Boom de Fakespace.

distinguer des flashes lumineux à 100, 200, 1000 ou 10000 Hz. Un autre avantage de ces caractéristiques fréquentielles est que la perception auditive peut fonctionner sur plusieurs échelles de temps simultanément. Nous pouvons distinguer des signaux à 5000 Hz, 500 Hz, 50 Hz, 5 Hz, 0.5 Hz (tous en même temps) et percevoir des changements minimes dans chacun d'eux. Enfin, notons la propriété des relations harmoniques : lorsque deux signaux ont un ratio entier de fréquence (e.g, 100 Hz et 200 Hz ou 1 Hz et 0.5 Hz), ils se combinent pour former un seul son, de timbre ou de rythme complexe [Malinowski, 2000].

La hauteur subjective d'un son (la tonie) n'est pas reliée directement à la fréquence. Le volume sonore est également à prendre en compte : un chuchotement est de l'ordre de 20 dB et on sait aujourd'hui qu'un niveau sonore continu ne doit pas dépasser 50 dB sous peine de devenir une nuisance sonore, et qu'au-delà de 65 dB, la situation est considérée comme pénible, si l'exposition est longue. Une exposition à un son de plus de 120dB, même pendant un court instant, peut entraîner des lésions irréversibles de notre système auditif.

Le système acoustique possède également des capacités intéressantes de localisation d'un son dans l'espace, même si la discrimination spatiale auditive est faible comparée à la vision [de Cheveigné, 2003]. Ainsi, notre acuité angulaire auditive est de 1° en azimut (3° pour des sources en mouvement) et de 10° à 20° en élévation. En

vision, cette acuité est de l'ordre de 5 secondes d'arc. L'acuité auditive est dégradée pour certains stimuli ou en présence de réverbération. Les indices de localisation sont la différence de temps d'arrivée du son entre les deux oreilles (ITD pour interaural time difference), la différence d'intensité des sons (ILD pour interaural level difference) et le décalage de phase.

En RV, le rendu audio est effectué soit par plusieurs haut-parleurs disposés dans l'espace, soit par casque. Ce rendu peut être mono, stéréo ou tridimensionnel. La spatialisation peut être effectuée en distribuant les émissions sonores sur différents haut-parleurs, créant ainsi un champ sonore, ou en utilisant la synthèse binaurale : avec un simple casque ou deux enceintes, on peut reconstituer le son qui arrive aux oreilles de l'utilisateur par l'intermédiaire d'une "fonction de transfert de la tête" (HRTF pour Head-Related Transfert Function). Celle-ci est propre à chaque être humain, et dépend notamment de la forme de ses oreilles et de sa tête (figure 2.4). Les performances des dispositifs audio actuels sont bonnes en terme de qualité de rendu, mais perfectibles pour ce qui est de la spatialisation.

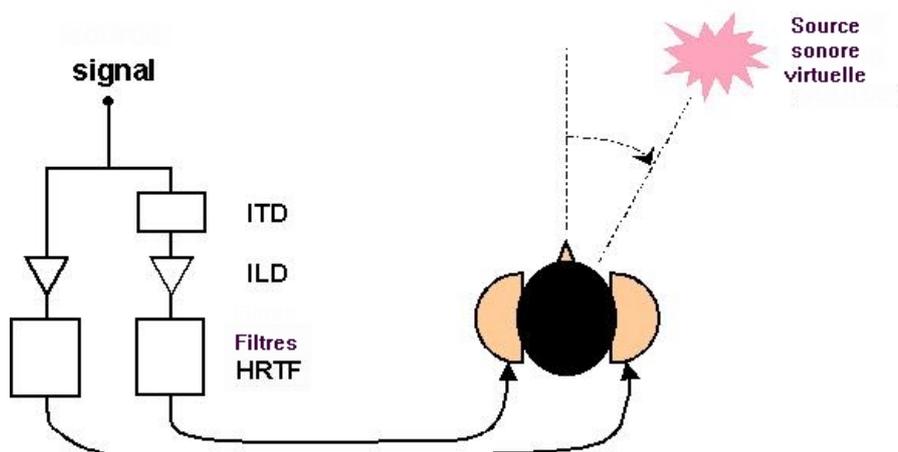


FIG. 2.4 – Principe du rendu audio spatialisé en utilisant les HRTF.

2.2.4 Rendu et interaction haptique

Les interactions haptiques sont à la fois en entrée et en sortie du système informatique : l'homme a une action motrice sur l'environnement et une perception sensorielle de celui-ci. Le canal haptique possède trois fonctionnalités [Ammi, 2005] :

- une fonction ergotique, celle de l'action matérielle, de la modification, de la transformation de l'environnement ;
- une fonction épistémique, qui nous permet de connaître, de percevoir l'environnement : contact, position et déplacement des objets, mais également leurs

propriétés : forme, masse, texture, etc. Cette deuxième fonction est la plupart du temps simultanée à la première ;

- une fonction sémiotique d'émission d'information vers l'environnement (par le geste).

Une autre caractéristique qui rend les interactions haptiques particulières est qu'elles mettent en jeu deux sous-canaux sensori-moteurs. Le sens tactile permet de percevoir, par l'intermédiaire de la peau, les propriétés surfaciques des éléments extérieurs : géométrie, courbure, texture, vibrations, température, glissement, etc. Le sens kinesthésique permet lui, par l'intermédiaire des muscles et des tendons, de percevoir nos propres mouvements et les efforts du monde extérieur sur notre corps : lorsque nous saisissons un objet, nous percevons sa dureté, son poids, son inertie, les forces de contacts entre nos doigts et sa surface.

Les interfaces haptiques sont donc elles aussi divisées en interfaces tactiles et kinesthésiques (ou interfaces à retour d'effort). Les interfaces tactiles sont rares. Elles utilisent souvent une matrice de petits éléments (picots) qui peuvent se soulever de quelques microns (figure 2.5). Elles peuvent également faire appel à des signaux électriques ou des coussins d'air agissant sur la pulpe des doigts, ou encore à des vibreurs situés sur les phalanges (gants vibro-tactiles). Les interfaces à retour d'effort sont aujourd'hui beaucoup plus répandues. On rencontre des systèmes à réaction interne, i.e. dont la structure est portée par l'utilisateur (exosquelettes), ou à réaction externe, i.e. dont la structure est implantée dans l'environnement. Les structures mécaniques de ces interfaces peuvent être séries ou parallèles, les unes limitées en effort mais pouvant avoir un grand espace de travail, les autres autorisant de grands efforts mais de faibles débattements. Il existe également des dispositifs avec des structures mixtes. Il est à noter que les interfaces kinesthésiques peuvent en outre créer des stimuli tactiles de type vibration, à l'aide des hautes fréquences. Les figures 2.6 et 2.7 présentent des exemples de bras haptiques, qui sont des interfaces série à réaction externe, ainsi que des exemples d'exosquelettes de main.



FIG. 2.5 – Dispositif tactile de type matriciel (Université de Karlsruhe) et gant vibro-tactile CyberTouch (Immersion).



FIG. 2.6 – Dispositifs haptiques à réaction externe : bras Virtuose de Haption (structure série) et interface Delta de Force Dimension (structure parallèle).

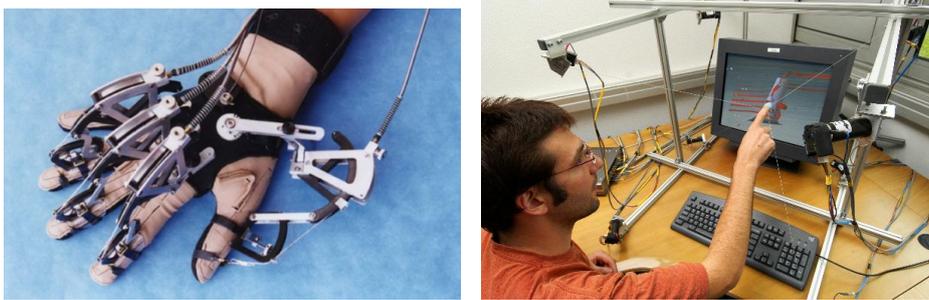


FIG. 2.7 – Exosquelette de main Cybergrasp de Immersion et interface Spidar à base de câbles (image de l'INRIA).

Les caractéristiques des interfaces haptiques à prendre en compte sont nombreuses : la capacité et la résolution en efforts, la raideur, la transparence (perception la plus faible possible de l'interface : poids, inertie, etc.), l'espace de travail et la résolution en position et en orientation, les degrés de liberté et d'effort, la posture et la prise en main.

La perception tactile est sensible à des fréquences de 1 à 300 Hz pour des objets mous et jusqu'à 10 kHz pour des objets durs, tandis que la perception kinesthésique varie de 1 à 10 Hz selon la nature du stimuli [Bayle, 2003]. Cependant, pour des raisons mécaniques, électroniques et automatiques, les forces virtuelles doivent avoir un taux de rafraîchissement supérieur à 1 kHz pour avoir des interactions haptiques stables en temps réel. Ceci en fait une contrainte essentielle lors de l'intégration d'un périphérique haptique dans une application de RV. Le système doit être capable de générer les informations haptiques à cette vitesse, alors que le rendu visuel ne réclame en général qu'une boucle de rafraîchissement à 100 à 120 Hz. Il existe toutefois des

solutions techniques et logicielles permettant de garantir la stabilité haptique même à de faibles fréquences, au prix d'autres inconvénients. Ainsi, le couplage virtuel entraîne une plus grande viscosité des mouvements, tandis que l'extrapolation et l'interpolation ont respectivement pour conséquence un décalage des rendus dans l'espace et dans le temps.

La perception tactile est sensible à différents phénomènes : pression (résolution de 2 g/mm^2), température (fenêtre d'excitation $15\text{-}48^\circ$, discrimination 2%), vibrations (résolution de 0.4Hz pour la fréquence, $0.1 \mu\text{m}$ pour l'amplitude), etc.

Les interfaces tactiles actuelles posent de nombreux problèmes de performances, tant en taille des surfaces de contact qu'en précision de la stimulation. Des travaux sont cependant en cours pour combler ces lacunes : l'IBISC-CNRS a par exemple couplé une matrice tactile avec des stimulations thermiques, le tout sur une plateforme haptique à 3 degrés de liberté [Drif et al., 2005]. Les interfaces kinesthésiques semblent moins restrictives pour leur intégration dans des systèmes immersifs, malgré un espace de travail réduit par rapport à la taille des écrans ainsi qu'un encombrement qui les rend invasifs dans un environnement immersif. D'un autre côté, les calculateurs permettent d'avoir la fréquence de rafraîchissement suffisante pour la stabilité dans de nombreuses scènes. Comme pour le visuel, cette fréquence dépend de la complexité de la scène, et de la précision demandée pour les contacts. Elle profite énormément des avancées théoriques et logicielles, notamment dans les algorithmes de détection des collisions et la parallélisation des calculs.

Dans la suite de ce manuscrit, nous emploierons le terme "haptique" pour toutes les interactions liées aux canaux tactile et kinesthésique, bien que les paradigmes étudiés ou donnés en exemple seront le plus souvent kinesthésiques.

2.2.5 Utilisation des rendus

A travers ces trois présentations des canaux et des interfaces sensori-motrices, nous voyons bien que chacun à ses avantages et ses inconvénients. Ces différences portent évidemment sur les types des stimuli, mais également sur leurs caractéristiques spatiales, temporelles, fréquentielles, etc.

Friedes [Friedes, 1974] explique ces différences par le fait que chaque canal sensoriel possède une méthode unique de traitement de l'information, et qu'ainsi chaque modalité sensorielle est adaptée à un certain type d'information. Cette théorie psychophysique est appelée Modal Specific Theory (ou Théorie des Spécificités Modales). La figure 2.8 compare ainsi les trois canaux décrits dans les sections précédentes. La vision est présentée comme un canal capable d'interpréter des relations spatiales. L'audition est un canal adepte de l'interprétation de relations temporelles. Le sens haptique est, lui, propice à la reconnaissance de mouvements, qui sont des phénomènes à la fois temporels et spatiaux.

Cette vision semble néanmoins caricaturale. En effet, concernant les phénomènes

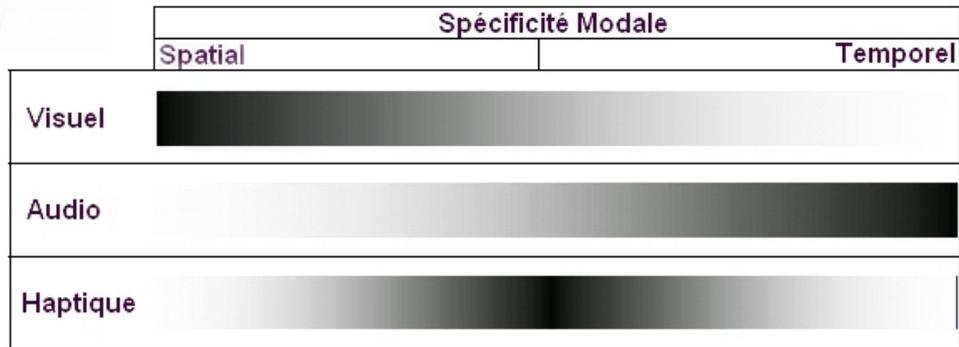


FIG. 2.8 – Comparaison des modalités selon la Théorie des Spécificités Modales [Friedes, 1974].



FIG. 2.9 – Relativisation de la Théorie des Spécificités Modales.

de mouvements, la vision sait distinguer le mobile par rapport au statique, même si les déplacements propres de l'observateur ainsi que la distance du mobile peuvent être perturbateurs. L'audition peut, elle, d'une part percevoir les mouvements grâce à l'effet Doppler, d'autre part localiser spatialement les événements. De son côté, la perception des mouvements par l'haptique est relative, car elle est extrêmement locale. En contrepartie, l'haptique peut localiser les objets et le tactile est dans une certaine mesure sensible à des stimuli fréquentiels. En appliquant ces remarques à la figure 2.8 et en dilatant les curseurs, nous pouvons obtenir une nouvelle représentation (figure 2.9).

Ce qui est valable pour les canaux est bien sûr valable pour les modalités qui sont transmises par ces canaux. Cette dernière figure justifie donc deux aspects de la multimodalité en sortie :

1. L'optimisation des rendus est possible car les canaux et les modalités ont des caractéristiques qui les rendent propices à la transmission de certaines informations ;

2. Dans les situations où l'utilisation des canaux et des modalités optimaux est impossible, on peut envisager de substituer ceux-ci par d'autres modalités sur les mêmes ou sur d'autres canaux.

Le premier problème que l'on rencontre lorsque l'on désire faire du rendu multimodal est donc celui de l'allocation modale. André le décrit comme suit : "Etant donné un ensemble de données et un ensemble de modalités, comment trouver la combinaison de modalités qui communique efficacement toutes les données pour une situation donnée ?" [André, 2000]. Les réponses à ce problème vont prendre la forme de "règles de projection" (mapping rules) ou de "fonctions transfert" (transfert functions) entre les modalités et les données.

Dans ce contexte, il convient d'étudier plus finement les apports de chaque canal et modalités sensori-moteurs. Les paragraphes suivants détaillent l'utilisation qui est faite des caractéristiques des canaux et des modalités auditives et haptiques dans des applications de RV.

2.2.5.1 Sonification de données

La sonification de données est la représentation de données (numériques notamment) dans le domaine acoustique dans un but d'interprétation et de compréhension du domaine étudié. Plus généralement, c'est l'utilisation du rendu audio non vocal pour transmettre des informations. Comme nous l'avons vu dans les sections précédentes, le son possède un ensemble de propriétés uniques qui peuvent être exploitées pour la présentation de données. Il attire l'attention, car le canal auditif est toujours ouvert. Il peut donner du relief au canal visuel en présentant des données multivariées. En tant que phénomène temporel, il peut naturellement représenter des données variant au cours du temps. Le son peut révéler des motifs, des répétitions, des périodicités, des signatures dans les données. Il peut également palier la déficience du canal visuel (surcharge, déformation, occultation, etc.).

Plusieurs techniques coexistent pour utiliser les modalités sonores dans une application. La plus simple est la transmission directe de la donnée en tant que son [Kramer, 1994]. Il s'agit par exemple de la diffusion d'événements sismiques enregistrés lors d'un tremblement de terre [Hayward, 1994]. Il faut néanmoins veiller à l'audibilité du son par certains filtrages. Une deuxième méthode est la projection des attributs des données sur des dimensions sonores (volume, durée, fréquence, timbre, localisation, etc.). Enfin, l'on peut faire correspondre un son et une information de manière causale, métaphorique ou arbitraire. Par exemple, le déplacement d'un objet téléopéré peut être accompagné d'un son de frottement [Ammi, 2005] (relation causale); associer un bruit de porte qui se ferme à la fermeture d'une application est métaphorique; émettre un 'bip' simple lorsque l'utilisateur a réussi à atteindre une cible haptique est arbitraire, et ne peut être compris que par expérience ou apprentissage.

2.2.5.2 Haptisation de données

L'utilisation d'une interface haptique dans une application de RV consiste à transposer les fonctions ergotiques, épistémiques et sémiotiques du monde réel (section 2.2.4) vers l'environnement virtuel. Les fonctions ergotiques et épistémiques nécessitent une interaction directe de l'opérateur sur les données de la scène. Elles reposent sur l'utilisation d'un modèle géométrique de la scène et d'une librairie de détection de collision [Ammi, 2005]. Cela permet à l'utilisateur de ressentir les efforts des objets entre eux, les chocs, les frottements, mais également les propriétés des objets : poids, inertie, texture, etc. L'exploration de sols virtuels en est un exemple : Green [Green and Salsbury, 1998] simule la manipulation d'une charrue à travers différents sols sableux (densité, cohésion et angle de friction interne). Il utilise également le retour d'effort pour faire ressentir la friction statique et les déviations de surface [Green, 1997]. Ceci permet à l'utilisateur de ressentir les textures de surfaces simulées, en plus de les voir.

Cependant, une autre catégorie intéressante d'haptisation de données existe qui peut être qualifiée d'interaction indirecte. Celle-ci concerne les données abstraites en général et les guides virtuels en particulier. Comme leur nom l'indique, les guides virtuels sont une métaphore de l'assistant, du maître, du coopérateur. Il s'agit d'apporter des fonctions haptiques à l'opérateur, tel le guidage des opérations manuelles ou encore des contraintes sur le mouvement, afin d'alléger la complexité de l'activité, d'améliorer le confort, la précision, la qualité de la tâche. Mais l'haptisation peut aussi servir à la communication d'informations qui ne sont pas de nature haptique dans le monde réel. Tout comme la sonification, plusieurs techniques existent. Si les données le permettent, elles peuvent être directement transformées en signal haptique (moyennant filtrages) : un champ de vecteurs peut être transformé en forces (ainsi, un champ électromagnétique donnerait l'illusion à l'utilisateur de manipuler un aimant) tandis qu'une variable numérique peut être communiquée par vibrations. D'autres données peuvent nécessiter des représentations intermédiaires ou la transposition de métaphores existantes, afin de créer des effets haptiques de ces données sur un effecteur virtuel manipulé par l'utilisateur. Par exemple, une courbe graphique peut être haptisée pour les aveugles : ceux-ci manipulent un stylet et suivent la courbe comme ils suivraient du doigt les arêtes d'un solide. Ou encore, des données abstraites peuvent être représentées par des isosurfaces, qui sont ensuite haptisées comme des surfaces de solides.

On retrouve donc ici une des justifications de la multimodalité énoncée dans le chapitre précédent : il semble judicieux d'exploiter au mieux les caractéristiques des différentes modalités en fonctions des propriétés des informations que l'on veut communiquer ou du rôle que l'on veut faire jouer au système informatique. C'est cette idée que de nombreuses équipes ont mise en oeuvre dans des recherches ou des

applications finales. Nous en présentons certaines dans la section suivante.

2.3 Applications multimodales en RV

De nombreuses applications de RV combinent les rendus sensori-moteurs, mais les travaux réellement multimodaux sont moins répandus. Comme nous l'avons vu dans le chapitre précédent (section 1.3), la multimodalité se distingue de la multi-sensorialité par la prise en compte de la sémantique des informations dans le choix des rendus. Nous allons nous intéresser ici aux applications ou aux recherches orientées tâches, c'est-à-dire intégrant une ou plusieurs modalités en sortie, naturelles ou non, dans une optique d'amélioration de l'activité de l'utilisateur humain. Pour cela, les efforts sont portés sur la recherche de la meilleure représentation, de la meilleure projection entre modalités et informations, et non pas sur la simple diversification des rendus. Les applications qui nous intéressent ne sont pas seulement multi-sensorielles mais multimodales, elles ne cherchent pas seulement à combiner les rendus visuels, sonores et haptiques mais relient ces rendus aux informations transmises du système informatique vers les utilisateurs humains.

Les trois grandes catégories d'applications que nous avons identifiées sont : l'apprentissage et la formation ; l'exploration de données ; l'assistance en environnement augmenté. Nous présentons un aperçu de différents travaux dans ces domaines.

2.3.1 Apprentissage, formation

Dans les divers travaux utilisant la RV pour l'apprentissage et la formation, l'objectif est généralement de recréer les conditions réelles (ou réalistes) de l'activité et d'y faire agir l'apprenant afin qu'il acquiert ou perfectionne son expérience. Les possibilités de la RV sont ensuite exploitées pour faire varier les conditions et les tâches à effectuer, en fonction des objectifs et du rythme de l'apprenant, et en particulier pour créer des conditions limites, des situations à risque, pour tester son comportement. Les techniques de RV peuvent enfin servir à aider l'apprenant, à la fois dans son geste s'il y en a un, et dans sa compréhension de l'activité ou des phénomènes.

L'apprentissage et la formation sont donc des domaines logiquement orientés vers la tâche : l'amélioration recherchée ici est celle de l'expérience, de l'habileté, de la compréhension et des connaissances de l'apprenant.

Burkhardt, Lourdeaux et Mellet d'Huart présentent dans le quatrième tome du Traité de la RV [Burkhardt et al., 2006] de nombreux apports de l'utilisation de la RV pour l'apprentissage.

La première fonctionnalité qui à notre sens concerne directement la multimodalité en sortie est la flexibilité des modes de représentation. La RV permet tout d'abord

l'exploitation de modes d'interactions et d'assistance impossibles dans le monde réel. Elle rend notamment possibles des représentations reconstruites de toutes pièces et qui permettent de comprendre des phénomènes non perceptibles ou des causalités non intuitives dans le monde physique. Elle autorise également l'ajout, la superposition d'informations d'assistance avec la scène qui fait l'objet de l'étude ou à l'inverse le filtrage, la sélection des informations pertinentes pour l'apprentissage, parmi toutes les données de l'activité réelle.

Le deuxième apport de la RV à l'apprentissage est l'ancrage sensori-moteur de l'apprenant dans son activité. La combinaison multi-sensorielle peut augmenter sensiblement la compréhension des phénomènes et l'utilisabilité des applications, et particulièrement dans des situations non réalistes, ce qui est justement la première fonction dont nous avons parlé.

Enfin, la troisième problématique intéressante pour la multimodalité en sortie est la coopération nécessaire entre l'apprenant et l'EV. Elle est fortement liée aux deux points précédents. Le système va notamment jouer un rôle pour répondre au besoin d'adapter la représentation au contexte, qui est ici prépondérant. Celui-ci est constitué des objectifs globaux et locaux de l'apprentissage, des actions de l'apprenant, de son style d'apprentissage, etc.

Enseignement Dans le domaine de l'enseignement, le projet ScienceSpace est un ensemble de trois mondes virtuels destiné à améliorer la compréhension par des collégiens de concepts physiques, en leur donnant à percevoir des phénomènes invisibles dans le monde réel [Salzman et al., 1996] [Dede et al., 1999]. Les interfaces utilisées pour immerger les étudiants sont un rendu visuel stéréoscopique, un rendu audio stéréo, une souris 3d et une veste délivrant des sensations vibro-tactiles à l'aide de caissons de basses. NewtonWorld est un environnement permettant d'explorer les lois cinématiques et dynamiques du mouvement unidimensionnel (conservation de l'énergie et dynamique des déplacements). La scène est un corridor formé par des colonnes et terminé par un mur, dans lequel les étudiants peuvent lancer et attraper des balles de différentes masses, qui entrent en collision. Ils peuvent également se téléporter en différents points de vue. Un exemple de scénario multimodal est le lancement simple d'une balle. Avant le lancement (figure 2.10), l'énergie potentielle est représentée par un ressort et des vibrations dans la veste. Lorsque la balle est lancée, l'énergie potentielle se transforme en énergie cinétique, ce qui stoppe le rendu précédent. Pendant le déplacement, la taille des ombres de la balle sont proportionnelles à l'énergie cinétique. Lors de l'impact, l'énergie redevient potentielle puis cinétique, ce qui se traduit par une brève vibration. Les colonnes aident également à juger les vitesses et les distances par des clignotements et des tintements lors du passage des balles. MaxwellWorld fournit l'occasion d'étudier les forces et champs électrostatiques ainsi que les potentiels électriques. Par exemple, les étudiants peuvent placer des charges

positives et négatives de différentes magnitudes dans l'espace 3d et visualiser ensuite les lignes du champ électrique induit par leur configuration. PaulingWorld permet d'examiner des structures moléculaires sous différentes représentations, ainsi que leurs interactions.

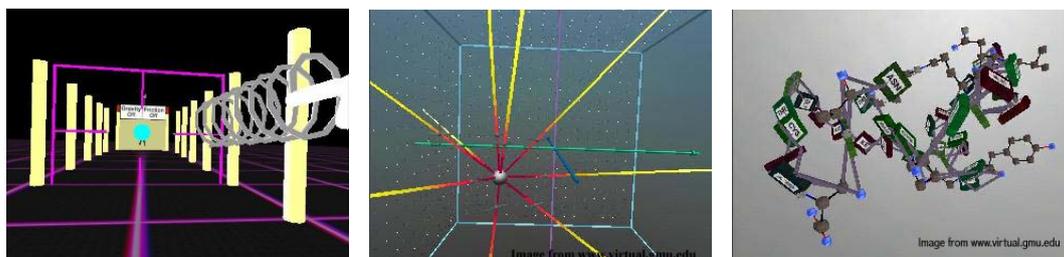


FIG. 2.10 – ScienceSpace : énergie potentielle avant le lancement dans NewtonWorld, lignes de champ pour une charge unique dans MaxwellWorld, représentation atomique et acides aminés dans PaulingWorld.

Dans le domaine de la biologie, Jones [Jones et al., 2004] a couplé un joystick haptique à un microscope à force atomique (AFM), permettant ainsi à des étudiants de manipuler des virus vivants. Ceci leur a permis de mieux appréhender les modèles de virus, comme par exemple leur morphologie spatiale (et non pas seulement en deux dimensions) ou à mieux se rendre compte de leur échelle. Minogue a appliqué le même principe pour des écoliers et des cellules animales [Minogue et al., 2006]. Pour l'étude de l'ADN, Garcia [Garcia-Ruiz et al., 2004] a combiné un rendu visuel 3d de la molécule, du texte et des earcons pour identifier les bases.

En chimie, Harvey et Gingold [Harvey and Gingold, 2000] ont développé une simulation visuo-haptique pour pallier la difficulté qu'ont les étudiants d'intégrer mentalement une ou plusieurs images pour se créer une représentation de l'orbite atomique. L'haptique permet de percevoir la densité des électrons en complément des méthodes de visualisation traditionnelles (isosurfaces, couleurs, plans de coupes, etc.). Richard [Richard et al., 2006] est l'une des rares à combiner le rendu haptique (Spidar), le rendu audio et le rendu olfactif, appliqué à l'étude du modèle atomique de Bohr et à la quantification énergétique.

Formation Dans le domaine de la formation professionnelle, Virtual Technical Trainer [Crison et al., 2005] (figure 2.11) vise à faire découvrir à des personnes en formation les principes de base de l'usinage et à leur faire construire les représentations mentales des processus. Ce qui est intéressant ici est l'utilisation de la modalité haptique sur un périphérique dédié pour faire ressentir physiquement le processus d'usinage, de manipuler directement et de comprendre le comportement de l'outil de coupe sur la pièce et les efforts en jeu. L'accent est mis ici sur l'ancrage sensorimoteur et la simplification de la tâche. L'intérêt est également la déportation de

l'utilisateur de sa tâche future réelle i.e. la commande d'une machine numérique d'usinage, vers l'outil d'usinage, afin de ressentir, de comprendre les efforts et les contraintes de l'outil sur la pièce usinée.



FIG. 2.11 – Multimodalité pour l'apprentissage : Virtual Technical Trainer (Clarté) et interface visuo-haptique pour l'apprentissage scientifique (Univ. Colorado).

Recherche Aux frontières de l'enseignement et de la recherche, Infed étudie la combinaison des rendus visuel et haptique pour la compréhension des champs électromagnétiques, de la mécanique des fluides, ainsi que pour des tests d'efforts sur des matériaux [Infed et al., 1999]. Dans ces trois problèmes, la visualisation se montre limitée dans certains cas, par exemple certaines données sont invisibles selon certains points de vue, et une rotation de la scène ne suffit pas à les comprendre dans leur globalité. Une interface haptique spécifique est utilisée pour rendre les champs de vecteurs et de tenseurs sous forme vibratoire et kinesthésique (figure 2.11). Ces travaux sont à mi-chemin entre l'apprentissage, puisque les utilisateurs ciblés sont des étudiants, et l'exploration de données multidimensionnelles et massives, que nous allons voir dans la section suivante.

Gherbi et al. [Gherbi et al., 2006] citent également l'utilisation de la multimodalité pour la compréhension de la relativité. Ainsi, on peut d'une part simuler les images que verrait un utilisateur naviguant à 98% de la vitesse de la lumière, mais encore imaginer associer des émissions sonores pour souligner les variations d'écoulement temporel entre les objets au repos et l'observateur relativiste, ou bien renforcer par l'haptique l'inertie des objets lorsque leur vitesse approche celle de la lumière.

2.3.2 Exploration de données

Tout comme celui de l'apprentissage, le domaine de l'exploration de données est vaste. Ils ne diffèrent pas beaucoup sur les plan des interactions et des modalités envisagées, mais plutôt sur le public et les objectifs visés. L'exploration de données

concerne principalement les recherches scientifiques ou industrielles : simulations d'écoulements gazeux ou liquides, pour l'aéronautique ou le développement durable, exploration géophysique, pour l'exploitation de forages pétroliers ou les études sismiques, observation de données médicales, etc. Mais ce domaine englobe aussi l'analyse de données abstraites : étude statistique des échanges boursiers, surveillance de variables d'état pour le fonctionnement d'une unité de production, debuggage de programmes informatiques, etc.

Même si ces travaux ne se déroulent pas tous dans le cadre de la RV immersive, il convient de s'y intéresser. A leur apparition, les applications scientifiques en RV étaient principalement des adaptations des applications bureautiques. Les choix de représentations visuelles étaient repris en immersif, avec les contraintes et les avantages supplémentaires de la RV : stéréoscopie, interaction, superpositions visuelles, etc. Il y a donc eu un double choix de visualisation à faire. Le premier est issu de l'application d'origine : quelle modalité pour quelle information. Le second est lié au passage à la RV : comment adapter la modalité d'origine. L'analyse de données en géoscience en est un exemple.

Cependant, on souhaite s'intéresser ici plus précisément à l'utilisation de plusieurs canaux sensori-moteurs pour l'exploration de données. L'utilisation des modalités visuelles est bien sûr prédominante. L'audio est également répandu et largement étudié. Vient ensuite l'haptique, dont l'usage est beaucoup plus récent. L'utilisation des trois canaux est encore rare. Pour toutes les applications d'exploration de données, il s'agit en général d'améliorer le rendu visuel, non optimal pour certaines tâches, par du rendu audio et/ou haptique.

2.3.2.1 Sonification

Nous l'avons vu, la sonification de données est la représentation de données dans le domaine acoustique pour compléter, améliorer la communication des informations. Les travaux dans ce domaine sont très nombreux.

Electronique Rabenhorst [[Rabenhorst et al., 1990](#)] s'est intéressé à l'utilisation du son pour l'interprétation de données multi-dimensionnelles complexes. Il applique son étude à la simulation de semi-conducteur. Trois champs scalaires sont calculés : densité des électrons, densité des trous et potentiel électrostatique. Afin de comprendre les relations spatiales entre ces trois champs, deux champs sont visualisés. Ensuite, l'utilisateur déplace un curseur 3d sur l'un des champs, tandis le gradient 3d de l'autre champ au même point est sonifié. Un travail important est effectué sur la méthode de sonification afin qu'elle soit non seulement complémentaire de la visualisation, mais qu'elle soit également assez intuitive et assez rapide à apprendre

pour permettre d'apporter quelque chose aux représentations mentales de l'observateur. La sonification est conçue afin de pouvoir déterminer les minima et maxima locaux du champ sonifié sans avoir à le visualiser. Chaque composant (signé) du vecteur à sonifier est projeté à la fois sur la balance stéréo et sur le ton d'une note, ce qui produit une suite de trois notes. La balance permet de percevoir le signe du composant mais sans grande précision de valeur, tandis que le ton raffine cette perception de valeur.

Biologie moléculaire Dans le domaine de la biologie moléculaire, les protéines ont fait l'objet de sonifications sur le modèle du morse. Les acides aminés sont associés à des tons [Dunn and Clark, 1999] ou des musiques courtes [Shi et al., 2007] afin de constituer la "musique de la protéine". Dans [Dunn and Clark, 1999] le registre (grave/aigu) est également déterminé par la solubilité des acides aminés, les moins solubles ayant une tonalité grave. Ces études sont toutefois plus proches de démarches artistiques que scientifiques.

PROMUSE est un projet de l'Université de Californie qui utilise la combinaison de la représentation visuelle et de rendus audio pour l'analyse de l'alignement structural des protéines [Hansen et al., 1999]. Ces connaissances sont utiles pour comparer les protéines entre elles, construire des modèles de leur évolution, ou comprendre les relations entre leurs formes et leurs fonctions. Cela permet également d'étudier quels acides aminés peuvent être retirés sans que la structure globale change, ce qui conduit à améliorer la prédiction des structures des protéines inconnues. La sonification se veut musicale pour réduire la fatigue et la gêne auditive. Quatre paramètres de la protéine sont examinés : la polarité, la structure secondaire, l'exposition et le "goodness-of-fit". Chacun est projeté sur un instrument de jazz, composant ainsi un quartet. Selon les trois valeurs possibles pour chaque paramètre, il y a trois instruments possibles mais qui gardent le même rôle dans le quartet (respectivement : accompagnement, basse, batterie, mélodie). Les sons sont également distribués sur quatre positions de l'espace entre les diffuseurs sonores droit et gauche (ici un casque). Les tests ont été réalisés en visualisant une protéine sous Rasmol et en demandant aux sujets de cocher la bonne caractéristique pour chacun des quatre paramètres (figure 2.12). Les résultats montrent que la perception des instruments facilite la reconnaissance de certains cas ambiguës.

Mécanique des Fluides En Mécanique des Fluides, Childs sonifie une étape particulière du processus d'analyse d'écoulement [Childs, 2001] [Childs, 2002]. En effet, un écoulement est régi par une équation, qu'un ordinateur va tenter de résoudre itérativement, en convergeant vers une solution. La sonification peut être utilisée pour suivre l'évolution de la solution et également contrôler les paramètres

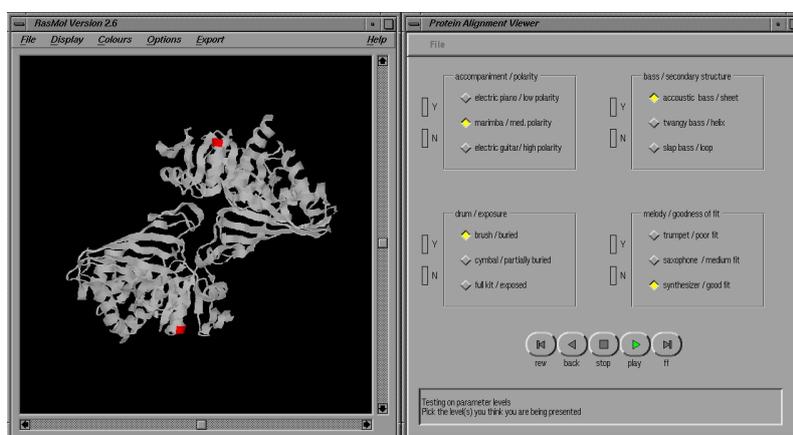


FIG. 2.12 – Visualisation et sonification de protéines dans le logiciel PROMUSE (Univ. de Californie).

de la simulation. Sur un écoulement simple et connu, Childs conçoit une stratégie de sonification (ton, enveloppe, timbre, durée) pour trois variables : la vitesse, la composante transverse de la vitesse et la pression.

McCabe [McCabe and Rangwalla, 1994] s'intéresse au même domaine, mais pour l'analyse des résultats de la simulation.

Médecine Dans le domaine médical, Hermann [Hermann et al., 2000] présente une méthode pour contrôler une multitude d'images par le son. Ces travaux sont appliqués à l'évaluation de piles d'images médicales issues d'un microscope à fluorescence. La recherche se fera en faisant défiler les images tout en écoutant la sonification, qui facilitera l'identification des motifs et la discrimination de motifs similaires.

Fitch [Fitch and Kramer, 1994] utilise la sonification afin de compléter le rendu visuel de certains paramètres physiologiques d'un patient (fréquence de respiration, température corporelle, fréquence cardiaque) par un rendu sonore approprié.

Programmation L'une des applications d'origine de la sonification est la programmation informatique [Jameson, 1994]. Le son est utilisé comme outil de débogage, par exemple pour repérer des boucles infinies (son qui ne s'arrête pas) ou une fuite mémoire (son dont la fréquence augmente sans arrêt). Le projet Sonnet d'IBM poursuit dans cette voie. C'est un langage de programmation visuelle, qui permet de développer des applications temps réel en connectant des blocs fonctionnels. Les événements au sein des programmes ainsi créés sont convertis en signaux sonores ce qui permet à l'utilisateur de surveiller le déroulement et de réagir à des imprévus, de la même manière qu'un conducteur ne prête attention au son du moteur que

s'il fait un bruit anormal. Par exemple, dans le domaine boursier, un bloc Sonnet pourrait recevoir une cotation, la transmettre à un autre bloc qui la convertirait en un son. Les investisseurs pourraient ainsi développer une perception des événements financier et réagir auditivement plutôt qu'à l'aide d'une analyse. Sonnet peut aussi être utilisé dans les systèmes temps réel pour répondre aux problèmes de parallélisme (problème également traité par Jackson [[Jackson and Francioni, 1994](#)] dans ou d'applications distribuées.

Peep [[Gilfix and Couch, 2000](#)] est un logiciel qui permet la sonification de réseaux, remplaçant ainsi la surveillance visuelle de l'activité. Des sons sont associés à des événements spécifiques (charge, trafic excessif, spam, etc.) et l'utilisateur peut alors facilement les détecter lorsqu'ils se produisent.

Conception industrielle Dans le domaine de la conception industrielle, Eckel [[Eckel, 1998](#)] a effectué des travaux pour la conception automobile, en particulier pour le problème du positionnement des conduits de ventilation dans une voiture. Un modèle 3d permet de visualiser l'habitacle ainsi que la disposition des flux d'air, tandis qu'une sonde peut être disposée dans l'espace afin d'écouter l'amplitude du flux sous la forme d'un bruit de vent plus ou moins puissant.

Geophysique Enfin, divers travaux existent dans les champs de la géophysique. Barass [[Barass and Zehner, 2000](#)] a employé des signaux audio pour rendre les propriétés de données pétrolières, sur la métaphore du compteur geiger. L'utilisateur observe le modèle visuel du puit de pétrole, coloré selon certaines données, tout en écoutant les caractéristiques d'autres données à l'aide d'une sonde audio.

Des travaux sont également conduits en volcanologie. Les projets EGEE et EELA ont servi à la sonification de données sismiques provenant de volcans (Etna en Sicile et Tungurahua en Equateur) afin de prévoir leurs éruptions. Les signaux issus de sismographes numériques sont échantillonnés puis traités afin de les rendre audibles tout en conservant leurs motifs d'origine. Diverses sonifications sont ensuite possibles, par exemple des morceaux de piano. Ainsi, la partition issue des données de l'Etna est rapide, trépidante, avec des pauses fréquentes, alors que la partition du Tungurahua est plus lente et plus fluide.

Le projet SonEnvir [[de Campo et al., 2005](#)], débuté en 2005, réunit quatre universités autrichiennes. Il aborde la sonification dans de nombreux domaines scientifiques, pour lesquels une même plate-forme logicielle est utilisée. Un thème original est l'analyse de données sociologiques. Par exemple, pour la conception de toilettes pour personnes handicapées, diverses données sur les interactions des usagers avec un prototype ont été enregistrées dans des fichiers. Puis les différents événements ont

été sonifiés : ouverture et fermeture de porte par un bruit ambiant, hauteur et inclinaison du siège par un ronronnement dont le ton varie, actionnement des commandes de réglage du siège par des *glissando*, chasse d'eau par un bruit caractéristique et alarme par une cloche. Ces sons sont accompagnés par une représentation schématique (fenêtre et widgets). L'étude de ces enregistrements permet de comparer les comportements, former des groupes d'utilisateurs et ainsi améliorer le prototype. D'autres thèmes scientifiques sont abordés : la physique théorique (classification des baryons), la neurologie (surveillance de patients en temps réel, grâce à la sonification d'électro-encéphalogrammes, par exemple pour préciser un diagnostic d'épilepsie), le traitement du signal (classification de bruits), etc.

2.3.2.2 Haptisation

L'haptisation des données est particulièrement utile lorsque ces données sont volumiques et/ou qu'elles ne peuvent pas être visualisées correctement. Ces données sont le plus souvent statiques, et ne sont pas modifiées par l'interaction, car il faudrait alors recalculer toute la simulation en temps réel tout en respectant la fréquence de rafraîchissement haptique.

Une large majorité des travaux traite de l'exploration d'écoulements, qu'ils soient liquides ou gazeux, ou de champs de forces. D'autres exploitent l'haptique pour améliorer la qualité, la finesse d'une exploration.

Micro-manipulation Seeger [Seeger and Chen, 1997] propose d'utiliser une interface haptique pour contrôler la sonde d'un microscope et ainsi percevoir des informations supplémentaires sur les interactions entre la sonde et les surfaces observées. Le système permet de visualiser de manière stéréoscopique des surfaces, colorées en fonction de différentes données. L'utilisateur peut également ressentir la forme de la surface et le frottement de la sonde.

Mécanique des Fluides Bourdot [Bourdot et al., 2006a] propose d'haptiser des isosurfaces lors de l'exploration d'un écoulement en Mécanique des Fluides. L'objectif est de ressentir les aspérités des données et les particularités de la simulation, par exemple la direction des vitesses d'écoulement.

En aéronautique, Lundin [Lundin et al., 2005] utilise le rendu haptique pour combler les lacunes de la visualisation volumique (occultations et surcharge d'informations) et améliorer la compréhension de l'aérodynamique. Le rendu visuel est composé d'un rendu volumique de la pression autour de l'appareil, et de rubans colorés représentant les flux d'air, qui peuvent être positionnés librement dans l'espace

3d. L'haptique possède différents modes qui peuvent être actifs ou inactifs. Le premier permet de ressentir la surface de l'avion. Le second mode guide l'utilisateur le long du flux d'air autour du fuselage : le mouvement est libre si l'utilisateur déplace le stylet dans le sens du flux, et visqueux dans le cas contraire. Le dernier mode permet de ressentir les tourbillons (vortex) (figure 2.13).

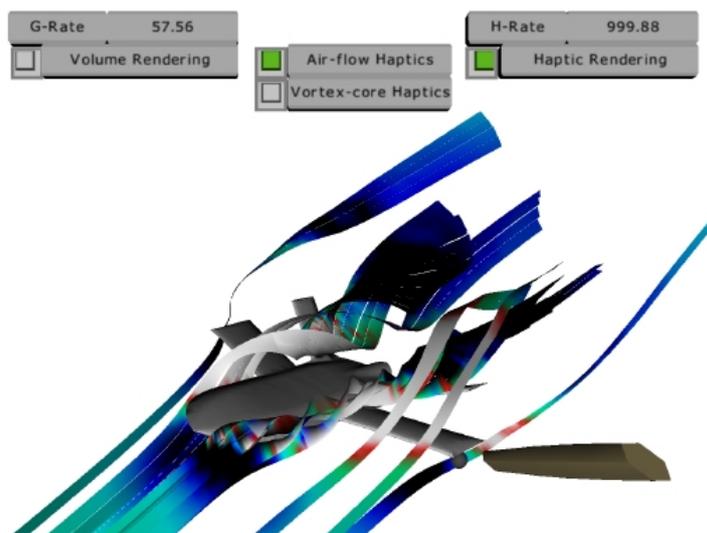


FIG. 2.13 – Haptisation de données de Mécanique des Fluides : le stylet permet de suivre les flux d'air autour du fuselage de l'avion (NVIS, Suède).

Biologie moléculaire Maciejewski [Maciejewski et al., 2005] étudie l'apport de la "perceptualisation multimodale" au docking moléculaire. La tâche consiste à trouver la meilleure configuration entre deux protéines (actine et cofiline) à l'aide d'un rendu visuel stéréoscopique et d'un rendu haptique. L'une des particularités de cette étude est l'emploi de plusieurs fonctions transfert entre la valeur à rendre et la raideur du rendu haptique à l'aide d'un procédé interactif IVPS (Interactive Volume Perceptualization System). Les résultats montrent cependant que, dans les conditions de l'expérience, l'ajout de l'haptique ne diminue pas les erreurs de docking, tout en augmentant la durée de la tâche. La cause est imputée aux fonctions transfert inadaptées.

Hansen, Lodha et al. ont prolongé leur travail abordé dans PROMUSE (section précédente) dans le logiciel PROFEEL, qui permet l'analyse des alignements structurels protéiniques à l'aide d'un joystick à retour d'effort [Hansen et al., 2000]. L'information de structure secondaire crée le mouvement du joystick (linéaire, rotatif ou irrégulier). L'exposition correspond à une vibration à 25Hz avec une variation

périodique de l'amplitude. La polarité est une métaphore du compteur Geiger avec des "clics" dont la fréquence varie. Enfin, comme la variable "goodness-of-fit" suit le modèle physique de la température, qui elle-même correspond à une vibration plus ou moins forte de molécules, elle est transmise par une variation de fréquence élevée.

Geophysique McLaughlin [McLaughlin and Orenstein, 1997] utilise le retour d'effort pour assister l'interprétation sismique de données géophysiques. Cette interaction aide l'utilisateur à ressentir des éléments subtils et occultés par le modèle visuel.

2.3.3 Assistance en environnement virtuel et augmenté

L'un des rôles du système informatique en RV&A est d'assister l'opérateur dans ses tâches, que ce soit l'aide à la sélection et la manipulation d'objets virtuels ou l'apport d'informations pour améliorer la compréhension d'un problème.

Interfaces Les interfaces graphiques des logiciels informatiques sont parfois complexes à maîtriser par les utilisateurs, à cause par exemple de la profusion des widgets¹, ou de la difficulté à les sélectionner. C'est l'objet de certaines des recherches du Multimodal Interaction Group de l'Université de Glasgow. Leurs travaux se portent d'une part sur la sonification [Lumsden et al., 2002], d'autre part sur l'haptisation de ces interfaces [Oakley et al., 2002]. La sonification est établie grâce à des earcons (dont la traduction officielle est "carillon") non vocaux associés soit à des widgets particuliers, soit à des événements que ces widgets génèrent. Ces événements peuvent être une action (i.e. un résultat immédiat de l'interaction) ou bien un état (i.e. la progression d'une tâche en cours). La sonification permet dans un cas de renforcer le widget visuel, dans l'autre d'informer uniquement auditivement de l'événement en libérant le canal visuel de cette charge. Dans le cas où les widgets sont nombreux, les auteurs préconisent une cohérence des earcons dans l'interface, à la manière des styles de bureau d'un O.S., ainsi qu'une attention particulière sur les compatibilités entre les sons. Dans le cas contraire, le risque est que la sonification soit ressentie comme une gêne plutôt qu'une aide. Dans le cas de l'haptisation, ici encore la multimodalité peut réduire les performances des utilisateurs si elle n'est pas conçue correctement. Différentes stratégies de retour d'effort sont possibles. Les widgets peuvent ainsi devenir des boutons poussoir à enclencher ou être entourés de frontières haptiques qui les délimitent. L'interface peut également attirer l'utilisateur

¹Un widget est un composant d'interface graphique avec lequel l'utilisateur peut interagir. Il peut être un objet comme un bouton, un menu... mais également une fonctionnalité, comme le glisser-déposer.

vers la cible ou freiner son mouvement lorsqu'il s'en approche, ce qui suppose alors une inférence sur cette cible.

Dans le cadre de l'assistance à la navigation et à la manipulation d'objets, il existe le concept de grille haptique, transposé du concept de la grille dans les logiciels de CAO ou de dessin. Yamada [Yamada et al., 2002] crée ainsi une grille régulière de points 3d vers lesquels l'outil ou l'objet manipulé par l'utilisateur est attiré haptiquement. Ceci augmente la précision du geste et le positionnement de formes en trois dimensions. Fritz [Fritz, 1996] utilise le même concept, mais cette fois-ci pour renforcer la perception des distances et des échelles de déplacement chez les aveugles. Une grille composée de plans 3d est haptisée : lorsqu'un plan est traversé, l'utilisateur ressent une légère force de frottement. Picon [Picon et al., 2007] a appliqué divers profils d'attraction, mais cette fois-ci pour assister à la sélection des éléments topologiques des objets (faces, arêtes, sommets), avant leur modification.

Réalité Augmentée En dehors des environnements informatiques ou virtuels, les techniques d'assistance peuvent également être appliquées en environnement augmenté. La Réalité Augmentée (RA) est, selon la définition de Milgram, une portion du continuum réel-virtuel (figure 1.6 p.12). Elle a pour objectif "d'augmenter la rétroaction naturelle de l'opérateur avec le monde réel à l'aide d'indices virtuels" (traduction de Didier dans [Didier, 2005]). Fuchs [Fuchs and Moreau, 2004] généralise cette définition : "la RA regroupe l'ensemble des techniques permettant d'associer un monde réel avec un monde virtuel". De notre point de vue, la RA est un cas particulier de la RV. Celle-ci vise à mettre au point des systèmes informatiques qui donnent à l'humain la capacité de percevoir et d'interagir de façon multi-sensori-motrice avec des données numériques ou mondes virtuels. Quand en plus, ces données numériques intègrent une "virtualisation" d'une partie de l'univers réel et permettent ainsi de gérer des interactions entre des objets totalement virtuels et des objets réels, on parle alors, tantôt de Réalité Augmentée, tantôt de Virtualité Augmentée, voire de Réalité Mixte, selon la part des données virtuelles et réelles.

La RA est donc par essence un domaine qui fait appel à des choix de représentations multimodales. L'objectif essentiel est d'ajouter du contenu informationnel à la perception du monde réel, ce qui pose deux questions principales : quelles informations apporter à l'utilisateur en plus de ce qu'il perçoit déjà ? sous quelle forme lui présenter ces informations ?

Une partie des informations à apporter à l'utilisateur est l'ensemble des informations d'assistance, pour diverses tâches sur des données et des objets réels. Ainsi, pour les applications de maintenance qui sont typiques de la RA, le système peut apporter à l'opérateur des informations de manière temps réel et interactive. Reiners a travaillé par exemple sur la tâche d'insertion d'un verrou dans une portière automobile [Reiners et al., 1998]. La RA permet de mettre en situation un ouvrier

sur le matériel réel afin de lui apprendre les gestes spécifiques à effectuer. Le système affiche sur des lunettes immersives la scène réelle et montre successivement à l'opérateur la pièce à insérer, l'endroit où la pièce devait venir, la trajectoire d'insertion ainsi que le sens de vissage des divers écrous pour fixer la pièce (figure 2.14).



FIG. 2.14 – Multimodalité pour la Réalité Augmentée : aide au montage (Fraunhofer Institute).

Une autre objectif peut être de remplacer une documentation technique papier volumineuse et peu pratique par des informations contextualisés visuelles, ce qui libère les mains de l'opérateur. C'est l'objet des travaux de Dutoit [Dutoit et al., 2001], appliqués aux centrales nucléaires. Un casque et un écran tactile permettent d'afficher la liste contextualisée des opérations successives à accomplir pour effectuer une tâche.

Nano-opération et Télé-opération L'assistance peut également prendre une forme audio ou haptique. Ammi [Ammi, 2005] a implémenté des métaphores de guidage de l'opérateur pour la manipulation de micro-billes à l'aide de la pointe d'un microscope à force atomique (AFM). L'haptique est utilisé pour aider le déplacement des billes d'une configuration initiale à une configuration finale. Dans un premier mode, l'opérateur est attiré sur la trajectoire optimale qui minimise les déplacements et évite les obstacles. Un second mode existe dans lequel les obstacles créent des

champs de forces qui informe l'utilisateur de leur présence. En ce qui concerne la sonification, elle sert à communiquer des grandeurs physiques, comme la distance entre la pointe AFM et la bille cible ou l'effort exercé par la pointe. La modulation du son est fonction de la fréquence ou de l'amplitude de la grandeur. Un second mode consiste à ajouter du "pseudo-son" au micro-environnement. Des icônes auditives spatialisées communiquent les événements physiques qui s'y déroulent, comme les frottements ou les contacts.

Dans le domaine de la téléopération/téléprésence, le projet SACARI porte sur la supervision d'un véhicule autonome à partir d'un environnement immersif multimodal [Tarault et al., 2005]. En téléopération, le téléopérateur perçoit l'environnement distant et doit réagir et effectuer des commandes comme s'il était sur le site distant. Le projet a entre autres donné lieu à une étude de la localisation de sources audio distantes [Katz et al., 2007]. Dans SACARI, le son est capturé sur le véhicule distant, et doit être transmis et restitué de manière spatialisée à l'opérateur dans un environnement immersif. L'étude a porté sur une tâche de recherche de cibles audiovisuelles dans l'espace, le son aidant à la localisation visuelle. Elle a comparé deux méthodes de capture, binaurale et ambisonique, couplées à une méthode de restitution binaurale. Les résultats ont montré une capacité certaine de la perception auditive à aider à la localisation des sources dans l'espace (les temps de ciblage étant inférieurs en présence de rendu sonore) ainsi que des différences entre les deux techniques (temps inférieurs pour la capture binaurale).

2.3.4 Discussion

Les applications que nous venons de présenter ont toutes le même objectif : exploiter au mieux les spécificités des trois principaux canaux sensori-moteurs, les capacités cognitives humaines et la grande variété des techniques d'interaction de RV pour améliorer la performance ou le confort des utilisateurs ou leur apporter de nouvelles fonctionnalités. Celles-ci peuvent aider l'apprentissage, la formation, la compréhension de phénomènes. Elles peuvent éclairer d'une autre manière des problèmes complexes, dont l'étude est limitée en mode visuel. Elles peuvent améliorer les tâches des opérateurs en environnement augmenté ou virtuel. La plupart des expériences ont montré un apport certain de la multimodalité aux utilisateurs, que ce soit une augmentation du savoir, du savoir-faire ou des performances.

La figure 2.15 résume la part des canaux sensoriels utilisés dans les travaux rencontrés. Les applications combinant le visuel avec un autre canal sont de plus en plus nombreux. Les domaines les plus concernés sont la Biologie Moléculaire et la Mécanique des Fluides. Bien qu'il n'existe que peu de travaux combinant les trois canaux, il semble cohérent d'imaginer que des applications de ce type sont possibles et vont voir le jour, tant il y a des thèmes et des expériences récurrents.

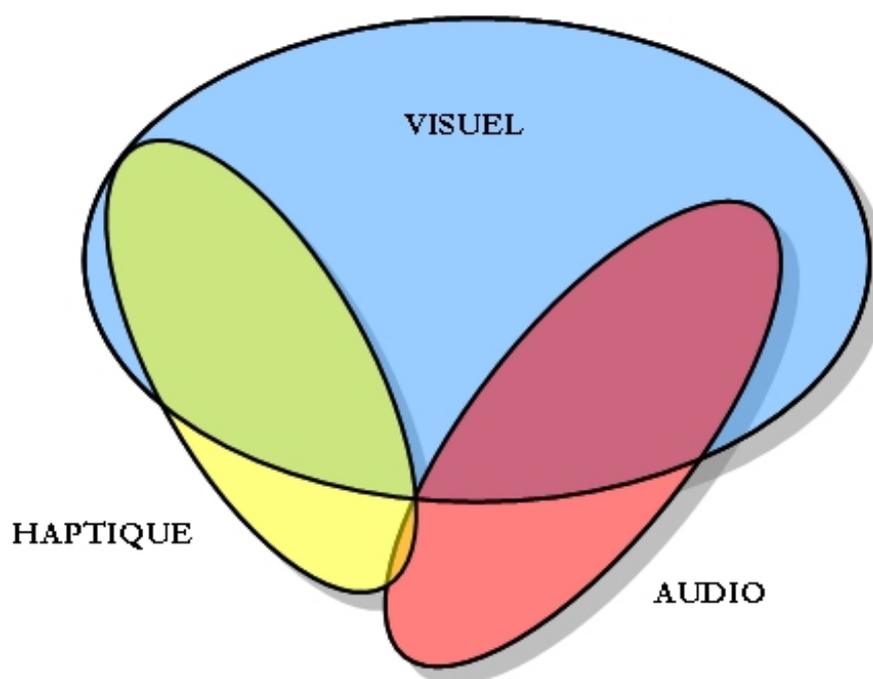


FIG. 2.15 – Répartitions des applications multi-sensorielles orientée tâches existantes.

Au fil de cet état de l'art des applications multimodales, des réponses aux problèmes de la multimodalité en sortie émergent. Tout d'abord, il apparaît que les domaines et les types de données qui peuvent être communiquées à l'utilisateur de manière multimodale sont nombreux. Ensuite, les travaux comportant une évaluation montrent régulièrement un apport de la (ou des) modalité(s) supplémentaire(s), même si certains révèlent parfois une baisse des performances, par exemple une augmentation de la durée de la tâche lors de l'ajout de l'haptique. Toutefois, les bénéfices subjectifs et ergonomiques de la multimodalité pour les utilisateurs et les apports à long terme pour les domaines scientifiques ou industriels cités sont pour l'heure encore inconnus.

Même si ces travaux présentent une utilisation conjointe des rendus visuels et sonores ou haptiques, ou du remplacement du visuel par l'un ou l'autre des deux canaux, il faut noter qu'aucun ne traite de la combinaison réelle des modalités, en particulier des effets qu'elles ont les unes sur les autres. Il faut également ajouter que si le rendu multimodal s'est avéré pertinent dans ces études, aucune n'a suivi ou mené à un modèle de la multimodalité en sortie, ou à un cahier des charges des combinaisons multimodales à faire ou à ne pas faire. Enfin, les rendus ont dans tous les cas été choisis à l'avance par les concepteurs, en fonction des utilisateurs

visés et du domaine, mais jamais le rendu n'est dépendant du contexte, le système ne s'adapte pas automatiquement à de nouvelles situations. Nous pouvons donc dire que l'"intelligence" qui fait de cette multi-sensorialité de la multimodalité reste humaine : la pertinence du rendu est garantie par les concepteurs pour des situations figées, mais le système n'est pas adaptatif.

La suite de l'état de l'art ainsi que les réalisations de cette thèse se concentrent sur les trois limites identifiées précédemment. Tout d'abord, même si les choix de projections entre les informations à transmettre et les modalités visuelles, sonores ou haptiques varient d'une étude à l'autre (en fonction du domaine, des objectifs et de l'activité de recherche), des correspondances, des points communs apparaissent et pourraient être répertoriées au minimum dans un "cahier des charges", au mieux dans un modèle du rendu multimodal. Ces correspondances peuvent être l'utilisation d'un même rendu pour communiquer un même type d'information dans différents domaines, ce qui mène à penser que ce choix de rendu est adéquat. L'utilisation de guide haptique pour l'assistance au geste, ou la variation du ton d'un son pour une variable réelle dont on veut connaître la variation en sont deux exemples. On peut remarquer également pour le même type de tâche des choix de rendus différents, qui s'ils étaient évalués bénéfiques pourraient constituer des alternatives, et donc conduire à de la substitution multimodale validée applicativement. D'autre part, même si les travaux présentés précédemment n'ont pas fait appel à des modèles du rendu multimodal ni à des études psychophysiques, il en existe. C'est l'objet des sections suivantes.

2.4 Modèles du rendu multimodal

2.4.1 Espaces de conception

Un espace de conception est une structure permettant d'organiser le raisonnement, de clarifier les concepts, de classifier des éléments, de prévoir les cas possibles, etc. dans un domaine donné. Le plus souvent, il précède la conception proprement dite d'un système. Dans le domaine des interfaces multimodales, les espaces de conception sont aussi nombreux que les définitions de la multimodalité. Ils visent à clarifier la terminologie utilisée, classifier les interfaces passées et présentes, afin d'aider au design d'interfaces futures. Ils essaient notamment d'identifier les médias, les modalités et leurs caractéristiques, les combinaisons possibles ou les éléments constitutifs du contexte.

Schomaker et al [[Schomaker et al., 1995](#)] développent toute une taxonomie de l'IHM dans le contexte du projet MIAMI (Multimodal Integration for Advanced Multimedia Interfaces). L'IHM est tout d'abord la relation entre deux agents : un

humain et une machine (figure 2.16). Chacun de ces agents peuvent échanger des informations à travers deux processus basiques : la perception d'un côté et le contrôle d'un autre. Ceci mène à distinguer quatre composants : les canaux d'entrée humains, les canaux de sortie humains, les modalités d'entrée du système et les médias de sortie du système. Au sein de chaque agent a lieu un processus de "cognition" chargé de traiter les entrées et les sorties. Pour chacun de ces quatre composants, quatre niveaux d'observation sont définis :

- niveau physique ou physiologique (respectivement pour les médias ou les canaux) : les caractéristiques des dispositifs techniques et des canaux sensorimoteurs humains (résolutions des appareils, mesures physiques de la perception, bandes passantes sensorielles, etc.) constituent les limites et les possibilités de l'interaction ;
- niveau des informations théoriques : les capacités sémantiques de transmission des informations des canaux et des médias ("bande passante informatique") ;
- niveau cognitif : permet de clarifier de manière syntaxique et sémantique les aspects représentationnels et procéduraux (langage utilisé, mémoire à court terme, apprentissage) ;
- niveau intentionnel : explication des buts et des croyances de chaque agent.

Les auteurs développent chacun de ses aspects, ce qui permet de classifier et de caractériser les dispositifs et les techniques générales qui existent ou existeront en IHM. Ils mettent en place une partie des acteurs qu'il est nécessaire d'étudier pour exploiter le rendu multimodal.

Bernsen [Bernsen, 1994] présente une catégorisation des modalités de sortie en IHM. Dans son modèle, chaque modalité est représentée par un couple (média, profil) où le média est le support sensoriel de l'information (image, son, toucher) et le profil est un ensemble de quatre propriétés binaires qui caractérisent la modalité :

- linguistique / non linguistique : une modalité linguistique est basée sur un langage, c'est-à-dire sur des relations conventionnelles entre des éléments lexicaux et syntaxiques qui font sens ;
- analogique / non analogique : une modalité analogique (on parle également de modalité iconique ou isomorphique) ressemble à la réalité, s'inspire de l'existant, le rendu fait l'objet d'une transposition du réel vers l'informatique ;
- arbitraire / non arbitraire : une modalité arbitraire est différente de toute représentation existante, a besoin d'une connaissance préalable pour être interprétée ;
- dynamique / statique : une représentation dynamique change dans le temps.

Les combinaisons de tous les médias et les profils donnent donc 48 modalités basiques possibles. En enlevant les incohérences (une modalité analogique ou linguistique ne peut pas être arbitraire, une modalité de son ou de toucher ne peut pas être statique), Bernsen obtient 28 combinaisons possibles qu'il appelle "modalités pures" et pour

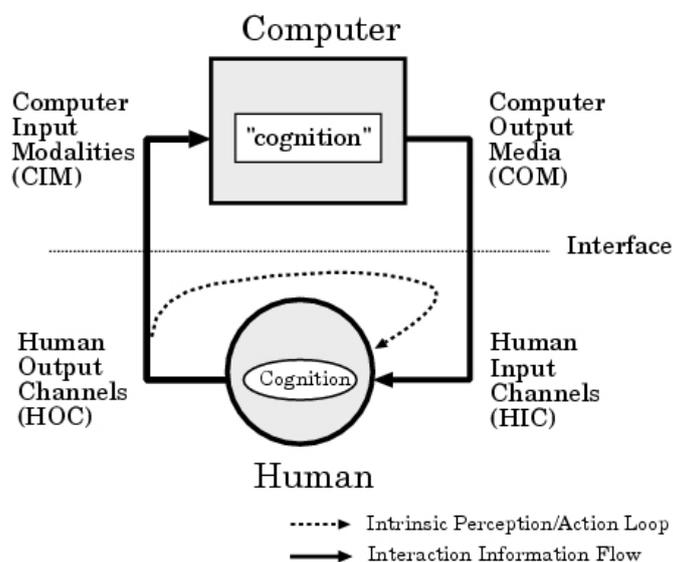


FIG. 2.16 – L'IHM : une double boucle Perception-Commande entre un agent humain et un agent système [Schomaker et al., 1995].

lesquelles il donne des illustrations concrètes quand cela est possible (figure 2.17). Par exemple, le profil d'un texte est (langagier, non analogique, non arbitraire, statique) tandis qu'une photographie est représentée par (non langagier, analogique, non arbitraire, statique) et un icône auditif par (non langagier, non analogique, arbitraire, dynamique).

Cette classification est utilisée dans de nombreux travaux en multimodalité. Toutefois, les quatre critères proposés ne correspondent pas forcément aux propriétés des modalités qui sont analysées lorsqu'un concepteur veut améliorer la tâche d'un utilisateur en RV. L'attention se porte sur la correspondance de ses propriétés avec celles des informations qui doivent être communiquées. Ainsi, le critère dynamique est approprié, mais il manque un critère de spatialité aux modalités. Les critères analogique et arbitraire ne sont pas non plus indépendants, et la RV ferait plutôt appel au caractère "naturel" d'un rendu ou à son affinité avec l'immersion (ce qui par exemple n'est pas le cas des affichages visuels proposés).

Bernsen aborde également le problème de la représentation optimale de l'information échangée entre le système et l'utilisateur pendant une tâche, qui est au coeur même du rendu multimodal. Il introduit le concept de "représentation multimodale", et considère qu'une modalité n'est pas une capacité sensorielle humaine mais un système de représentation. Ce concept peut être résumé par la relation : *Ce qui doit être représenté* \leftrightarrow *Principes de projection* \leftrightarrow *Représentations*. Cependant, Bernsen ne propose aucune règle de projection ni de méthode pour s'assurer que la

Modality	Well-known atomic types
1. Static analogue graphic language	Hieroglyphs. Rarely used.
2. Dynamic analogue graphic language	Gestural language. Dynamic hieroglyphs would appear anachronistic.
3. Analogue spoken language	Part of everyday spoken language.
4. Analogue touch language	Apparently none.
5. Static non-analogue graphic language	Letters, words, numerals, other written language signs, text, logograms (e.g. arrows), special-purpose notations (e.g., programming languages, formal logic, music). Sequential, list and tabular orderings.
6. Dynamic non-analogue graphic language	Dynamizations of (5). Graphically viewed spoken language discourse.
7. Non-analogue spoken language	Spoken letters, words, numerals, other spoken language signs, discourse. List orderings.
8. Non-analogue touch language	Touch letters, numerals, words, other touch language related signs, text, list and table orderings. Example: Braille.
9. Static diagrammatic pictures	Pure diagrams, maps, cartoons. Sequential, list and tabular orderings.
10. Static non-diagrammatic real-world pictures	Photographs, naturalistic drawings, holograms. Sequential, list and tabular orderings.
11. Static graphs	1D, 2D or 3D graph space containing geometrical forms or other elements. Pure charts (dot charts, bar charts, pie charts, etc.). Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
12. Animated diagrammatic pictures	Pure animated diagrams. Pure standard animations.
13. Dynamic real-world pictures	Pure movies, videos, realistic animations.
14. Dynamic graphs	Pure graphs (see 11) evolving in graph space. Non-iconic use difficult due to the absence of linguistic annotation.
15. Real-world sound	Sound 'pictures'. Sound signals. List orderings.
16. Diagrammatic sound	Sound 'pictures'. Sound signals. Many possibilities, synthetic or manipulated, exist. Music? List orderings.
17. Sound graphs	E.g. Geiger counters.
18. Real-world touch	Single touch representations, touch sequences.
19. Diagrammatic touch	Apparently none, but many possibilities exist.
20. Touch graphs	1D, 2D or 3D graph space containing geometrical forms. Pure charts (dot charts, bar charts, pie charts, etc.).
21. Arbitrary static diagrammatic forms	Diagrams consisting of geometrical elements. Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
22. Animated arbitrary diagrammatic forms	Diagrams consisting of geometrical elements. Non-iconic use difficult due to the absence of linguistic annotation.
23. Arbitrary sound	Sound signal icons.
24. Arbitrary touch	Touch signals of different sorts.
25. Static graphic structures	Form fields, frames, table grids, line separations, trees, windows, bars. Sequential, list and tabular orderings. Non-iconic use difficult due to the absence of linguistic annotation.
26. Dynamic graphic structures	Dynamic frames, windows, scroll bars. Non-iconic use difficult due to the absence of linguistic annotation.
27. Sound structures	Apparently none.
28. Touch structures	Form fields, frames, grids, line separations, trees.

FIG. 2.17 – Taxonomie des modalités de Bernsen [Bernsen, 1994] : modalités pures et exemples associés.

représentation multimodale est optimale.

Le groupe IIHM du laboratoire CLIPS-IMAG à Grenoble a mené de nombreux travaux constituant un vaste espace de conception orienté vers les problématiques logicielles et techniques soulevées par la conception d'interfaces multimodales.

Dans cet espace (figure 2.18), le problème central est la sélection d'une ou plusieurs modalités afin de communiquer des informations à un ou des utilisateurs.

Les modalités et leurs combinaisons forment l'espace des possibilités, tandis que l'information à véhiculer et le contexte constituent les critères du choix.

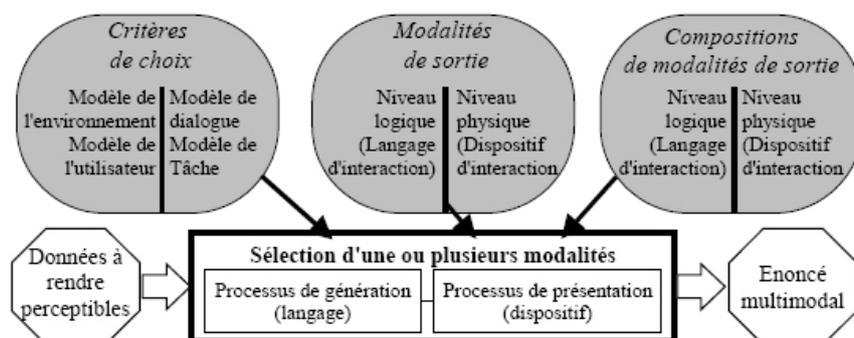


FIG. 2.18 – Espace de conception de la multimodalité en sortie [Vernier, 2001].

Les acteurs du choix, qui vont sélectionner la ou les modalités pour communiquer la ou les informations, sont le concepteur, le système ou l'utilisateur. Dans le cas où c'est le concepteur qui choisit la distribution, alors il n'y aura aucune adaptation possible du rendu. Lorsque l'utilisateur peut configurer le rendu, alors on parle d'adaptabilité du rendu (le système est adaptable). Par exemple, le possesseur d'un téléphone portable peut choisir que les appels soient signalés par une sonnerie ou une vibration. Enfin, si le système est capable de détecter le contexte et décide automatiquement du rendu en fonction de cette connaissance, on parle d'adaptativité du rendu (le système est adaptatif). Par exemple, on peut imaginer que le téléphone portable de l'exemple précédent perçoive le contexte physique de l'utilisateur, et que selon que l'environnement est une salle de cinéma ou une rue, décide d'utiliser respectivement le mode vibratoire ou le mode sonore.

Comme nous l'avons vu dans la section 1.3, ces auteurs définissent une modalité comme un couple (dispositif physique, langage d'interaction). Cette représentation de la modalité permet d'observer à la fois sa dimension physique (capacités sensorielles concernées, lieu d'interaction, temporalité du rendu i.e statique ou éphémère) et sa dimension logique, représentationnelle, reprise de Bernsen (linguistique, analogique et arbitraire). La sélection de l'énoncé multimodal se compose donc de deux étapes : la génération du rendu sous forme de langage, puis sa présentation à l'aide des dispositifs physiques.

Les projections possibles entre modalité(s) et information(s) ainsi que les relations entre ces modalités sont formalisées par les propriétés CARE : Complémentarité, Assignment, Redondance, Equivalence [Coutaz et al., 1995]. La complémentarité repose sur l'utilisation de plusieurs modalités distinctes représentant chacune une partie de l'information. L'assignation (ou spécialisation) restreint la représen-

tation d'une information à une unique modalité. La redondance est l'utilisation de plusieurs modalités dont au moins une représente la totalité de l'information. L'équivalence indique une alternative de représentation de l'information : plusieurs modalités peuvent être utilisées de manière comparable.

Ces quatre compositions possibles des modalités sont généralisées par Vernier [Vernier and Nigay, 2000] dans un tableau (figure 2.19) qui différencie le schéma et l'aspect de la composition. Les schémas sont issus des relations de Allen [Allen, 1983] et indiquent la manière dont les modalités se composent (séquence, enchaînement, recouvrement complet, recouvrement partiel et parallélisme) tandis que l'aspect représente la dimension sur laquelle elle se compose : temporel, spatial, articulatoire, syntaxique et sémantique. L'aspect temporel concerne les enchaînements possibles des modalités (c'est l'objet des travaux de Allen). L'aspect spatial porte sur l'agencement des modalités dans l'espace. L'aspect articulatoire s'adresse à la composition des flux d'information bas niveau transitant par les dispositifs associés aux modalités. L'aspect syntaxique s'intéresse au niveau des langages d'interaction associés aux modalités. Enfin, l'aspect sémantique se réfère à la composition des informations véhiculées par les modalités.

Composition					
Temporelle	Anachronique	Séquentielle	Concomitante	Coïncidente	Parallèle / Simultanée
Spatiale	Disjointe	Adjacente	Intersectée	Imbriquée	Recouvrante
Articulatoire	Indépendante	Fissionnée	Fissionnée et dupliquée	Partiellement Dupliquée	Dupliquée
Syntaxique	Différente	Complétion	Divergente	Extension	Jumelle
Sémantique	Concurrente	Complémentaire	Complémentaire et redondante	Partiellement redondante	Totalement redondante

FIG. 2.19 – Schémas et aspects de composition des modalités [Vernier and Nigay, 2000].

Ces études ont été appliquées à des systèmes de RA [Nigay, 2001] : CASPER est un système de chirurgie augmentée où un guidage visuel est superposé à une image réelle issue d'une caméra pour le positionnement d'une aiguille de ponction ; MAGIC est un dispositif pour l'aide à l'archéologie d'urgence qui fournit le relevé de terrain et aide au partage des trouvailles. Elles ont également été exploitées pour la visualisation sur ordinateur de grandes quantités d'informations [Vernier, 2001]. Ces trois exemples d'IHM montrent des situations où les concepteurs ont dû faire des choix de représentations visuelles pour communiquer des informations pertinentes

pour l'utilisateur. Plus généralement, quelques règles intéressantes de sélection et de composition des modalités ont été proposées dans [Nigay, 2005] :

- Les modalités de sortie doivent être en accord avec les préférences de l'utilisateur ;
- Les modalités de sortie doivent faire intervenir un dispositif disponible ;
- Les modalités qui véhiculent le centre d'intérêt doivent être précises et non déformées ;
- Les modalités globales doivent déformer le moins possible la structure des informations ;
- Si la tâche est critique, la combinaison de modalités doit être redondante ;
- Si une modalité est arbitraire, une autre modalité complémentaire et non arbitraire (linguistique de préférence) doit accompagner la modalité arbitraire pour expliquer la correspondance ;
- Une modalité précise et une modalité globale doivent toujours être présentes en parallèle ;
- La continuité spatiale doit être vérifiée : aucune discontinuité lorsqu'on balaye l'écran des yeux ;
- La continuité temporelle doit être vérifiée : aucune rupture visuelle à un endroit donné de l'écran au cours du temps.

Les trois espaces de conception que nous avons choisi de présenter sont représentatifs des études sur la multimodalité. Ils montrent les nombreux éléments à prendre en compte pour faire de la multimodalité : la psychophysique (caractéristiques des canaux sensoriels, processus cognitifs, représentation interne de l'information, etc.), les techniques (architectures, périphériques, modalités, etc.), l'information (ce qui doit être représenté, les tâches, les objectifs, etc.) et le contexte (l'ensemble des variables influençant l'interaction).

L'étude la plus aboutie en matière de rendu multimodal en RV est probablement celle de Nesbitt [Nesbitt, 2003], qui pose la question suivante : "comment concevoir de bons dispositifs de rendu multi-sensoriel pour des données abstraites (marchés boursiers, bases de données, trafic routier, données scientifiques, etc.)?".

Nesbitt a élaboré une taxonomie, des conseils et un processus afin de concevoir des dispositifs multi-sensoriels pour les données abstraites. Il cherche à aller plus loin que la conception ad-hoc, qui consiste souvent à ajouter du son ou de l'haptique à un rendu visuel existant. Sa thèse fournit d'abord une catégorisation des espaces de conception multisensoriels, afin de fournir une structure aux concepts qu'un concepteur doit prendre en compte. L'auteur développe ensuite des guides pour la projection des attributs des données sur les propriétés perceptives humaines. Enfin, il propose un processus de conception des interfaces multisensorielles, tenant compte des demandes des utilisateurs, de la nature des données abstraites, des limitations

des outils de développement et de la nature spécifique des environnements ciblés.

L'apport principal de ces travaux est que le rendu multimodal n'est plus adhoc mais se veut optimal et générique. De plus, Nesbitt propose tout un processus de conception, comportant notamment un ensemble détaillé de guides pour le choix des métaphores et la projection des données sur des rendus.

Cependant, ces travaux comportent quelques limites. Tout d'abord, cette étude se concentre sur des données abstraites, et sur des tâches d'exploration, d'observation, d'analyse de ces données, par exemple l'étude de données boursières. Ensuite, le processus de conception et les guides de métaphores multimodales proposés se veulent universels, alors qu'ils n'ont pas été évalués. Enfin, le rendu multimodal créé à partir des recommandations est choisi à l'avance, en fonction de critères pertinents, mais il reste le même pendant l'utilisation de l'application, le contexte dynamique ne faisant pas partie des critères envisagés.

2.4.2 Gestionnaires de rendu multimodal

La majorité des architectures de rendu multimodal (contextuel ou non) se fondent sur une description de toutes les possibilités par un langage de type XML. Ceci requiert de définir toutes les possibilités et toutes les règles de distribution modale.

CICERO [Arens and Hovy, 1995] est un gestionnaire d'interactions destiné à faire le lien entre l'application et les informations d'une part, et les médias et les utilisateurs d'autre part. Il est composé d'un intégrateur de données en entrée et d'un planificateur de présentation en sortie. Etant donné un ensemble d'informations à présenter à l'utilisateur, CICERO construit une présentation coordonnée et cohérente en utilisant diverses modalités. Pour cela, il se fonde sur un ensemble de modèles sémantiques, dont certains sont issus de travaux antérieurs :

- modèle de l'utilisateur : préférences de présentation, expertise, intérêt et opinion sur le domaine, capacité de langage, etc.
- modèle de la tâche : traite en particulier des objectifs de celui qui produit la présentation (en l'occurrence le système dans le cas de la sortie) par rapport à celui qui la reçoit (ici l'utilisateur). Par exemple, comment la connaissance de l'utilisateur est modifiée, de quelle manière ses objectifs sont altérés, comment varie sa position ou son implication par rapport au sujet, etc.
- modèle des informations : urgence, ordre, volume, décomposition en unités d'informations, etc.
- modèle des médias : caractéristiques matérielles et logicielles, capacités sensorielles et cognitives associées, etc.
- modèle de la communication : structure, cohérence, relations entre les différents composants du discours

L'intérêt de CICERO réside dans le fait que le lien entre les données à présenter et les médias disponibles pour la présentation n'est pas statique. A partir des propriétés des données, des objectifs de communication, du contexte de l'interaction, le gestionnaire déduit les propriétés requises par les médias, puis choisit ceux qui satisfont ces propriétés.

MAGIC (Multimedia Abstract Generation for Intensive Care) [Dalal et al., 1996] est conçu pour générer automatiquement des informations sur le statut post-opératoire de patients, après une chirurgie cardiaque : signes vitaux, médicaments administrés, cardiogramme, données sur d'éventuelles prothèses, etc. Ces informations sont contenues dans le système informatique d'un hopital et doivent être présentées au personnel soignant. Les modalités de présentation possibles sont le texte, la parole et l'image. MAGIC s'intéresse principalement à la coordination temporelle (ordre et durée) de ces modalités. Il faut par exemple s'assurer que des références vocales à des parties du corps du patient soient prononcées en même temps que celles ci sont mises en valeur sur une image du corps. Cette coordination est effectuée par la négociation entre les différents médias, qui aboutit à un ordre de présentation, et qui garantit la cohérence.

AVANTI (AdaptiVe and Adaptable iNteraction in multimedia Telecommunications applications) [Stephanidis and Savidis, 2001] est un troisième exemple d'application dans laquelle l'interface peut s'adapter à certains paramètres. Ici, il s'agit d'un navigateur web qui varie en fonction des capacités de l'utilisateur (valide, non-voyant, handicapé moteur), de son expertise, ou encore des différents contextes d'utilisation (domicile, bureau, etc.).

2.4.3 Adaptativité au contexte

Puisque nombre de travaux préconisent de tenir compte du contexte de l'interaction pour concevoir un rendu multimodal, il convient de définir ce qu'est ce contexte. L'Encyclopédie Larousse définit le contexte comme "l'ensemble des circonstances dans lesquelles se produit un événement, se situe une action". Dans le domaine linguistique, le contexte d'un mot ou d'un texte est l'énoncé dans lequel il figure, l'ensemble des termes qui le précèdent ou le suivent (Grand Dictionnaire Terminologique). C'est ce contexte qui permet d'éclairer la pensée de l'auteur lorsqu'elle est complexe, ou encore qui aide à comprendre le sens d'une expression inconnue, que ce soit dans une langue maternelle ou étrangère.

Dans le domaine de l'informatique et particulièrement de l'informatique ambiante, de nombreuses définitions du contexte ont été introduites. Certains définissent le contexte par des exemples, des énumérations des éléments qui le composent : localisation et identité des personnes, heure et température actuelles, objets

proches, etc. D'autres positionnent le contexte par rapport à des entités de référence : contexte de l'utilisateur, contexte de l'application, etc.

Une définition plus générale et largement acceptée est celle de Dey et Abowd [Dey and Abowd, 2000] [Dey et al., 2002] : "le contexte est toute information pouvant être utilisée pour caractériser la situation d'une entité considérée comme pertinente pour l'interaction entre un utilisateur et une application, y compris l'utilisateur et l'application eux-mêmes." Cette définition ne sépare pas les informations contextuelles données explicitement par l'utilisateur de celles identifiées automatiquement par le système. Cette distinction oppose de nombreux chercheurs. Actuellement, des technologies permettent de savoir où se déroule l'interaction (gps), les conditions météorologiques, le niveau de lumière et de bruit (capteurs), le nombre et l'identité des utilisateurs (biométrie, reconnaissance d'image). Cependant, pour obtenir certaines informations, il est peu probable de pouvoir se passer totalement de commandes précises de l'utilisateur, concernant : son humeur, son handicap, ses objectifs, son degré d'expertise, son processus de raisonnement, etc. Néanmoins, l'adaptativité du système à un contexte ne dépend pas de la manière dont celui-ci est déterminé (données explicite vs. identification automatique). Les deux problèmes sont indépendants, c'est pourquoi nous pouvons nous intéresser spécifiquement à la partie aval de la détermination du contexte qui est son exploitation en vue d'un rendu intelligent. Toutefois, nous devons malgré tout prendre en compte cette double source d'informations contextuelles, en prévoyant dans notre modèle de supervision l'intégration d'un observateur automatique du contexte.

Dey et Abowd proposent trois entités pour illustrer leur définition : les lieux, les personnes et les choses. Ces entités sont caractérisées par quatre catégories d'informations : l'identité, la position, le statut (ou l'activité) et le temps. Enfin, Dey et Abowd précisent comment les applications peuvent utiliser le contexte sous la forme de trois catégories de fonctions : la *présentation* d'informations ou de services (actions), l'*exécution* automatique de service et le *stockage* d'informations contextuelles en relation avec des données (pour un usage ultérieur).

Par analogie avec le contexte linguistique, l'intérêt de capturer le contexte d'interaction réside dans la difficulté pour un système à déterminer quels sont les objectifs de l'utilisateur, qu'est-ce qu'il tente d'accomplir.

Dey et Abowd précisent les trois étapes nécessaires pour qu'une application soit sensible au contexte. Premièrement, il faut capturer le contexte, sous la forme de données de bas niveau (capteurs, temps, paramètres physiques de l'environnement, etc.). Puis il faut interpréter ces données capturées, construire une représentation de haut niveau manipulable par l'application. Enfin, il faut fournir cette représentation à l'application, qui la combine à ses propres données pour créer un comportement adapté.

Chaari [Chaari and Laforest, 2005] [Chaari et al., 2006] préconise une séparation

claire entre les données propres à l'application et les données qui constituent le contexte. L'objectif est de pouvoir concevoir le coeur fonctionnel d'une application en faisant abstraction du contexte, et de développer l'adaptation au contexte dans une seconde étape. Cette frontière entre les données de l'application et les données contextuelles dépend du domaine observé. Par exemple, la localisation GPS de l'utilisateur fait partie intégrante de l'application pour un système de régulation du trafic routier, tandis que cette donnée est une partie du contexte pour une application de télé-médecine. Pour garantir cette séparation, le contexte est défini comme étant "l'ensemble des paramètres externes à l'application qui peuvent influencer le comportement de celle-ci en définissant de nouvelles vues sur ses données et ses services". Ces paramètres sont dynamiques dans le temps et doivent être transparents à l'utilisateur. Une nouvelle situation contextuelle correspond à une nouvelle instance de ces paramètres qui ne modifie pas les données de l'application mais peut mener à les traiter différemment.

Rousseau [Rousseau, 2006] propose un modèle conceptuel ainsi qu'une plateforme de rendu multimodal adaptatif. Ses travaux complètent l'espace de conception de Coutaz, Nigay et Vernier présenté dans la section précédente en s'attachant à rendre les applications adaptatives. Afin de répondre à la diversité croissante des contextes d'interaction, il intègre ce paramètre contextuel dans le processus de décision. Tout d'abord, il décrit le modèle WWHT qui caractérise le cycle de vie d'une présentation multimodale :

- What : quelle unité d'information présenter ;
- Which : quelles modalités de présentation allouer ;
- How : comment les instancier ;
- Then : comment faire évoluer la présentation.

Ensuite, il présente la plate-forme ELOQUENCE qui permet de concevoir un système de présentation multimodale, à travers trois étapes : l'analyse, la spécification et la simulation. Ces outils sont appliqués dans trois domaines : la présentation de l'appelant en téléphonie mobile, la présentation d'informations de ciblage en avionique militaire et le contrôle aérien.

2.4.4 Discussion

Les espaces de conception et les questionnaires que nous venons de présenter ont en commun des éléments essentiels à prendre en compte pour faire du rendu multimodal : l'utilisateur, l'environnement, les informations à communiquer, le rendu à effectuer.

Les travaux d'IHM se concentrent sur les capacités de l'utilisateur et cherchent souvent à générer un rendu adapté à ses éventuels handicaps afin qu'il puisse obtenir les mêmes informations qu'un utilisateur valide. Concernant l'environnement, là

encore, il s'agit en IHM (par exemple pour l'informatique mobile ou l'informatique ambiante) de pouvoir obtenir les mêmes informations quelque soit l'environnement dans lequel l'utilisateur se trouve. Les problèmes qui se posent sont l'emplacement de cet environnement (i.e. où et par quel moyen faire parvenir l'information) ainsi que les caractéristiques de l'environnement : conditions visuelles, auditives, temporelles, spatiales, etc. Celles-ci sont statiques, composées d'éléments dont l'état est le même pendant une certaine période. Les applications visées sont le plus souvent événementielles : elles peuvent être vues comme une succession d'événements qui ont un déclenchement et une fin ponctuels, mais les événements peuvent être simultanés. L'interaction de l'utilisateur avec le système suit également ce schéma événementiel : les interfaces sont conçues pour fournir au système des commandes et à l'utilisateur des présentations (souvent graphiques).

Sans oublier l'utilisateur ni l'environnement, nos travaux s'attache davantage à la tâche. En effet, comme nous l'avons vu, les tâches en IHM et notamment en RV sont complexes, et les éléments constitutifs de ces tâches ne sont pas vraiment pris en compte dans les études existantes. La sémantique de l'activité, les objectifs de l'utilisateur et le rôle que doit jouer le système par rapport à ces objectifs sont délicats à modéliser et à manipuler.

Les éléments du contexte en RV sont aussi très différents de ceux en IHM. Tout d'abord, l'utilisateur interagit de manière naturelle avec l'environnement virtuel et pas seulement avec les interfaces du système. Par exemple, sa position et son orientation par rapport à la scène sont à prendre en compte pour le rendu, tandis que sur un écran de téléphone portable la position de l'affichage ne va pas en dépendre. Ensuite, la charge mentale sur chacun des canaux sensoriels va intervenir fortement, d'autant plus que les interactions de RV sont non ponctuelles, et donc nombreuses et simultanées : on peut d'ailleurs considérer que l'utilisateur est toujours en état d'interaction, même quand il n'agit pas, car il est toujours soumis aux stimuli de l'environnement virtuel.

Ni les applications, ni les espaces de conception, ni les questionnaires de rendu n'ont proposé de règles génériques de projection entre des informations et des modalités. La section suivante étudie une nouvelle source d'inspiration pour ces règles : les expériences psychophysiques.

2.5 Etudes psychophysiques

Une part essentielle de la supervision du rendu multimodal que nous voulons concevoir réside dans l'ensemble des règles de distribution. Le superviseur artificiel comme le concepteur humain ont tous deux besoin de savoir quelle est la correspondance optimale entre informations et modalités, et quels critères sont nécessaires

pour choisir un rendu plutôt qu'un autre pour la réalisation d'une tâche.

Le modèle de supervision des rendus multimodaux que nous présentons dans le chapitre 4 se veut appliqué à la RV. Or comme nous l'avons vu, les systèmes de RV sont construits par et pour des utilisateurs humains, qui sont au coeur des interactions. Nous pensons donc que les connaissances et les expériences psychophysiques sont la clé d'un rendu multimodal pertinent. La conception d'un superviseur multimodal doit prendre en compte ces aspects au niveau des règles de distribution, à la fois pour :

- essayer de déterminer la meilleure correspondance possible entre un type d'information et un rendu ;
- savoir quels critères, quels éléments du contexte peuvent intervenir dans la décision ;
- quels problèmes, quels effets peuvent se produire lorsque deux modalités sont utilisées en même temps.

Les illusions sensori-motrices forment une partie des connaissances psychophysiques particulièrement intéressante à étudier dans le cadre de nos travaux. En effet, les illusions arrivent lorsque la perception ou le cerveau nous font croire des choses qui ne sont pas réelles, et c'est exactement le principe sur lequel s'appuie la RV : on peut dire en caricaturant qu'un dispositif de RV est conçu pour leurrer continuellement les sens de l'humain. Pour Gigante [Gigante, 1993], un système de RV est l'illusion de la participation à un environnement synthétique plutôt que l'observation externe d'un tel environnement. C'est pourquoi nous imaginons nous en inspirer pour faire du rendu multimodal. D'une part, elles peuvent nous indiquer quels canaux ou modalités sont incompatibles car générateurs de biais sensori-moteurs, d'erreurs ou de malaises. D'autre part, certaines combinaisons peuvent au contraire être à l'origine d'interactions novatrices, en créant des stimuli utiles. Un second champ d'études qu'il semble utile de connaître pour réaliser des interactions adaptées et profitables à un utilisateur humain concerne la charge cognitive. En effet, des indications telles que la quantité globale d'informations qui peuvent être traitées simultanément par un humain, ou encore le nombre de stimuli sensoriels simultanés et différenciables, seraient très instructives pour créer des rendus multimodaux pertinents.

2.5.1 Illusions sensori-motrices

Les illusions sensori-motrices sont liées aux capacités perceptives humaines évoquées dans la section 2.2 et au fonctionnement cognitif. Les illusions visuelles, ou illusions d'optique, sont les plus connues. La majorité le sont depuis l'antiquité, et leur connaissance est d'ailleurs parfois à l'origine de certains travaux philosophiques (comme l'allégorie de la caverne de Platon). Elles trompent notre perception des mouvements, des contrastes, des couleurs, des transparences et des géométries.

L'illusion Vertical-Horizontal et l'illusion de Müller-Lyer présentent deux segments de taille identique qui semblent différents (figure 2.20), soit à cause de leur disposition dans l'espace, soit en raison d'ajouts visuels perturbants. La seconde fonctionne également avec des cercles ou des carrés aux extrémités. Dans le même style, les illusions de Titchener/Ebbinghaus et de Delboeuf portent sur la taille de cercles positionnés à l'intérieur d'un ou plusieurs autres (figure 2.21). Là encore les cercles centraux paraissent différents et sont identiques. L'illusion donne les mêmes résultats. Un dernier exemple intéressant est l'illusion géométrique de Bourdon, dont les deux triangles alignés selon un côté semblent non alignés (figure 2.22).

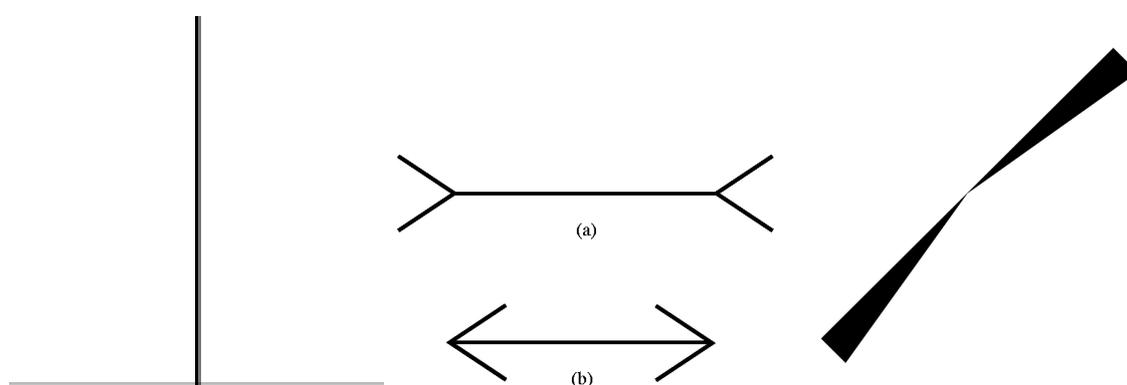


FIG. 2.20 – L'illusion visuelle Vertical-Horizontal et celle de Müller-Lyer.

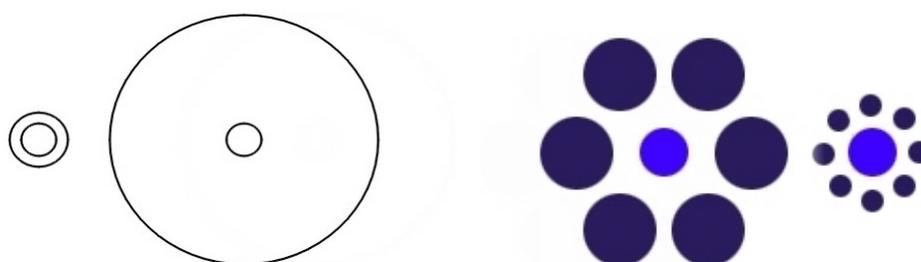


FIG. 2.21 – Les illusions visuelles de Delboeuf et de Titchener/Ebbinghaus.

Les illusions optiques sont très spécifiques, et servent principalement à l'enseignement, la culture ou les loisirs. Pour une application de RV, ce qu'il nous apparaît intéressant de retenir est que comme le canal visuel est prédominant, il convient d'éviter les cas qui prêtent à confusion. La stéréoscopie est notamment un moyen d'améliorer la perception de la géométrie, comme les distances ou les parallélismes.

Les illusions haptiques sont beaucoup moins connues que les précédentes, mais ce ne sont pas les moins intéressantes pour la RV. Leurs études sont récentes et sont sou-

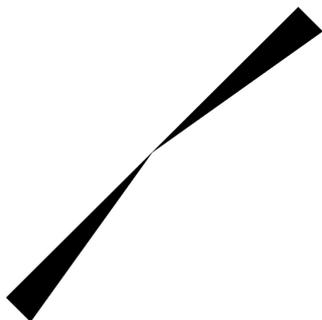


FIG. 2.22 – L'illusion visuelle de Bourdon.

vent une tentative de calquer les illusions visuelles. Gentaz [[Gentaz and Hatwell, 2004](#)] présente les résultats en haptique des illusions de Müller-Lyer, de Delboeuf et la Vertical-Horizontal. La première s'applique également à la perception haptique et est vraisemblablement le résultat de mêmes processus dans les deux modalités. La troisième existe dans les deux cas, mais est due pour partie à des processus similaires (bisection) et pour autre partie à des traitements spécifiques (anisotropie du champ visuel, et surestimation des mouvements exploratoires manuels radiaux par rapport aux tangentiels). Enfin, l'illusion de Delboeuf semble n'exister qu'en visuel, probablement car l'exploration par le doigt exclut la perception du contexte, qui est à l'origine de l'illusion. Day [[Day, 1990](#)] a de son côté montré que l'illusion de Bourdon existe également en haptique.

Les psychophysiciens qui effectuent ces expériences imputent généralement les résultats obtenus aux méthodes humaines de traitement de la perception et de l'action. Ce qu'il nous faut retenir pour une éventuelle application de ces études à la RV, est tout d'abord qu'elles ont été réalisées avec le canal haptique seul, sur des sujets voyants. Ce sont donc là encore des effets très spécifiques qui ne préjugent pas de l'efficacité de l'haptique pour des tâches d'exploration, en complément d'un rendu visuel. Par contre, le fait que certaines illusions ne fonctionnent plus en haptique (cf Delboeuf) confirme que le rendu haptique est propice à des tâches d'exploration ou de discrimination locales.

Des illusions "nouvelles" ont également été démontrées, cette fois-ci non pas pour montrer des limites de la perception haptique ou pour découvrir des processus de traitement, mais pour observer les effets réalisables par des interfaces haptiques de RV. Robles-De-La-Torre a ainsi montré que nous pouvons créer l'illusion de formes (trous et bosses) sur un plan en utilisant des forces de freinage et de propulsion [[Robles-De-La-Torre and Hayward, 2000](#)] [[Robles-De-La-Torre and Hayward, 2001](#)]. Gosline [[Gosline et al., 2001](#)] a créé des illusions haptiques de courbure en mélan-

geant des indices de force et des indices géométriques pour faire ressentir des formes différentes des formes de l'objet virtuel. Par exemple, en augmentant les frottements sur deux parties opposées d'un cercle lors de son exploration haptique, l'utilisateur a tendance à percevoir une ellipse.

La première de ces deux études trouve son application dans le rendu haptique de textures. Elle montre qu'il peut être plus aisé d'apporter à l'utilisateur l'illusion d'interagir haptiquement avec des objets ayant des textures particulières (rainures, érosion, etc.) plutôt que d'effectuer des modélisations détaillées des objets et des calculs complexes de points de collision ou de forces de frottement. C'est l'exemple de l'utilisation d'une nouvelle modalité pour communiquer des informations sur une scène. Dans le même esprit, mais moins appliquée, la seconde étude montre que l'on peut modifier la perception des objets d'une scène par l'exploration haptique : ils peuvent avoir une forme visuelle donnée, par exemple un parallépipède, et être perçus haptiquement comme des formes différentes, par exemple une forme ellipsoïdale. Ce dernier cas se rapproche d'ailleurs plus des illusions multimodales que nous allons présenter maintenant.

2.5.2 Illusions et effets multi-sensoriels

Les illusions multi-sensorielles apparaissent lorsqu'un canal sensori-moteur trompe la perception que l'on a sur un autre canal. Elles semblent donc essentielles à connaître pour développer des applications multimodales, soit pour éviter des effets indésirables entre les rendus, soit au contraire pour créer des interactions nouvelles. Les sections suivantes présentent les trois combinaisons possibles.

2.5.2.1 Visuel-Audio

L'illusion audio-visuelle la plus connue est sans aucun doute l'effet ventriloque, dans lequel la perception visuelle nous fait croire que la voix du ventriloque vient de son ventre et non pas de sa bouche. L'effet McGurk [McGurk and MacDonald, 1976] est un autre phénomène perceptif qui montre une interférence entre l'audition et la vision lors de la perception de la parole. Par exemple, si l'on regarde une personne qui prononce le phonème "ga", tandis que le son "ba" est diffusé, on a l'impression d'entendre le son "da". Ces deux illusions audio-visuelles montrent clairement un effet de la modalité "message vocal" et de la modalité "animation" entre elles. En particulier, l'effet ventriloque souligne que des incohérences dans la localisation spatiale d'événements auditifs et visuels synchrones peut conduire à une localisation erronée de la source sonore. L'étude de ces effets est essentiel si l'on veut ajouter du son dans un environnement virtuel visuel.

Plus récemment, des chercheurs plus spécialisés en IHM ont commencé à étudier la combinaison de différentes modalités en RV. Storms [Storms and Zyda, 2000] étu-

die par exemple l'influence de la qualité individuelle des rendus visuel d'une part et audio d'autre part, sur la qualité perçue globale de la scène.

Le projet POEMS (Perceptually Oriented Ego-Motion Simulation) a également montré que l'audio peut créer artificiellement chez l'utilisateur la perception du déplacement [Väljamäe et al., 1995].

2.5.2.2 Visuel-Haptique

Pour les illusions visuo-haptiques, un des résultats les plus frappants est l'effet pseudo-haptique étudié en détail par Lécuyer [Lécuyer et al., 2000]. Il montre que le rendu visuel peut influencer la perception haptique, alors même que l'on manipule une interface haptique passive. Par exemple (figure 2.23), on peut jouer sur le déplacement visuel d'un piston de seringue pour faire percevoir des efforts de différentes valeurs sur la Space Mouse, qui a pourtant toujours la même raideur. Un autre exemple est le déplacement d'un cube dans un "marais visuel", qui incite l'utilisateur à forcer plus sur la Space Mouse. Paljic [Paljic et al., 2004] a par la suite montré que l'effet pseudo-haptique fonctionne également pour les torsions.

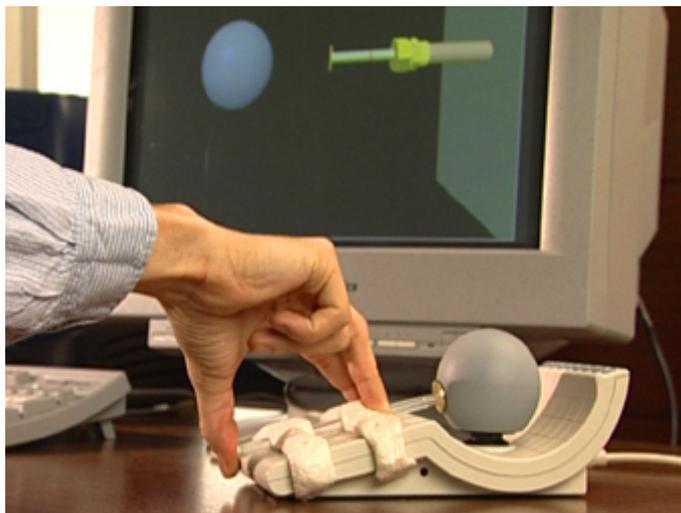


FIG. 2.23 – Le pseudo-haptique [Lécuyer et al., 2000]

Heller se concentre lui sur la perception des propriétés spatiales d'objets virtuels [Heller, 1982] et Wu sur la perception de raideur [Wu et al., 1999].

Nous pouvons également ajouter à ces illusions les illusions vestibulaires. Par exemple, de nombreux simulateurs de véhicules créent l'illusion d'une accélération ou d'une décélération par la combinaison du rendu visuel adéquate et de l'inclinaison de quelques degrés du siège de l'utilisateur, vers l'arrière ou vers l'avant.

2.5.2.3 Audio-Haptique

Concernant les effets entre les canaux audio et haptique, les expériences sont tournées vers la perception de texture, de rugosité ou de chocs. McGee propose de pallier les insuffisances des bras à retour d'effort à communiquer des informations de texture [McGee et al., 2000] et de rugosité [McGee et al., 2001] en ajoutant des signaux sonores réalistes. Cependant, le choix des modalités supplémentaires sont empiriques et la modification de la perception haptique n'est pas quantifiée. Toujours dans la perception haptique de rugosité, Guest [Guest et al., 2002] oriente ses recherches vers les sons utilisés. Les résultats de tests de discrimination montrent qu'ils sont influencés par la composante fréquentielle des stimuli sonores. En particulier, atténuer les hautes fréquences sonores augmente la perception de lissage de la surface. Retarder le rendu sonore provoque également une modification la perception de rugosité. Enfin, Bresciani et al. [Bresciani et al., 2004] étudient l'influence de l'audio sur la perception tactile de chocs. En diffusant un certain nombre des sons simples (bip) pendant un certain nombre de battements tactiles sur la pulpe d'un doigt, ils montrent que le rendu audio a tendance à augmenter le nombre de stimuli tactiles perçus.

2.5.3 Charge cognitive

Le terme de charge cognitive désigne les ressources mentales occupées à un instant donné chez un être humain lors de la résolution d'un problème, la conduite d'un raisonnement ou l'accomplissement d'une tâche (y compris un tâche de perception, de mémoire, de langage, etc.). Cette charge dépend bien sûr des problèmes ou des tâches considérés, mais aussi des caractéristiques individuelles, et de facteurs sociaux et environnementaux. Récemment, les recherches en théorie de la charge cognitive, popularisées notamment par Sweller et Chandler [Sweller and Chandler, 1994], se sont concentrées en particulier sur les facteurs limitants de l'attention lors de l'apprentissage, et sur la notion de "mémoire de travail". Ces deux aspects ont été identifiés comme les goulets d'étranglement principaux lorsqu'un humain traite des informations.

Il semble évident que l'on apprend mieux lorsque l'on peut tirer partie de ce que l'on connaît déjà. Plus l'on doit apprendre de choses dans un court laps de temps, plus il est difficile de traiter les informations. Par exemple, la charge cognitive lors de la compréhension d'un texte dans une langue étrangère est plus élevée que dans sa langue natale, car le cerveau doit simultanément traduire la langue et tenter de comprendre les informations du texte.

Une hypothèse abordée par la théorie de la charge cognitive est que les êtres humains travaillent avec deux types de mémoire : la mémoire à court terme, qui est limitée, et la mémoire à long terme, qui semble illimitée. Lors d'un processus

d'apprentissage, la mémoire à court terme, ou mémoire de travail, est largement sollicitée. Or celle-ci ne peut traiter qu'un nombre limité d'informations (le nombre couramment admis est d'environ 7 éléments à la fois). Sweller énonce que l'apprentissage optimum est obtenu lorsque la charge de la mémoire à court terme est minimale, afin de faciliter les changements dans la mémoire à long terme. Le processus d'apprentissage nécessite l'acquisition de schémas, qui organisent l'information, en ne gardant que les informations spécifiques. Si la charge cognitive est trop grande, c'est-à-dire que la mémoire de travail est saturée, ce processus d'acquisition des schémas est bloqué. Sweller et Chandler identifient deux facteurs qui augmentent la charge cognitive :

- le "split-attention effect" : l'attention de l'apprenant est distraite par la mauvaise organisation de l'information. Par exemple, dans le domaine de l'informatique, l'apprenant passera sans arrêt du manuel à son ordinateur. Second exemple : au lieu de séparer un manuel en textes et en schémas, il vaut mieux condenser l'information et intégrer le texte essentiel au schéma.
- le "redundancy effect" : la même information se retrouve à deux endroits différents. Ainsi, la charge cognitive est augmentée inutilement car la même information doit être traitée deux fois.

Si l'on utilise suffisamment les schémas que l'on a acquis, ceux-ci sont appliqués de manière automatique, sans que le sujet n'ait besoin d'un contrôle conscient. Ainsi, les experts utilisent moins de mémoire de travail que des novices pour résoudre un problème, car ils ont pu regrouper des connaissances. Ils identifient un état particulier d'un problème et savent immédiatement quelles sont les opérations à appliquer pour le résoudre. Au contraire, les novices doivent d'abord décomposer un problème en sous-but, avant d'aller en avant vers la résolution d'un problème.

La théorie de la charge cognitive peut être ainsi utilisée pour prédire les performances des humains lors de l'utilisation de nouvelles interfaces, et pour aider à concevoir des interfaces qui minimisent la charge. Les théoriciens séparent notamment la "complexité externe" associée à une interface de la "complexité interne" liée à la tâche principale effectuée par l'utilisateur, et proposent d'évaluer la première en comparant les performances obtenues par les utilisateurs avec différentes interfaces. L'objectif est bien sûr de diminuer la charge cognitive externe, afin que les ressources mentales des utilisateurs soient dédiées à leur tâche principale.

2.5.4 Discussion

Les études que nous citons dans les paragraphes précédents montrent que les questions sur les mécanismes de perception, d'intégration sensorielle ou de cognition mises en jeu dans des interactions de RV sont nombreuses. La réutilisation de telles expérimentations pour concevoir une base de connaissances menant à un rendu multimodal pertinent est cependant difficile. Tout d'abord, les illusions sensorielles

ou multi-sensorielles ne sont pas toutes exploitables en terme de tâches concrètes en RV, en particulier lorsqu'elles apparaissent dans des cas extrêmement précis ou limités. Elles informent cependant des configurations particulières à éviter, comme par exemple l'illusion de Delboeuf en visuel, ou encore celle de Müller-Lyer en visuel et en haptique, pose des conditions sur l'exploration d'objets proches ou la discrimination de tailles de solides. L'effet McGurk quant à lui peut trouver une application lors de la synthèse combinée de voix et d'expressions faciales d'un humanoïde. D'autre part, les expériences n'ont pas toutes conduit à des évaluations et celles qui ont été évaluées ne permettent pas de garantir qu'un rendu est plus pertinent qu'un autre pour une tâche ou une information donnée. Par exemple, l'ajout de signaux sonores à la perception haptique de chocs ou de frottements montrent une incidence d'un canal sur l'autre mais cet effet n'est pas quantifié, et aucune recommandation concrète ou ni règle de distribution multimodale n'en émergent. On aurait en effet pu imaginer que la diffusion de sons de collisions métalliques et de frottements réalistes puisse améliorer la performance en terme de qualité et de rapidité d'un opérateur pour une tâche de montage virtuel.

Paradoxalement, il semble donc plus sûr de réexploiter les résultats "négatifs" qui indiquent des situations ou des combinaisons potentiellement pénalisantes pour un utilisateur en RV, tout en conservant les résultats prometteurs comme des choix de distribution à évaluer après la réalisation d'un moteur de rendu multimodal.

Les recherches dans le domaine de la charge cognitive, initialement conduits dans le domaine de l'apprentissage, fournissent des éléments intéressants lorsqu'elles sont appliquées aux IHM. C'est notamment le propos d'Oviatt [Oviatt, 2006]. Dans le cadre qui nous concerne, il semble intéressant dans un premier temps d'intégrer dans notre moteur certaines règles pratiques de rendu d'information établies par les théoriciens de la charge cognitive (organisation spatiale des modalités, restriction des redondances, continuité temporelle, etc.), puis à terme, lors des évaluations des interactions multimodales concrètes, de mesurer la charge cognitive effective (par des données psychométriques telles que les pulsations cardiaques, la résistance de la peau, ou par des mesures de performance dans la tâche) tout comme la charge subjective perçue par les utilisateurs (par exemple par une interface de saisie spécifique). Les choix de distribution multimodale pourraient alors être révisés.

2.6 Conclusion

La communauté de la RV est convaincue que la combinaison intelligente de modalités selon différents canaux sensori-moteurs procure des bénéfices aux applications et aux utilisateurs. Les études se multiplient tant dans les domaines applicatifs (apprentissage, exploration de données scientifiques, assistance à l'opérateur) que dans les champs de la modélisation et de la psychophysique.

L'état de l'art montre des études qui implémentent ou évaluent une ou des modalités non habituelles (mais souvent naturelles) pour des informations ou des tâches diverses. Cependant, les rendus sont proposés dans un objectif précis et sont statiques pendant l'exécution de l'application. Toutes ces applications utilisent la multimodalité en sortie, mais les choix de représentation, les règles de distribution, sont effectués en amont, avant le développement, par les concepteurs. A aucun moment ces choix ne sont réalisés par un quelconque système informatique intelligent, à partir d'une connaissance de la tâche. Cette thèse aborde ce problème particulier et tente de regrouper ces travaux. Notre objectif est de concevoir un moteur de rendu multimodal générique permettant : 1) de réaliser un rendu adapté à diverses informations et tâches, 2) de tenir compte du contexte dynamique de l'exécution, lié en particulier à la succession et la simultanéité des interactions. Le chapitre suivant présente trois applications qui ont servi de support pour nos travaux de modélisation et d'implémentation.

Chapitre 3

Analyse d'applications de référence

Résumé. *Ce chapitre présente en détail trois applications qui pourraient potentiellement bénéficier de rendu multimodal : l'exploration d'écoulements en Mécanique des Fluides, l'appariement de protéines en Bio-Informatique et la conception de formes assistée par la RV. Dans un premier temps, nous décrivons les problématiques multimodales des trois domaines et nous imaginons des interactions multimodales utiles. Puis, dans le cadre du projet CoRSAIRe, les deux premières applications font l'objet d'analyses ergonomiques : les tâches et les besoins des utilisateurs sont identifiés et des recommandations multimodales sont formulées. Ces deux analyses détaillées, complétées d'un expertise plus pratique en CAO, doivent nous servir à bâtir un modèle des interactions qui soit générique et représentatif des activités réelles des utilisateurs, et donc un superviseur multimodal utile et concrètement utilisé.*

3.1 Introduction

Pour éviter d'avoir un modèle qui ne soit pas concrètement implémentable en RV, notre démarche a été de considérer les applications dès le commencement de notre étude. Nous avons fondé nos travaux sur trois applications disponibles au laboratoire : la manipulation et l'observation de protéines dans le cadre du problème d'appariement en Bio-Informatique ; l'exploration et l'analyse scientifique d'écoulements en Mécanique des Fluides ; la manipulation et l'édition de formes en Conception Assistée par Ordinateur. Les applications de Bio-Informatique et de Mécanique des Fluides sont au coeur du projet CoRSAIRe (Combinaison des Rendus Sensori-moteurs pour l'Analyse Immersive de Résultats) labellisé par l'Agence Nationale de la Recherche en décembre 2005. Elles ont fait l'objet d'une analyse ergonomique poussée réalisée par J. Nelson et J.-M. Burkhardt du Laboratoire Ergonomie, Comportement et Interactions de l'Université Paris Descartes.

Notre objectif est de compléter ces applications ou de les adapter au domaine de

la RV immersive et multimodale en sortie. Pour chacune d'entre elles, nous commencerons par décrire l'existant et les contextes d'utilisation. Puis nous présenterons les analyses ergonomiques détaillées qui ont été réalisées sur ces activités, et les recommandations qui en ont découlé pour la conception d'applications et d'interactions immersives et multimodales. Ces recommandations mettent en évidence les activités, les besoins, particuliers ou communs aux différents domaines, qui vont demander un rendu multimodal et qui vont devoir être intégrés dans les futures applications.

3.2 Potentialités de la multimodalité dans les applications visées

Nous présentons dans cette section une description du domaine, du contexte et des problématiques principales pour chacune des trois applications visées. Nous tentons également d'imaginer les apports possibles des rendus multimodaux pour les utilisateurs, afin de les aider à effectuer leurs tâches ou pour améliorer les résultats.

3.2.1 Appariement de protéines en Bio-Informatique

3.2.1.1 Description de l'application

Les recherches et développements en Bio-Informatique sur l'architecture spatiale des génomes et les interactions moléculaires dans les cellules des organismes vivants relèvent le défi majeur de la compréhension de l'organisation dynamique de l'activité cellulaire à différentes échelles. L'exploration spatio-temporelle, la représentation et la visualisation deviennent alors des étapes clés afin d'aider les biologistes pour mieux cibler leurs recherches et leurs études.

L'étude des rapports entre structures et fonctions a toujours été le thème récurrent, explicite ou sous-jacent de la plupart des recherches biologiques. Connaître la structure tridimensionnelle d'une protéine est une étape nécessaire mais souvent insuffisante pour comprendre un certain nombre de mécanismes fonctionnels. Ainsi, la plupart des protéines assurent leur fonction sous la forme de complexes : deux (ou plus) protéines, appelées "partenaires", se combinent pour former une seule entité. Dans ce cas, c'est la structure du complexe qui devient l'objet d'intérêt. Prédire la structure des complexes protéiques est une démarche tout aussi délicate que de prédire la structure des partenaires. Actuellement, cette prédiction s'opère par le biais de méthodes dites de "docking" (traduit par "amarrage" ou encore "appariement") entre les protéines participantes. Les biologistes disposent d'abord de logiciels de visualisation comme Pymol, Rasmol ou ICM (figure 3.1) qui leur permettent d'étudier les topologies et les caractéristiques des protéines. Ensuite, concernant les complexes protéiques, il faut une équipe de plusieurs personnes et des calculs intensifs de pré-

diction pouvant durer plusieurs semaines, pour espérer tomber sur un complexe viable de deux protéines. De plus, ces prédictions échouent souvent, obligeant ainsi à réitérer tout le processus avec de nouvelles valeurs de paramètres ou en changeant de méthode de prédiction. Il se trouve que tous les algorithmes de prédiction explorent de manière combinatoire tous les cas possibles d'arrimer une protéine avec une autre. Quand on sait qu'il y a entre 500000 et 1000000 de protéines, on peut imaginer les grandes difficultés de mener à bien ce type de simulation "en aveugle".

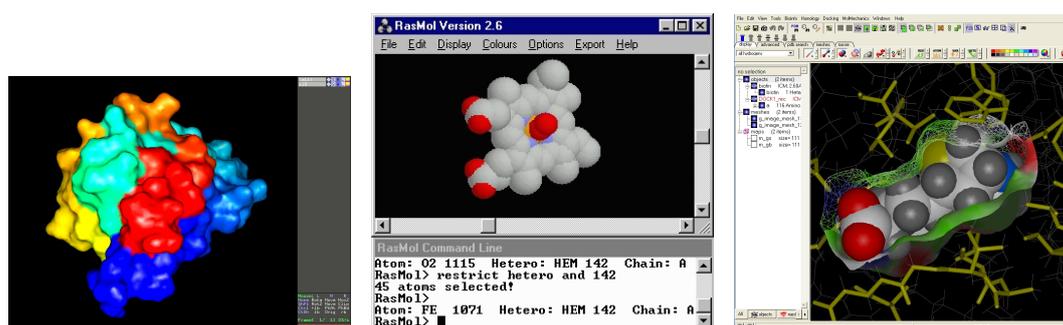


FIG. 3.1 – Logiciels de visualisation moléculaire : PyMOL, Rasmol et ICM.

3.2.1.2 Interactions multimodales pressenties

Le docking virtuel de protéines est un terrain très prometteur pour les recherches en interaction 3d multimodale. En effet, une autre approche possible pour répondre au problème est de faire intervenir le biologiste dans la boucle de simulation. Celui-ci pourrait interagir de manière virtuelle avec les molécules pour simuler les dockings potentiels, et améliorer ainsi considérablement la fiabilité et la durée de prédiction de ces dockings. Il fournirait, par son expérience, un nombre restreint de dockings potentiellement valides, sur lesquels les calculs lourds seraient alors ciblés. Nous pensons que la meilleure façon pour fournir de manière naturelle à un biologiste ce nombre restreint de dockings potentiels, est de lui présenter virtuellement les deux protéines (voire plus) à grande échelle et de lui donner la possibilité de les manipuler gestuellement avec ses deux mains. En effet, par le biais d'une présentation et d'une interaction multi-sensorielle, on peut :

- d'une part permettre au biologiste d'effectuer son (ses) docking(s) en utilisant par exemple le canal gestuel pour la manipulation et le canal vocal pour les commandes ou le contrôle du paramétrage ;
- d'autre part, faire en sorte que le système présente les informations (sur les protéines ou sur les variables issues des calculs de prédiction) pour les unes sous forme visuelle et pour d'autres sous forme haptique, voire sonore 3d.

Ainsi, on exploiterait la bande passante multi-sensorielle de l'utilisateur pour lui présenter simultanément les informations optimales susceptibles de l'aider à cibler son étude. Des travaux centrés sur la visualisation immersive de molécules (ADN ou protéines) ont déjà été menés (figure 3.2).



FIG. 3.2 – Logiciels de visualisation moléculaire immersive : ADN-Viewer [Hérisson et al., 2004] (LIMSI-CNRS) et ARC Docking (LORIA).

Notre approche suppose une poursuite de ce type de travaux sur les rendus audio et haptique et l'interaction gestuelle avec les protéines, combinés avec les algorithmes de docking. Les informations à présenter au biologiste lors de la recherche d'une configuration de docking sont de types, d'échelles et de qualités variés. Elles peuvent concerner une seule protéine ou être liées à une configuration en cours de test : topologie d'une protéine ou complémentarité topologique des deux, champ électrostatique d'une molécule ou forces électrostatiques entre deux, niveau d'énergie du complexe, etc. Les capacités sémantiques des canaux et des modalités pourraient être utilisées pour communiquer simultanément et efficacement au chercheur ces informations calculées par le simulateur¹.

3.2.2 Exploration de cavité en Mécanique des Fluides

3.2.2.1 Description de l'application

Actuellement, l'étude des caractéristiques des structures dynamiques tridimensionnelles générées au sein d'un écoulement devient un enjeu conséquent en Mécanique des Fluides, notamment dans le domaine du contrôle. Les problèmes de traînées en sont un exemple. Dans le domaine aéronautique, ces structures jouent

¹Les informations "élémentaires" sont de plus calculables plus rapidement que la résolution complète de l'algorithme de docking.

un rôle prépondérant sur les performances aérodynamiques d'un système. Dans les problèmes d'environnement urbain, elles ont une influence certaine sur le confinement des polluants ou le renouvellement de l'air dans les rues.

L'étude d'un écoulement tridimensionnel instationnaire se révèle extrêmement complexe. Néanmoins, les outils d'analyse actuels pour ce type d'étude s'avèrent quelque peu limités. En effet, ici, l'information à traiter consiste en un ensemble de vecteurs représentant des quantités physiques (vitesse, pression, température, gradients, etc.) échantillonné sur une grille 3d. Le premier problème qui se pose est celui de la simulation numérique (ou des mesures expérimentales) de ces données. Vient ensuite un travail de post-traitement des données qui peut s'avérer très lourd.

Les deux caractéristiques principales de ces données sont d'une part qu'elles sont à trois dimensions spatiales et une temporelle, d'autre part qu'elles sont par nature abstraites. Si les outils d'analyse scientifique dont les chercheurs disposent actuellement (figure 3.3) permettent parfois de visualiser des champs 3d, ceux-ci restent statiques, ce qui complique le suivi de leur évolution au cours du temps. De plus, il existe de grandes difficultés de visibilité des informations lors de l'exploration des champs. La conception d'animations 3d dans ce domaine est limitée car elle est complexe, fastidieuse et demande beaucoup de temps pour sa mise en forme. Ces animations sont généralement conçues pour mettre en valeur des résultats obtenus à partir d'études menées par des moyens plus modestes. En premier lieu, les champs instantanés sont explorés par l'intermédiaire d'un ensemble de coupes 2d, avec éventuellement une animation temporelle pour percevoir l'aspect instationnaire. L'expérience du chercheur oriente généralement la sélection des plans de coupe pour en limiter le nombre à considérer. Un travail conséquent d'analyse est alors mené pour aboutir à la compréhension des caractéristiques 3d instationnaires de l'écoulement. Après avoir ciblé la nature des résultats de l'étude, le recours à l'animation 3d peut alors être effectué pour mettre en valeur, affiner ou obtenir un supplément d'information sur les résultats obtenus. Compte tenu des problèmes de post-traitement, les études sur les structures dynamiques des écoulements tridimensionnels instationnaires se limitent à des configurations géométriques relativement simples.

3.2.2.2 Interactions multimodales pressenties

En mode visuel, la RV en milieu immersif offre de plus grandes possibilités dans la façon d'appréhender la perception des champs. Tout d'abord, l'espace de visualisation étant plus important, le chercheur peut bénéficier d'une plus grande résolution du rendu visuel, l'amenant à mieux percevoir les détails (figure 3.4). De plus, la possibilité de se mouvoir dans la scène facilite la tâche d'exploration des champs, en simplifiant la localisation et l'extraction des zones pertinentes de l'écoulement :

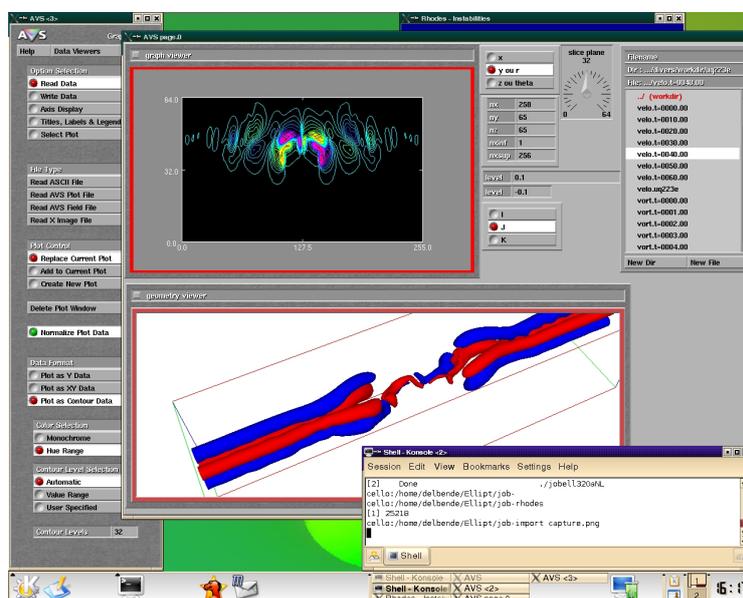


FIG. 3.3 – Logiciel de visualisation d'écoulement AVS. On distingue notamment une visualisation 3d (au centre, en bas), une coupe 2d (au centre, en haut) et une liste pour accéder à différents instants de la simulation (à droite, en haut).

soit par le déplacement propre de l'utilisateur de manière à observer l'écoulement sous différents angles (par stéréoscopie adaptative grâce au suivi de position de la tête), soit par la navigation du véhicule que constitue le dispositif (par exemple par un joystick ou une souris 3d). Enfin, l'analyse d'un écoulement repose souvent sur l'étude des champs de différentes grandeurs physiques (composantes de vitesse, masse volumique, énergie, pression, température, rotationnel, second invariant de vitesse, etc.), dont le nombre et l'importance varient suivant le type de phénomène que l'on veut mettre en avant. L'étude d'un écoulement nécessite souvent de pouvoir visualiser plusieurs champs simultanément et sous différentes formes : par exemple, différentes valeurs des isosurfaces de pression combinées avec les vecteurs vitesse, ou encore la superposition de deux isosurfaces et d'un plan de coupe (figure 3.5).

Si l'immersion visuelle se révèle la modalité la plus attendue pour avoir une perception des informations résultantes des données de simulation (pour des raisons cognitives et parce que, par la force des choses, elle était jusqu'alors unique), elle admet des limitations essentiellement liées à la surcharge de scènes qui peuvent "noyer" les informations recherchées (trop d'isosurfaces ou de formes complexes qui limitent la profondeur de visualisation, par exemple). De plus, le mode visuel n'est pas forcément le plus adapté pour percevoir certains phénomènes de nature intermittente ou fortement localisés (intermittence de la turbulence, ruptures de structures résultant de l'étirement tourbillonnaire, etc.). Il est évident que ces phénomènes ne peuvent

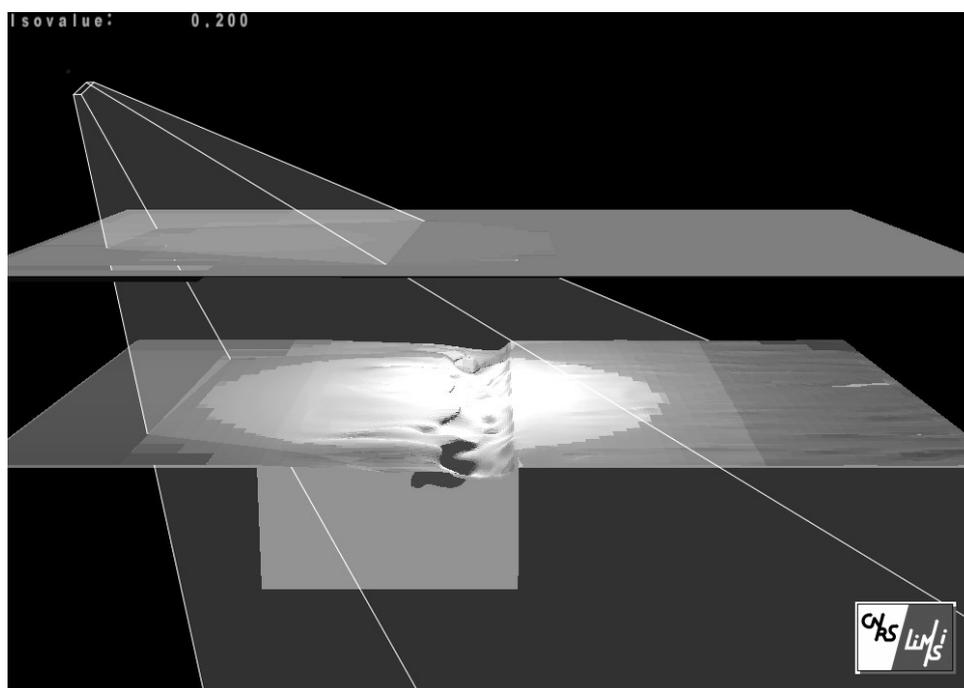


FIG. 3.4 – Visualisation d’une isosurface de Mécanique des Fluides avec différents niveaux de détail [Gherbi et al., 2006].

être perçus par le mode visuel que s’ils se trouvent dans le champ de vision et que les paramètres de visualisation sont correctement réglés (valeur d’isosurface pertinente, par exemple). En revanche, la perception auditive peut avantageusement être utilisée pour détecter ce type de phénomène dans l’écoulement et guider le chercheur sur leur localisation (qu’il pourra ensuite visualiser, connaissant l’existence du phénomène). Ajoutons à cela que le mode audio peut être un bon moyen de percevoir les phénomènes de nature périodique, ceux-ci étant caractérisés par une fréquence particulière.

De son côté, l’haptique pourrait être avantageusement utilisé pour aider à la manipulation des outils de l’application, par exemple pour contraindre le déplacement d’un plan de coupe selon un axe (figure 3.6) ou pour aider à le positionner sur certaines zones caractéristiques. Il pourrait s’avérer également utile pour la perception épistémique des données, par exemple pour ressentir la topologie des isosurfaces. Enfin, certaines fonctionnalités telles que la viscosité pourraient être utilisées pour donner à ressentir des gradients de valeurs ou des champs de vecteurs.

L’étude des écoulements nous semble un domaine très intéressant pour tous les problèmes d’exploration de données en RV. En utilisant l’immersion et la multimodalité, nous pensons qu’il est possible d’accroître de manière significative le potentiel

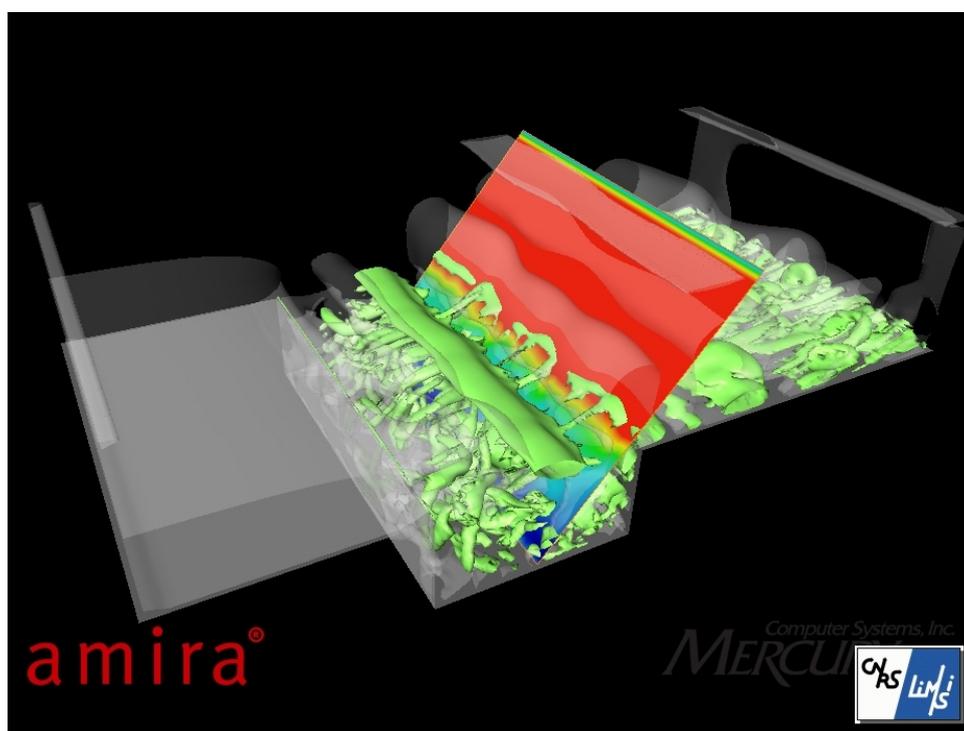


FIG. 3.5 – Visualisation d’une isosurface d’un facteur Q (verte) avec une isosurface de vitesse verticale (transparente) et un plan de coupe du module de la vitesse.

d’analyse des données complexes (tridimensionnels et instationnaires) issues de simulations numériques. Le développement d’outils multimodaux adéquats permettrait au chercheur d’optimiser l’analyse des écoulements aux caractéristiques dynamiques complexes. En fait, à long terme, l’idéal serait d’aboutir à la conception d’une soufflerie numérique virtuelle interactive. Les données numériques issues d’un code de calcul en cours d’exécution ou d’un banc expérimental seraient instantanément présentées dans un environnement immersif permettant ainsi au chercheur d’analyser l’évolution d’un écoulement instationnaire en temps réel à l’aide des différents outils de perception mis à sa disposition. Il pourrait aussi interagir avec le code de simulation dans le but de modifier directement les paramètres de simulation, contrôlant ainsi en temps réel les caractéristiques de l’écoulement par le biais d’interfaces multimodales. L’utilisation de la multimodalité dans le cadre d’une soufflerie numérique virtuelle interactive permettrait de concevoir de nouvelles approches pour l’étude des écoulements tridimensionnels instationnaires, se basant sur un enrichissement de la perception des informations via une augmentation de la bande passante sensorimotrice. L’état de l’art des applications multimodales (section 2.3) a montré des travaux de sonification et d’haptisation de flux, mais nous n’avons pas trouvé d’intégration coordonnée des trois rendus principaux. Le rendu visuel stéréoscopique et

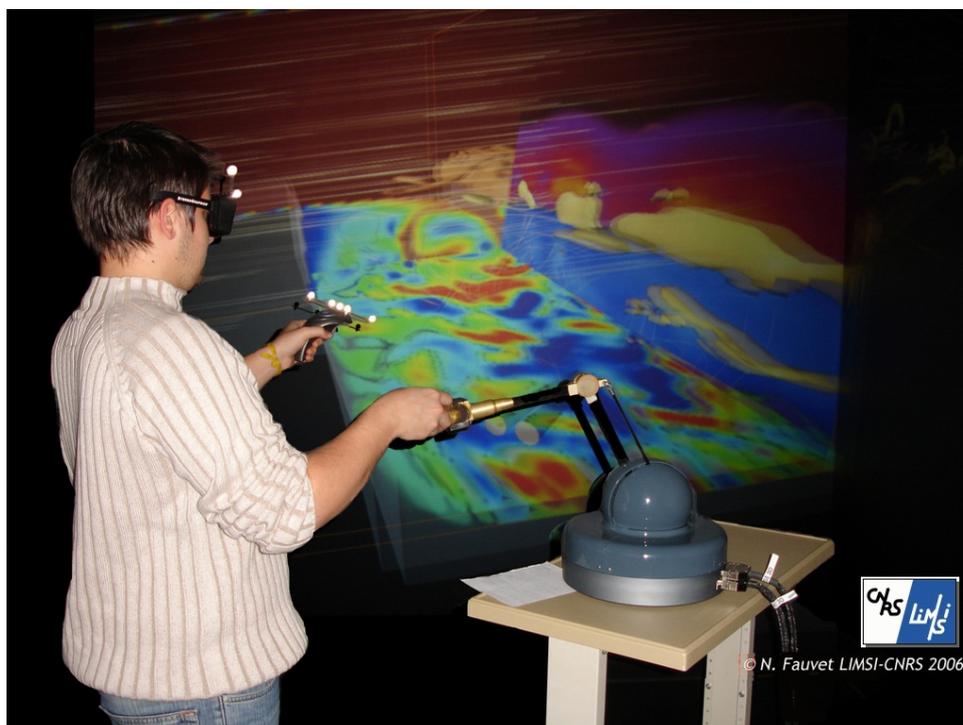


FIG. 3.6 – Interaction avec un plan de coupe et visualisation d’une isosurface de Mécanique des Fluides [Bouyer et al., 2007].

immersif serait utilisé principalement pour observer de manière globale les données, varier les représentations mais également pour faire le lien avec les habitudes des chercheurs. Des paradigmes audio seraient ajoutés pour reconnaître ou détecter des phénomènes particuliers, notamment temporelles. L’haptique pourrait également participer à cette détection, et aider à la manipulation des outils de l’application.

A plus court terme, bien que les calculs de simulations numériques ne peuvent pas être à ce jour menés en temps réel, travailler sur la conception d’interactions multimodales immersives de leur exploration en différé ("hors-ligne") est pertinent. Les outils développés dans ce contexte seront alors disponibles lorsque les calculateurs et les modèles de Mécanique des Fluides seront compatibles avec le temps réel.

3.2.3 Conception Assistée par Ordinateur

3.2.3.1 Description de l’application

En dehors de la recherche et de l’exploration scientifique, une autre discipline disponible pour nos travaux est la conception de formes en environnement immersif.

Dans le monde réel, pour construire une forme, un opérateur humain va suivre habituellement deux étapes schématiques. D'abord, il va s'installer devant une planche à dessin et, à l'aide d'un crayon, de ses deux mains et d'un contrôle visuel de son activité, il va concevoir la forme, prévoir son rôle et ses interactions, planifier les différentes étapes de sa création. Ensuite, il va pouvoir passer à la création concrète de la forme, en manipulant des outils à l'aide d'actions motrices et de perceptions sensorielles diverses.

Dans le monde industriel, la Conception Assistée par Ordinateur (CAO) suit un processus logique et fait appel à des stations de travail et des logiciels qui utilisent de manière intensive un rendu uniquement visuel et une interaction classique d'ordinateur de bureau (paradigme WIMP : Windows, Icons, Menus and Pointing devices). Le logiciel CATIA en est un exemple (figure 3.7).

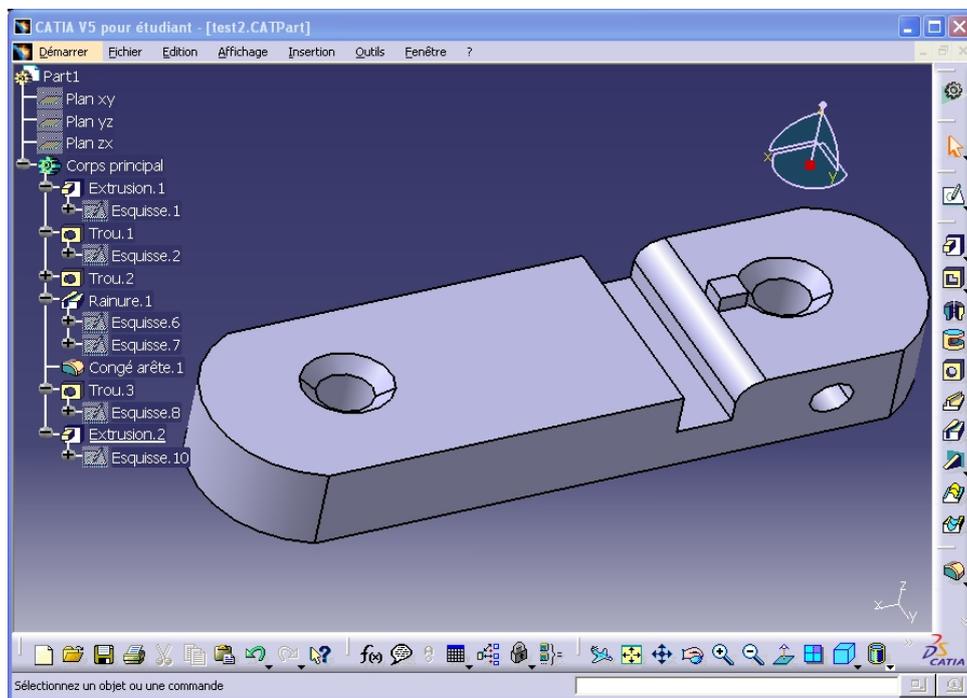


FIG. 3.7 – Logiciel de CAO avec paradigmes WIMP (CATIA).

Considérant qu'une telle interaction n'est pas adaptée à une tâche de conception d'objets 3d (nombreuses boîtes de dialogue, boutons, etc.), Convard et Bourdot ont tenté de rapprocher la CAO actuelle de la conception réelle originelle d'une forme [Convard and Bourdot, 2003]. L'application VRAD (Virtual Reality Aided Design) s'attelle à l'amélioration de l'utilisabilité des outils de CAO pour la conception de pièces géométriques en utilisant à la fois l'immersion visuelle stéréoscopique et l'interaction gestuelle et vocale pour la conception (figure 3.8). La multimodalité en

entrée bénéficie ainsi à l'activité du concepteur.

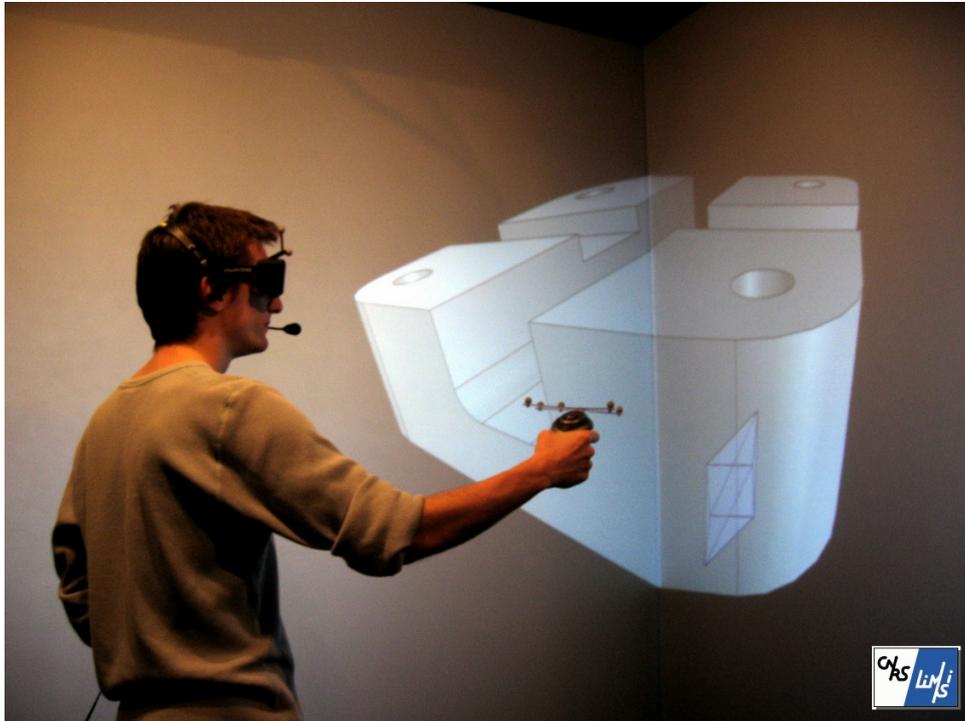


FIG. 3.8 – Interaction immersive et multimodale en entrée (voix, pointage) pour la conception dans VRAD [Convrad and Bourdot, 2003].

3.2.3.2 Interactions multimodales pressenties

La CAO peut également être un domaine d'application adéquat pour la multimodalité en sortie en RV. L'assistance à l'utilisateur peut prendre plusieurs formes. L'interaction haptique semble la plus naturelle, car c'est le canal sensori-moteur qui est prépondérant lors de l'activité réelle, que ce soit pour dessiner un plan de conception, faire une sculpture, un moulage, sentir la matière ou les finitions, etc. Un dispositif à retour d'effort peut être utilisé pour assister le modelage de la forme. Si le modèle de la scène le permet, par exemple s'il est basé sur des voxels, le modelage peut être effectué par creusage : le rendu haptique doit communiquer les efforts de collision, de pression, de creusage de la matière, suivant la métaphore de la sculpture, du tournage ou de la poterie. Si la représentation est à base de B-Rep, le modelage peut être réalisé par déformation des éléments topologiques : l'assistance haptique peut alors prendre la forme d'une force de rappel vers la position d'origine (métaphore du ressort), d'une force d'attraction vers des points de l'espace, par exemple pour positionner un élément à la même hauteur qu'un autre (grille

haptique ou points de référence), ou encore d'un guide contraignant la manulation selon un axe ou un plan. Cette manipulation suppose en amont une sélection des éléments topologiques, qui peut aussi être assistée haptiquement : l'effecteur virtuel ("proxy") est attiré (différents profils d'attraction sont possibles) par les éléments les plus proches (figure 3.9) ou bien contraint sur certains d'entre eux (goulotte pour les arêtes, creux pour les sommets). Une autre interaction naturelle en conception est le ressenti des surfaces. Pour cela, le système peut transmettre haptiquement à l'utilisateur le contact physique solide ou déformable entre l'effecteur virtuel et l'objet étudié, mais également les textures des surfaces de cet objet. Si plusieurs objets sont manipulés, par exemple pour la conception de pièces mécaniques destinées à s'emboîter ou à coulisser entre elles, il est possible de simuler les forces et les frottements pendant leur assemblage, et ainsi vérifier que les futurs opérateurs des chaînes de montage pourront effectuer leur tâche correctement. Un dernier rôle possible de l'haptique peut être l'aide à la saisie de valeurs numériques en remplacement d'un périphérique intrusif comme le clavier : reprenant la métaphore de la règle ou l'utilisation de la molette de la souris, il pourrait être utile d'ajouter des crans haptiques lors de la manipulation d'un élément, par exemple pour augmenter la précision d'une extrusion, ou du moins avoir un indice de distance et de vitesse. A plus haut niveau, ces interactions pourraient être exploitées pour aider à la formation des concepteurs et à l'apprentissage de techniques de conception.

Le rendu sonore pourrait substituer ou être combiné à ces rendus haptiques. Les assistances par ressort ou attraction haptique sont fondées sur des calculs de distances et doivent communiquer à l'utilisateur la variation de ces distances, le dépassement d'un seuil ou l'approche d'une valeur cible. Il serait donc tout à fait envisageable de transmettre ces informations par une variation (de fréquence, d'intensité ou de timbre) d'un son, ou par des earcons d'alerte. Le ressenti de textures et de frottement peut également être augmenté par des sons réalistes.

Dans le cadre de nos travaux, cette application n'a pas été utilisée au même niveau que les deux précédentes. VRAD étant la seule application disponible au début de la thèse, nous avons pu nous en servir en amont de nos recherches, pour imaginer des situations concrètes de multimodalité en sortie, des rendus réalisables et potentiellement utiles aux utilisateurs. Par la suite, ne faisant pas partie des applications visées par le projet CoRSAIRE (exploration de masses de données scientifiques), elle a été moins analysée au niveau ergonomique qu'au niveau pratique. VRAD apparaît donc aux deux extrémités de la chaîne de conception de notre superviseur multimodal : les premières idées menant à l'ébauche du modèle ; et la future évaluation du moteur, pour laquelle l'inconvénient cité précédemment devient un avantage : n'ayant pas été étudiée dans la boucle d'affinage du superviseur de rendu, elle sera utile pour juger du caractère générique ou universel du moteur.

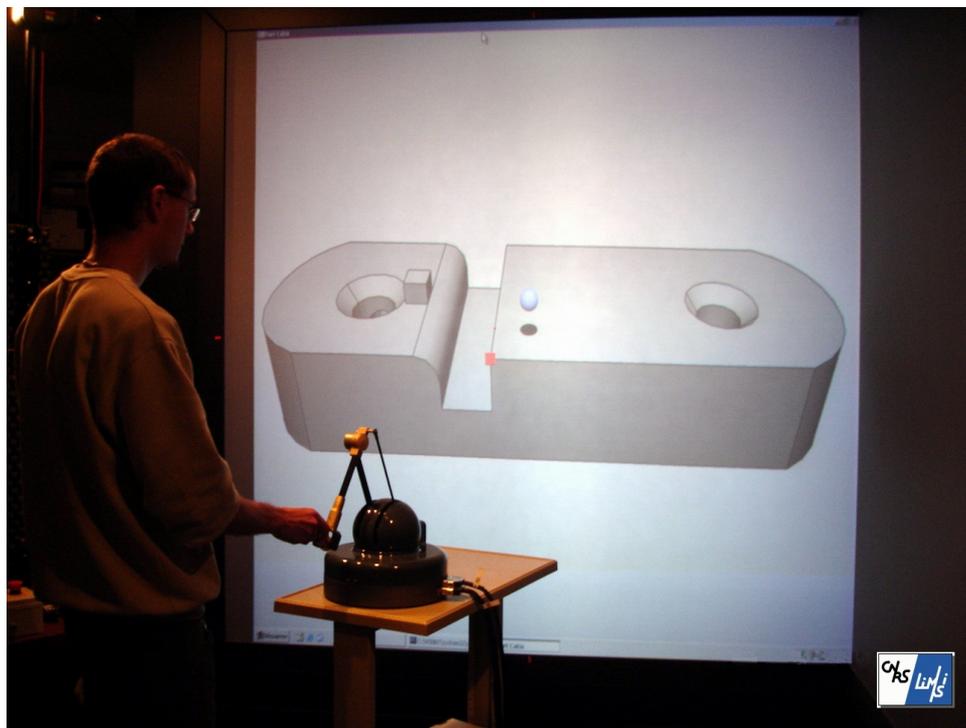


FIG. 3.9 – Interaction haptique pour la sélection de formes : l'effecteur virtuel est attiré par le sommet rouge [Picon et al., 2007].

3.3 Analyses ergonomiques

Après avoir imaginé certains rendus multimodaux dans les trois applications, une analyse ergonomique détaillée a été menée dans les deux applications scientifiques par Julien Nelson et Jean-Marie Burkhardt [Nelson, 2006c] (Unité d'Ergonomie de l'Université Paris 5) dans le cadre du projet CoRSAIRE. Les objectifs étaient tout d'abord de modéliser précisément l'activité des utilisateurs (dans notre cas, des chercheurs en Biologie Moléculaire et en Mécanique des Fluides) et de formaliser leurs besoins. Cette analyse fine de l'activité (nature des tâches, objectifs, besoins en terme d'informations et processus mentaux mis en oeuvre) a servi à formuler des recommandations pour la conception de nouvelles interactions multimodales et immersives pour qu'elle soient utilisables et utilisées.

3.3.1 Méthodes d'analyse

La première étape de l'analyse a été la modélisation des représentations mentales des utilisateurs et leur évolution dans le temps. En effet, les travaux du chercheur, dans les applications étudiées, vont lui demander de se construire une image mentale

de la masse des données qu'il explore, afin d'accéder correctement aux informations qui l'intéressent, puis de les manipuler. Une représentation mentale n'est ni fixe ni unique, elle va être reconstruite ou transformée en fonction de l'activité. Le concept d'image opérative [Ochanine, 1992] [Sperandio, 1988] est une représentation mentale particulière. En effet, dans les applications de masse de données, les opérateurs vont faire un travail important de filtrage et de hiérarchisation, afin de ne retenir que les informations qui les intéressent et de minimiser l'effort cognitif. L'image mentale qu'ils vont se construire va donc être spécialisée à une tâche, à un objectif précis. Ce concept est particulièrement intéressant pour la conception d'IHM : la succession de ces images opératives va permettre de connaître les étapes du raisonnement, leur enchaînement, les besoins en terme d'informations ou d'assistance pour chacune de ces étapes et les modalités nécessaires.

La formalisation de ces "images opératives" dans les applications a été faite en plusieurs étapes. D'abord, des entretiens semi-dirigés ont permis de comprendre les concepts centraux de l'activité et leur articulation, ainsi que les techniques et le langage opératif utilisés. Ensuite, l'observation de l'activité avec raisonnement à voix haute (technique des "protocoles verbaux" [Bisseret et al., 1999]) a servi à identifier les procédures et les enchaînements de tâches réalisés, les interactions entre l'homme et sa machine, les cas de charge mentale élevée ou d'image mentale confuse (indiqués par les temps morts). Les gestes réalisés par les opérateurs donnent également des indices de leurs images mentales "dynamiques". Enfin, une analyse de traces écrites a été réalisée sur des articles scientifiques, qui sont de véritables modèles réduits et finalisés du raisonnement. Les formules éventuelles donnent également une information sur les variables manipulées, et les représentations graphiques de différents types peuvent mener à des recommandations sur les modalités de sortie à utiliser. Les entretiens et les écrits ont été soumis à une ACD (Analyse Cognitivo-Discursive) [Kintsch and Dijk, 1978] [Ghiglione et al., 1998] à l'aide du logiciel *Tropes*. Cette analyse permet, à partir de la connaissance de divers champs lexicaux (ou univers sémantiques) d'identifier les concepts centraux du discours et leurs relations sous la forme d'un graphique : le diamètre des sphères est proportionnel à la fréquence d'apparition de la référence dans le texte, la distance entre les sphères est la distance moyenne entre les références dans le texte (figure 3.10).

Une fois les représentations mentales identifiées, une analyse de l'activité avec le formalisme HTA (Hierarchical Task Analysis) a permis d'obtenir un modèle de l'activité. Celui-ci présente les objectifs, les tâches, les procédures, mais aussi les besoins et les contraintes pour les données à présenter et leurs modalités de présentation, ainsi que le contenu et la forme de l'assistance à apporter éventuellement à l'utilisateur.

Une difficulté commune aux deux applications scientifiques que sont la Mécanique des Fluides et la Bio-Informatique est la grande diversité des tâches, due au grand volume de données et à l'existence de nombreuses problématiques de recherches. Cette

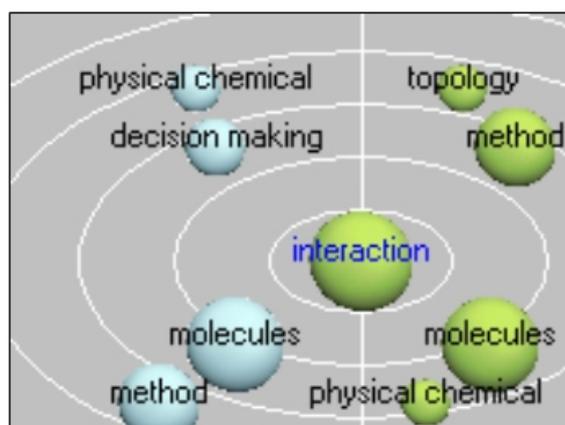


FIG. 3.10 – Exemple de résultat d'ACD par le logiciel Tropes en Bio-Informatique. Le graphique présente les concepts identifiés comme proches de l'univers "interaction" : celle-ci a lieu entre des "molécules" et fait l'objet d'une "prise de décision" à propos de sa qualité. Ce jugement est fondé en particulier sur les propriétés "physico-chimiques" et "topologiques" des protéines [Nelson, 2006a].

contrainte a conduit les ergonomes à suivre une approche par "scénarios d'utilisation" [Go and Carroll, 2003]. Le travail commence par l'élaboration d'un scénario (une tâche concrète et détaillée d'utilisation de l'application) qui permet d'identifier un certain nombre de besoins. Puis un prototype est développé sur la base de ces besoins. Il est ensuite enrichi en intégrant d'autres scénarios.

3.3.2 Appariement de protéines en Bio-Informatique

3.3.2.1 Analyse de l'activité actuelle

Actuellement, les recherches sur le docking utilisent massivement les capacités calculatoires des ordinateurs. L'activité de docking consiste pour un chercheur à valider un modèle de complexe entre deux protéines produit par un algorithme, auquel il aura fourni un certain nombre de paramètres (contraintes). L'algorithme de docking procède en trois étapes : la génération des modèles de complexe, le filtrage des solutions (clustering) et enfin le classement (par une fonction de score) des configurations possibles. Le chercheur valide ensuite ou non les modèles de score élevé en les comparant à un modèle prédictif ou expérimental du complexe. Le chercheur a donc une activité en amont de l'algorithme pour réduire les calculs, et une activité en aval pour la sélection finale du résultat.

Les séances d'observation de l'activité ont porté sur différentes étapes de cette recherche : la validation expérimentale d'algorithmes, la construction de modèles prédictifs et la prise en compte de la flexibilité des partenaires moléculaires.

L'analyse des tâches [Nelson, 2006a] a permis d'identifier les zones de focalisation de l'attention et les critères de qualité nécessaires pour juger le résultat d'un algorithme de docking. Tout d'abord, les zones de focalisation de l'attention des chercheurs sont les "interfaces" : ce sont les zones d'interactions probables entre les deux protéines. Leur topologie ainsi que les résidus chimiques qui s'y trouvent sont particulièrement informatifs. Ces résidus (acides aminés dans le cas des protéines) sont d'ailleurs appelés "hot spots" ou "points chauds" car ils ont une forte probabilité de participer à l'interaction. Ensuite, lorsqu'un complexe est proposé par l'algorithme, les autres critères observés par le biologiste sont :

- la stabilité énergétique du complexe, dont les acteurs sont les forces électrostatiques, les interactions hydrophobes et le coût de désolvatation ;
- la topologie des partenaires c'est-à-dire leur complémentarité et leur adaptation topologique (déformations et collisions) ;
- la topologie du complexe, représentée par les propriétés topologiques de l'interface (taille, symétrie, distance entre les résidus, etc.) et le RMSD ou "Root Mean Square Deviation" (écart topologique entre le complexe et le modèle).

L'objectif des travaux en cours est le remplacement de la boucle Génération-Filtrage-Classement effectuée par l'algorithme par une interaction plus naturelle, plus intuitive, du biologiste avec les protéines (figure 3.11). Celui-ci doit pouvoir les manipuler librement et les positionner dans différentes configurations en fonction de son expertise et du rendu multimodal des critères répertoriés ci-dessus. Des recommandations associées à cette transmission d'information ont été émises par les ergonomes.

3.3.2.2 Recommandations de conception et discussion

La caractéristique majeure d'une application de docking en RV multimodale, et l'apport indéniable par rapport aux outils actuels, doit être de permettre la manipulation directe des deux protéines (position et orientation) et le rendu multimodal en temps réel des critères de qualité pendant la manipulation. Cette interaction permettrait une coopération intense entre le système informatique et l'opérateur humain. Le premier effectuerait des calculs complexes et présenterait des informations compréhensibles au deuxième qui évaluerait immédiatement la qualité de la configuration courante, la modifierait en fonction de ses connaissances et du guidage virtuel, pour arriver au meilleur complexe possible. Les avantages de cette manipulation directe sont principalement d'explorer les solutions possibles issues de l'expertise qu'a le biologiste de ce problème et donc de paramétrer les algorithmes de score à des cas vraisemblables. A plus long terme, il s'agit de donner à percevoir la dynamique de formation du complexe.

Les fonctionnalités élémentaires permettant d'arriver à cet idéal sont :

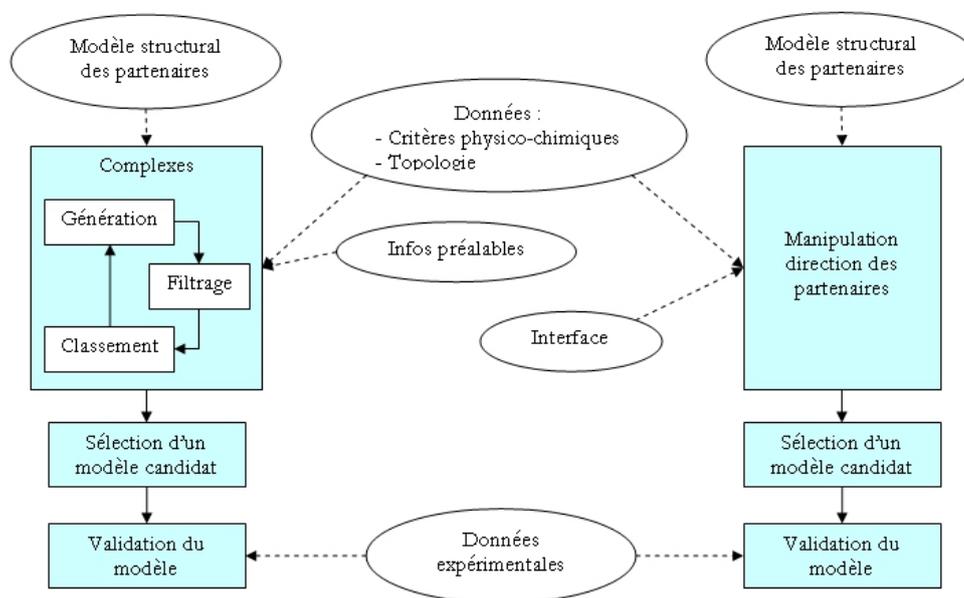


FIG. 3.11 – Comparaison du processus de docking classique et du processus visé en RV [Nelson, 2006a].

- Des interfaces permettant de manipuler les protéines dans l'espace et ainsi de modifier la configuration ;
- La présentation simultanée d'informations, de manière multimodale ;
- Le guidage vers les conformations pertinentes (via l'appel itératif à l'algorithme de score).

Enfin, il apparaît qu'un autre outil essentiel de l'application de docking serait un système de "snapshot", c'est-à-dire une sauvegarde d'état permettant au chercheur d'enregistrer puis de retrouver facilement les modèles pertinents ou les configurations encourageantes. Il faudrait pour cela enregistrer tous les paramètres de la simulation pour l'état courant du docking (espace, temps, échelle, point de vue, etc.). Il est pertinent de sauvegarder également les interactions qui ont précédé cette configuration intéressante : leur rendu composent une partie du contexte, des informations, qui a mené à la décision d'enregistrement. Il faudrait enfin pouvoir annoter la configuration. Comme pour tous les rendus, ce n'est pas le moteur multimodal qui va être en charge de la capture d'état réelle. Mais pour notre étude, il est important de connaître les paramètres à conserver et surtout à quel moment cette sauvegarde sera nécessaire. En effet, en dehors d'une commande de sauvegarde explicite de l'utilisateur, on peut imaginer que des situations particulières ou des événements puissent réclamer une capture, auquel cas le superviseur devrait 1) continuellement observer les éléments adéquats du contexte, 2) déclencher la capture au moment opportun.

L'analyse de l'activité par les ergonomes ne permet pas seulement d'identifier les fonctionnalités attendues, mais propose également des règles pratiques pour le rendu multimodal, qui peuvent servir de base avant l'évaluation. En particulier, nous pouvons distinguer trois caractéristiques des variables étudiées en Bio-Informatique, dont vont dépendre les modalités et canaux de rendus (figure 3.12) :

- La persistance. Les variables pérennes n'évoluent pas au cours de l'interaction et concernent les protéines isolées (dont les caractéristiques sont donc fixes). Les variables transitoires varient au cours du temps, en fonction du complexe courant manipulé par le chercheur : l'interaction entre les protéines modifie par exemple les forces électrostatiques et la stabilité énergétique de la configuration. Il convient d'éviter les modalités auditives lorsque la donnée à observer est pérenne : celles-ci ne variant pas au cours du temps seraient filtrées par l'oreille humaine ;
- La directionnalité. Les variables transitoires liées à l'interaction peuvent être favorables ou défavorables (ou les deux) à la formation d'un complexe. Ainsi, les forces électrostatiques sont bidirectionnelles, car elles attirent ou repoussent les protéines selon le signe des charges qui entrent en jeu. À l'inverse, les forces hydrophobes ne peuvent que favoriser la formation d'un complexe, en attirant deux zones hydrophobes. Cette dimension des variables doit donc être conservée par les modalités de rendu ;
- Le mode d'exploration des données. Le mode global a pour objectif de repérer des interfaces potentielles sur les protéines. Le mode local vise à valider l'hypothèse d'interaction entre deux zones repérées en mode global. Ces deux niveaux d'observation vont influencer le choix du rendu, en particulier leur résolution (ou niveau de détail), ainsi que leur échelle ou encore leur dimension spatiale.

Nous pouvons représenter autrement ces recommandations en fonction des caractéristiques. La figure 3.13 présente la répartition sur les trois canaux de rendu des critères particuliers manipulés par les biologistes, selon qu'ils concernent une protéine isolée, deux protéines en interaction ou un complexe formé. Schématiquement, les données d'interaction entre deux protéines peuvent être rendues auditivement ou haptiquement, car elles sont à la fois transitoires et directionnelles. Au contraire, les caractéristiques pérennes concernent les partenaires isolés et vont être transmises visuellement (le rendu visuel étant de plus traditionnel à la discipline, largement répandu et immédiatement compris).

Types des variables	Définitions / Exemple	Recommandations
Pérennes	Constantes durant l'interaction : caractéristiques des protéines isolées (ex : topologie, résidus)	Eviter la modalité auditive (car filtrage des signaux auditifs persistants)
Transitoires	Variables durant l'interaction : caractéristiques de la configuration de docking courante (ex : stabilité énergétique)	Aucune modalité à proscrire Ex : retour haptique direct, signal sonore périodique...
Bidirectionnelles	Peuvent favoriser ou bien défavoriser l'interaction suivant leur valeur (ex : charge électrostatique)	Modalité comportant une échelle et un point neutre Ex : attraction ou répulsion haptique, signal sonore variant par 2 paramètres (pitch et intensité, utilisation d'harmoniques...)
Unidirectionnelles	Ne peuvent influencer l'interaction que dans un seul sens (ex : hydrophobicité)	Ex : code couleur commun, pitch sonore aigu ou grave, amplitude forte ou faible
Globales	Concernent l'ensemble de la protéine ou l'ensemble d'une configuration de docking (ex : pour identifier des zones d'interaction)	Résolution faible Ex : - Topologie : contour visuel - Résidus : code couleur ou balise auditive - Hydrophobicité : patchs colorés - Electrostatique : non utilisé
Locales	Concernent une seule partie de la protéine ou d'une configuration de docking (ex : pour évaluer la force d'une interaction)	Résolution élevée Ex : Topologie : contour visuel Topologie, hydrophobicité, électrostatique : retour haptique

FIG. 3.12 – Recommandations de rendu en fonction des caractéristiques des données de Docking [Nelson, 2006a].

3.3.3 Exploration de cavité en Mécanique des Fluides

3.3.3.1 Analyse de l'activité actuelle

En Mécanique des Fluides, la simulation numérique d'écoulement est utilisée comme un outil pour la résolution de problèmes. Pour un problème donné, c'est-à-dire un phénomène non expliqué, le chercheur va tout d'abord essayer de trouver un modèle existant proche². Une simulation va ensuite être calculée à partir de ce modèle inadapté, et donner un ensemble de données. Pour un instant t de cet ensemble,

²Nous pouvons remarquer que le biologiste fait également des comparaisons avec des dockings connus, mais en fin de processus (validation des sorties de l'algorithme).

Information	Visuel	Audio	Haptique
Topologie : - Surface (partenaire isolé) - Résidus (partenaire isolé) - Collisions (interaction) - Complémentarité (interaction)	• • •	• •	• •
Electrostatique : - Champ (partenaire isolé) - Forces (interaction)	•	•	•
Hydrophobicité : - Patches (partenaire isolé) - Forces (interaction)	•	•	•

FIG. 3.13 – Distributions multimodales possibles de certains critères de docking (d'après [Nelson, 2006a]).

on va construire une image opérative sous forme de coupes 2d de l'écoulement (en calculant de nouvelles variables et en paramétrant leur présentation visuelle). Le chercheur va ensuite repérer des points d'intérêt dans cette représentation statique, puis faire un balayage temporel (à l'aide de nouvelles images opératives) pour suivre leur évolution. En comparant ces images avec celles du modèle théorique, il va inférer sur la cause des différences, puis essayer de transposer ses hypothèses dans le modèle.

L'analyse de l'activité [Nelson, 2006b] a montré que les éléments caractéristiques étudiés principalement (les zones d'attention) étaient les "vortex". Pour les observer, il faut calculer la "vorticité" à partir des données, puis la présenter sous forme d'une isosurface : les structures tourbillonnaires se traduisent alors dans l'image opérative par des tubes. Ces vortex sont dynamiques : une fois repérés, ils sont donc suivis dans le temps et l'espace.

Trois problèmes ont été identifiés en terme d'interface. Tout d'abord, il est difficile d'anticiper les besoins en information. Une fois la simulation calculée (ce qui prend des heures en terme de calcul et des mois en terme d'organisation matérielle), le chercheur peut se rendre compte qu'il n'a pas la bonne échelle ou les données suffisantes pour observer le phénomène. Ensuite, l'activité du chercheur est double : il va d'abord explorer la simulation numérique, c'est-à-dire suivre les données dans l'espace et le temps, puis traiter les données, formuler des hypothèses, faire des cal-

culs avec les moyens traditionnels (papier, crayon, etc.). Les parts de ces activités sont respectivement de 25% et de 75% du temps total du chercheur. Enfin, il est également difficile de trouver les bonnes représentations graphiques à utiliser, pour d'une part présenter les éléments et les caractéristiques pertinents, et d'autre part masquer ceux qui ne le sont pas.

Les difficultés touchent donc tous les secteurs de l'activité : les informations à présenter, les modalités de présentation et l'interface matérielle et logicielle qui doit respecter l'activité première du chercheur.

3.3.3.2 **Recommandations de conception et discussion**

L'analyse de l'activité a donc permis d'identifier les besoins des chercheurs et les limites des systèmes actuels, et donc d'imaginer les fonctionnalités et les interactions multimodales utiles à développer.

Tout d'abord, l'application idéale devrait être dynamique, c'est-à-dire permettre l'exploration de plusieurs instants de la simulation. Plus particulièrement, les chercheurs voudraient suivre les vortex dans l'espace et le temps. Le calcul et l'animation temps réel de la scène pose d'énormes problèmes techniques, en raison de la masse de données à simuler et à stocker (plusieurs jours de calcul et plusieurs Go de données pour chaque instant) puis à représenter, ce qui nécessite d'énormes ressources graphiques pour satisfaire la fluidité entre chaque instant. Ce rendu temps réel n'est pas possible à l'heure actuelle. De plus, en dehors de la rapidité, il faut que la navigation soit simple entre les différents instants, et que le chercheur puisse animer la scène dans les deux sens, ou encore puisse la stopper. Ces thèmes de recherches (animation visuelle temps réel de masses de données et paradigmes d'interaction pour l'animation) de ne sont pas concernés par cette thèse. L'étude multimodale prend cependant en compte cette contrainte, dans une dimension prospective, puisque le modèle doit prévoir cette fonctionnalité future, mais elle permet également de la contourner momentanément, en décomposant cette exploration idéale en sous-étapes d'exploration statique. Par exemple, une fonctionnalité moins ambitieuse mais accessible actuellement est le suivi de l'évolution temporelle de certaines variables, dont la valeur a été calculée sur plusieurs instants. L'utilisation des modalités audio est particulièrement adaptée à ce rendu.

La navigation dans le temps et l'espace peut être perturbante, notamment dans une activité d'analyse. C'est pourquoi il est nécessaire que l'interface permette au chercheur de se repérer dans l'espace (position et orientation) et le temps. Pour cela, des "balises" pourraient être positionnées par le chercheur dans l'espace 4d afin de structurer son activité. Puis, sur demande, la position de celles-ci serait indiquée au chercheur : par un rendu visuel (e.g. flèche), audio (source spatialisée) ou haptique

(guidage). La localisation de la balise dans le temps pourrait également être représentée par une échelle de temps visuelle comme une barre de défilement, ou par une échelle sonore localisée. Les modalités exactes les plus adéquates à cette "boussole multimodale" doivent être identifiées par l'évaluation future, mais le modèle doit prévoir cette fonctionnalité.

Comme en Bio-Informatique, une fonctionnalité intéressante de l'application d'écoulement virtuel serait un système de capture d'état et d'annotation, permettant au chercheur d'enregistrer puis de retrouver facilement des points d'intérêt dans une scène.

Enfin, la coexistence entre activité d'exploration et activité d'analyse doit être permise, puisque c'est la méthode de travail des chercheurs. Cette seconde activité pose des problèmes particuliers pour une application immersive. En effet, elle fait souvent l'objet de prise (ou d'examen) de notes à l'aide de fournitures traditionnelles de bureau (crayon, papier, gomme, tableau, etc.), qui en l'état est incompatible avec l'obscurité d'une salle de RV. La transposition de cette activité n'est pas immédiate et fait partie des problèmes traités dans le cadre de la multimodalité en entrée : utilisation de widgets, reconnaissance de gestes et de parole, etc. L'analyse fait également parfois appel à de la documentation externe, issue par exemple d'une base de données ou d'une autre ressource, et qui est communiquée sous forme multimédia. La cohabitation de rendus non immersifs avec des EV est un champ de recherche ouvert. Notre thèse ne s'intéresse pour le moment qu'à la partie exploration, c'est-à-dire aux 25% de l'activité. Si l'activité d'analyse de l'application est également effectuée dans un système de RV immersive, comme nous l'envisageons (et qu'elle ne reste pas confinée au papier et au crayon), la multimodalité pourra bien sûr y être intégrée, et même très intéressante à étudier. Elle dépendra des choix d'interaction et des interfaces préconisées par les experts et les utilisateurs, et portera notamment sur les transitions entre les deux activités, voire leur simultanéité.

D'autres points intéressants ont émergés de l'analyse. Ils donnent des cas concrets de données manipulées ou observées et de modalités de rendus. Tout d'abord, les zones d'attention principales des chercheurs sont les vortex. On peut donc supposer que l'application devra gérer ces objets. Cependant, ces vortex sont des propriétés locales qui ne sont pas des objets géométriques séparables ou modalisables par des équations mathématiques. Il n'existe pas de variable qui permet de caractériser exactement un vortex. Il en résulte que leur détection ou "segmentation" dans les données 3d n'est pas aisée ni même calculable. Il existe néanmoins des grandeurs, comme la vorticité, qui affichées sous forme d'isosurface correctement paramétrées (valeur, échelle) permettent d'identifier visuellement des structures tourbillonnaires, qui pourraient être le signe de la présence d'un vortex. Une autre approche envisa-

geable est l'utilisation de techniques de traitement du signal pour tenter de segmenter les données et ainsi d'isoler les structures pertinentes. Ces signaux pourraient ensuite être rendus par différents modalités : vibration ou viscosité haptique, variation sonore localisée, etc. En résumé, une application de Mécanique des Fluides ne manipule donc pas actuellement d'objet de type "vortex", mais toujours des grilles 3d de données, sur lesquelles il faudra appliquer des opérateurs, et présenter les résultats sous une forme compréhensible et propice au repérage de vortex. Il faut donc d'une part prévoir des outils d'application des opérateurs (menus haptisés, commande vocale, etc.), d'autre part permettre le rendu adéquat des résultats.

Information recherchée	Visuel	Audio	Haptique
« Point vitesse » - valeur numérique absolue - valeur numérique relative	●	●	●
Plan de coupe - structure dans le plan - structure dans le volume - plusieurs structures dans le volume	● (couleurs) ● (déplacement) ○ (limité)	● (localisation)	○ (local) ○ (local)
Isosurface - structure	●		● (direct)
Champ de vecteurs - structure - gradient	● (volumique)	●	● (viscosité)

FIG. 3.14 – Distributions multimodales possibles de certaines données d'écoulement.

Par analogie avec les recommandations de rendu pour la Bio-Informatique, nous pouvons réaliser un tableau de distribution des données de Mécanique des Fluides sur les trois canaux (figure 3.14). Les interactions et les données à représenter sont moins clairement identifiées que dans l'application précédente. Nous avons vu qu'à partir des masses de données de la simulation, les chercheurs appliquent des représentations, en vue de repérer certaines structures telles que les vortex. Les représentations principales sont les isosurfaces, les plans de coupe et les champs de valeurs ou de vecteurs. Les isosurfaces s'apparentent à des topologies, et peuvent être étudiées visuellement ou haptiquement (rendu épistémique). Les plans de coupe sont par

nature visuels. Pour isoler des structures 3d, il faut les déplacer. Ces déplacements peuvent être assistés haptiquement. Les champs de valeurs ou de vecteurs sont difficiles à représenter visuellement. Il semble plus intéressant de les parcourir et de sonifier ou haptiser les gradients des valeurs. En plus de ces données statiques, la simulation actuelle permet de connaître la valeur de la vitesse sur plusieurs instants en certains points de l'espace. Ces données temporelles vont pouvoir être sonifiées lorsque l'utilisateur le demandera ou lorsqu'il sera à proximité.

L'analyse ergonomique des besoins ainsi que les recommandations en terme d'interactions multimodales qui en ont découlé fournissent un socle solide pour la conception des deux applications scientifiques multimodales, mais également un cahier des charges, une base de connaissances et un corpus de test pour la réalisation d'un superviseur générique du rendu multimodal.

3.4 Conclusion

L'analyse ergonomique des besoins ainsi que les recommandations en terme d'interactions multimodales qui en ont découlé fournissent un socle solide pour la conception des deux applications scientifiques multimodales, mais également un cahier des charges et un corpus de test pour la réalisation d'un superviseur générique de rendu multimodal.

Chapitre 4

Supervision multimodale en RV

Résumé. *Dans ce chapitre, nous présentons notre contribution au problème de la gestion des rendus multimodaux en RV. Il s'agit d'un moteur de supervision chargé de coopérer avec l'application et l'utilisateur pour contrôler les interactions et leurs rendus multimodaux. Nous rappelons tout d'abord notre approche du problème. Puis nous présentons notre modèle des interactions, qui va servir de format de communication entre tous les acteurs du rendu. Nous décrivons ensuite en détail l'architecture de la supervision, le fonctionnement du superviseur et son implémentation. Nous évaluons enfin ce superviseur sur des interactions recommandées par les ergonomes dans les applications du chapitre 3.*

4.1 Introduction

Comme l'a montré l'état de l'art du chapitre 2, l'utilisation de plusieurs modalités de sortie dans des applications de RV commence à se répandre et les utilisateurs sont de plus en plus habitués à manipuler des interfaces haptiques ou à entendre des sons lorsqu'ils utilisent des systèmes immersifs. Dans ces applications, les rendus multimodaux générés pour les différentes interactions possibles sont le plus souvent prédéterminés par les concepteurs et implémentés de façon ad hoc au coeur même de l'application. Si cette intégration du rendu semble la plus efficace en terme de réactivité et de simplicité de mise en oeuvre, elle ne peut pas s'adapter à des contextes d'interactions différents, comme de nouveaux environnements d'utilisation ou de nouveaux utilisateurs, à moins que ceux-ci n'aient été prévus à l'avance.

Par exemple, pour réaliser une application permettant à un concepteur automobile de tester si un moteur de lève-vitre peut correctement s'insérer dans une portière de voiture, le développeur peut décider d'utiliser, en plus du système de projection stéréoscopique, un bras à 6 degrés de retour d'effort, afin que l'utilisateur ressente de manière réaliste les efforts exercés sur les différents éléments. Dans ce

contexte, l'application est spécifiée et programmée en conséquence en tenant compte de ces interfaces. Néanmoins, si l'application est utilisée sur un autre poste avec une souris 3d à la place de l'interface haptique, ou s'il survient une situation imprévue où l'haptique est inutilisable, il est crucial de conserver le principal intérêt de l'application (i.e. le test de montage). Il faudrait alors que le changement de situation soit prévu à l'avance, et que l'information haptique soit remplacée par une autre modalité, comme par exemple l'affichage visuel d'une trajectoire optimale, la mise en valeur des éléments en collision ou la sonification des frottements entre les pièces. Le concepteur va devoir compléter son programme pour tenir compte de ces choix de rendu multimodal. Parallèlement, imaginons un concepteur d'une application d'exploration haptique d'écoulement en Mécanique des Fluides. Le problème qui se pose à lui va être le même que précédemment, et il va devoir intégrer à son tour des modifications ad-hoc pour gérer les différents cas possibles.

Notre objectif est de répondre aux différents problèmes du rendu multimodal en RV de manière globale. Pour cela, nous avons modélisé et développé un moteur de supervision de la distribution multimodale. Celui-ci répond aux spécifications suivantes :

1. **Généricité et modèle.** Le superviseur multimodal doit d'abord être utilisable dans différentes applications. Pour cela, il doit être indépendant de la réalisation technique du rendu : gestion de la scène, synchronisation des différentes boucles de rendu, communication avec les dispositifs d'interaction, etc. Il doit de plus être adaptable et évolutif face à de nouveaux environnements et besoins d'interactions, de nouveaux utilisateurs, etc. L'application et le superviseur doivent également avoir un langage commun pour communiquer, et l'application doit fournir toutes les informations nécessaires au moteur. Cela suppose un modèle et des classifications communes pour représenter les interactions, ainsi que des spécifications claires et précises sur les entrées attendues et les sorties prévues du moteur.
2. **Prise de décision.** Toutes les interactions entre l'utilisateur et l'environnement doivent être cohérentes et répondre aux mêmes règles logiques. Ce contrôle global des interactions, doit être réalisé par une entité unique : le superviseur multimodal. Le rendu multimodal suppose une prise de décision "relativement intelligente". Pour cela, le superviseur doit connaître : ce qui sera représenté (les informations à communiquer à l'utilisateur pour l'assister dans ses tâches), les moyens de représentations (les modalités), l'ensemble des éléments du contexte qui influenceront la représentation et enfin les règles de correspondance, de distribution, entre les informations et les modalités. Ces règles ne doivent pas décrire explicitement tous les cas de rendus multimodaux possibles. Nous voulons développer une représentation générique et évolutive qui factorise et limite le nombre de connaissances nécessaires à la prise de

décision. D'une application à l'autre, il est souhaitable que les mêmes règles de distribution soient respectées. L'élaboration de ces règles à partir d'évaluations ergonomiques et psychophysiques doit donc mener à un cahier des charges théorique du rendu multimodal, en parallèle du superviseur.

3. **Coopération.** Le superviseur de rendu multimodal doit respecter l'essence de la RV, c'est-à-dire la coopération entre un (ou des) utilisateur(s) humain(s) et un système, et donc favoriser l'échange intuitif d'informations. Le superviseur doit également respecter les choix de rendu de l'utilisateur (lorsque c'est possible et sans danger) afin que celui-ci reste maître de sa tâche. Le rendu doit donc être géré par le superviseur, lui-même au service de l'utilisateur.
4. **Evaluation.** Nous pensons enfin qu'un superviseur multimodal serait un outil puissant pour évaluer l'utilisation de la multimodalité dans des applications de RV. Dans cette optique, il doit pouvoir enregistrer les choix de l'utilisateur et le niveau d'acceptabilité, d'utilisation et d'utilisabilité de la multimodalité de la part des utilisateurs finaux et des concepteurs d'applications.

Une première étape pour que notre superviseur respecte les spécifications n°1 et n°2 est d'établir quelles sont les interactions possibles dans des applications de RV. Sur la base des applications analysées dans le chapitre 3 et de notre expertise des EV, nous présentons notre modèle des interactions en RV. Il est composé d'un modèle des tâches (ce qui sera représenté) et d'un modèle du rendu multimodal (les moyens de représentations). Puis nous décrivons le principe général de la supervision, l'architecture du système, le fonctionnement du superviseur (la prise de décision) et son implémentation. Nous appliquons ce superviseur sur des interactions concrètes identifiées dans les applications du chapitre 3.

La figure 4.1 reprend les différents éléments qui participent à la conception du superviseur : 1) un moteur de rendu multimodal, utilisant un cahier des charges du rendu multimodal, composé 2) d'un ensemble de règles de distribution pouvant s'appliquer à différentes applications, ainsi que 3) d'un modèle des interactions servant de langage de communication entre les applications et le moteur, et donnant lieu à 4) des évaluations ergonomiques.

4.2 Modélisation des interactions de RV

4.2.1 Méthode et objectifs

Le concept de supervision suppose que les différents éléments de l'architecture communiquent des informations au sujet des interactions et des rendus (spécification n°1). Nous avons donc besoin d'un modèle pour que cette communication soit

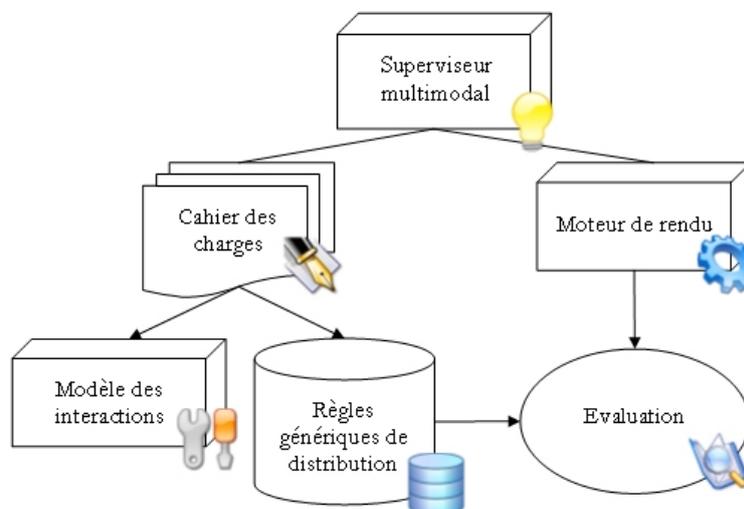


FIG. 4.1 – Rôles et retombées de la supervision multimodale.

possible et cohérente. Ce modèle doit également être tourné vers les critères de choix du rendu.

En étudiant les trois applications décrites précédemment, nous avons élaboré un modèle des interactions possibles dans les applications de RV. Celui-ci ne se veut pas exhaustif mais propose une représentation étendue des interactions qui peuvent nécessiter un rendu multimodal orienté tâche. Notre modèle des interactions est composé d'un modèle des tâches effectuées par les utilisateurs, et d'un modèle des rendus multimodaux, effectués par le système. Ces deux modèles permettent d'obtenir la traduction des interactions effectuées par les utilisateurs en un couple (tâche, rendu), composé d'éléments sémantiques simples, complets, compréhensibles et manipulables par le système¹. Ils résument le rôle que doit remplir le système vis-à-vis de l'humain, et ce rôle est ensuite concrétisé par un rendu multimodal.

Remarque Cette décomposition en éléments simples est proche du concept d'Unité d'Information Élémentaire (UIE) rappelé par Vernier [Vernier, 2001] : l'information à transmettre à l'utilisateur est décomposée au préalable et si c'est possible en sous-informations. Ce sont ces composants qui vont ensuite être communiqués.

¹NB : Les concepts que nous présentons dans la suite de cette section font l'objet d'une terminologie, qui résulte de compromis sémantiques visant à exprimer synthétiquement les définitions de ces différents concepts. En tant que compromis, les termes choisis sont discutables. Néanmoins, ils ont aussi un rôle opératif, puisqu'ils sont utilisés dans les programmes que nous avons développés, sous forme de labels.

4.2.2 Modèle de la tâche

Notre modèle des tâches est résumé dans la figure 4.2. Cette représentation est constituée d'un type de tâche, de paramètres de spécialisation et de données concernées par la tâche. Les sections suivantes présentent chacun des éléments en détail.

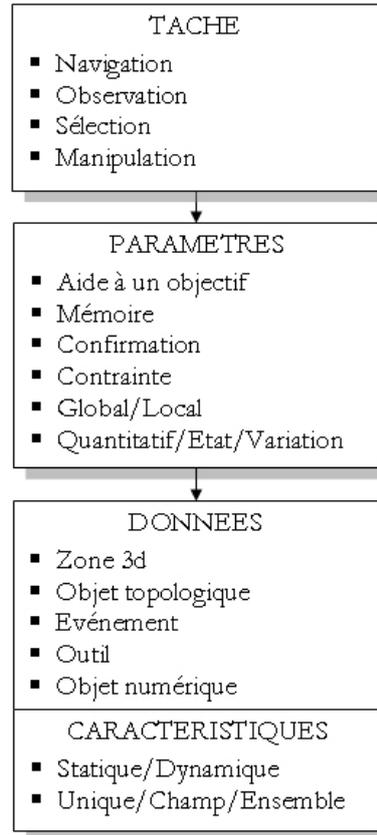


FIG. 4.2 – Modèle de la tâche en RV : un triplet (type de tâche, paramètre, donnée).

4.2.2.1 Types de tâches

Quatre types d'activités dominantes ont été identifiées dans les applications de RV orientées tâche² :

La navigation On peut séparer la navigation en une composante motrice et une composante cognitive ("recherche de chemin") [Tyndiuk and Thomas, 2001]. La composante motrice ("travel") définit le mouvement de l'utilisateur entre

²Certaines tâches de RV, par exemple le dialogue avec des humanoïdes virtuels, ont été volontairement écartées de notre modèle car elles posent des problèmes différents de choix du rendu

un point de départ et un point d'arrivée. Elle décrit sa position, son orientation et sa vitesse. Elle peut être associée à un déplacement physique (tracking de tête par exemple [Bourdote and Touraine, 2002]) ou non (joystick, commande vocale, etc.). Ce mouvement peut avoir un objectif explicite ou non. La composante cognitive ("wayfinding") est un processus de construction d'une représentation mentale de l'environnement ("carte cognitive"). Les simulateurs de conduite font largement appel à cette tâche, mais également les systèmes de nano-manipulation, afin de déterminer la meilleure trajectoire [Ammi, 2005].

L'observation C'est la perception multimodale et le traitement d'informations provenant de données de l'EV, sans modification de ces données. L'analyse d'écoulements en Mécanique des Fluides est une application privilégiée pour ce type de tâche.

La manipulation C'est l'action de déplacer spatialement ou de modifier les caractéristiques (forme, couleur, dimensions, comportement, état, etc.) d'une donnée de l'EV. La conception de formes est par exemple le siège de nombreuses manipulations topologiques : déformation, opérations booléennes, translations, etc.

La sélection C'est la désignation, à l'aide d'un dispositif de commande quelconque, d'un ou plusieurs éléments de l'EV. Cette tâche précède souvent une autre tâche, comme par exemple une observation (des caractéristiques des objets sélectionnés) ou encore une manipulation (e.g. sélection d'une face d'un objet topologique avant une extrusion).

Evidemment, dans la réalité, ces quatre catégories de tâches ne sont pas aussi clairement distinctes. En effet, l'utilisateur passe en général très rapidement d'une activité à l'autre, certaines phases ne durant parfois que quelques secondes, avant de revenir à la tâche précédente. Souvent même l'opérateur réalise deux activités simultanées. Par exemple, le déplacement dans une scène 3d à la recherche d'une cible visuelle sera traduite dans notre modèle par une combinaison de navigation et d'observation. De même, pendant une tâche de conception, la déformation d'une pièce mécanique suivant un axe pourrait faire l'objet d'une observation attentive d'une valeur numérique, comme la cote exacte de la pièce, ou sa distance par rapport à une autre pièce. Enfin, il existe également des périodes de l'activité entre deux tâches où l'utilisateur n'interagit pas précisément avec le système. Ces périodes que nous pouvons qualifier d'"intertâche" permettent de planifier, de se préparer mentalement aux tâches suivantes. Dans notre modèle, ces deux dernières nuances seront autorisées, d'une part par la possibilité pour une interaction donnée de contenir deux tâches simultanées, d'autre part par la possibilité de ne pas préciser de tâche.

Un troisième commentaire qui s'impose est que quel que soit le type de la tâche désirée par l'utilisateur, le rendu associé réalisé par le système résulte toujours en

une perception de l'utilisateur, qui est aussi une tâche d'observation. Cependant, cette observation peut avoir pour conséquence une réaction, une nouvelle commande. Cette boucle perception-décision-action va constituer la tâche désirée au départ.

Notons également qu'une tâche d'observation peut dans la plupart du temps être effectuée automatiquement par le système, sans avoir besoin d'informations issues de l'utilisateur. Ainsi le problème du choix du rendu peut être géré indépendamment des entrées du système. A l'inverse, la navigation peut par exemple faire appel aux interfaces permettant de se déplacer dans l'EV, comme le tracking de la tête, de la main, ou un joystick. La manipulation et la sélection vont elles dépendre des dispositifs de pointage utilisés : laserbeam, souris 3d, bras haptique, etc. On voit donc ici que l'observation pourra être le cas d'école qui nous permettra de débiter nos expériences sur notre superviseur, tandis que les autres tâches nous donneront des éléments de test de sa robustesse.

Enfin, un des problèmes qui se pose est la reconnaissance du type de la tâche. L'automatisation du procédé par inférence intelligente est très complexe. On peut cependant imaginer que cette inférence dépende de la connaissance par le système du domaine de l'application, ou encore qu'il reconnaisse les médias utilisés ou les mouvements de l'utilisateur. Le plus simple à ce stade de nos recherches est que l'utilisateur indique par une commande explicite (par exemple vocale) le type de la tâche qu'il entreprend.

4.2.2.2 Paramètres des tâches

En complément de son type, la tâche est paramétrée par un certain nombre de critères, qui correspondent aux besoins ou aux préférences de l'utilisateur, à la finalité de son action, à la manière dont il veut qu'elle se déroule. Ces paramètres vont donc influencer grandement sur le rendu que devra offrir le système à l'utilisateur.

Les paramètres possibles que nous avons identifiés sont :

Aide à un objectif Ce paramètre indique l'existence d'un objectif pour la tâche et le besoin d'aide pour l'atteindre de la part du système. L'objectif peut être complexe, subjectif, implicite (navigation dans un écoulement à la recherche d'un tourbillon "intéressant", ou bien observation passive en attendant que le tourbillon apparaisse) ou encore un but explicite, par exemple géométrique (extrusion jusqu'à la hauteur d'une autre pièce de CAO, déplacement d'un objet jusqu'à une cible). Le système doit alors assister l'utilisateur en améliorant son action ou sa perception. Il ne doit pas restreindre l'action de l'utilisateur (comme l'empêcher de se détourner du but) mais conduire à une meilleure réalisation (par exemple, l'optimisation d'une trajectoire de navigation). L'amélioration de la perception consiste à communiquer à l'humain des informations utiles à l'objectif (e.g. déclenchement d'un son de fréquence croissante lorsque

l'on s'approche d'une cible). Comme pour les types de tâches, l'inférence par le système de l'existence d'un objectif, tout comme la reconnaissance de celui-ci, est un problème en soit.

Mémoire La tâche nécessite de conserver en mémoire l'état précédent, soit pour que l'utilisateur puisse sans dommage annuler son action, soit parce que la tâche courante dépend des données précédentes. Ce second cas est alors lié à un paramètre "aide à un objectif" : l'objectif de l'action est dépendant des données conservées en mémoire (par exemple le déplacement d'un objet selon l'axe y d'une certaine distance par rapport à la position initiale). Ce paramètre demande un rendu particulier de l'état courant, de l'état précédent, et de la relation entre les deux états.

Confirmation Il s'agit d'assurer l'utilisateur (automatiquement ou à sa demande) que sa tâche a été correctement effectuée (par exemple déclenchement d'un son pour être sûr qu'une commande sans effet visuel a bien été traitée par le système).

Contrainte Il s'agit de la restriction des actions de l'utilisateur. Nous pouvons distinguer plusieurs types de contraintes :

- Des contraintes de guidage peuvent être induites par les besoins de l'application. Dans une optique de formation et d'apprentissage, on peut par exemple imposer des restrictions supplémentaires à un apprenant pour l'entraîner sur des manipulations spécifiques. Pour des raisons de sécurité de la tâche, pour limiter les erreurs de l'opérateur, on peut l'empêcher de déplacer un objet dans une zone 3d. Il peut enfin y avoir une volonté de sécurité matérielle, pour éviter d'endommager les dispositifs.
- La contrainte peut également être justifiée par la logique de la tâche. Par exemple, une extrusion peut être contrainte selon un axe, un déplacement d'un objet contraint par d'autres éléments fixes de la scène, etc.
- Enfin, la contrainte peut avoir pour but la recherche de réalisme. Par exemple, un moteur physique peut être intégré dans l'EV et ses calculs peuvent entraîner des contraintes spécifiques. La non-pénétration des objets en est un exemple : le rendu devra dans ce cas informer des collisions et empêcher l'interpénétration en contraignant le geste ou encore en effectuant un affichage visuel approprié de la scène.

Ce paramètre peut sembler identique à celui d'aide à un objectif. Néanmoins, la finalité et le rendu sont différents, puisqu'il ne s'agit pas d'aider l'opérateur, mais de l'empêcher d'effectuer une action.

Global / Local Ces paramètres exclusifs représentent à la fois l'étendue de l'attention de l'utilisateur et l'étendue spatiale des données concernées par la tâche. Par exemple, l'utilisateur peut être intéressé par la forme d'un objet ou par ses caractéristiques, par une valeur ou un champ de valeurs, etc. ;

Quantitatif / Etat / Variation Ces trois paramètres sont exclusifs et sont particulièrement utiles pour spécifier l'observation de données numériques. "Quantitatif" indique que l'on désire connaître la valeur explicite (ou absolue) de la donnée. "Etat" signale que l'on est intéressé par une caractéristique, une qualité de la donnée. "Variation" précise que l'on veut percevoir la fluctuation d'une donnée.

Pour une interaction donnée, il est possible de spécifier aucun, un ou plusieurs paramètres, chacun contraignant d'autant plus le rendu.

4.2.2.3 Données manipulées

Pour représenter convenablement la variété des tâches effectuées en RV, le troisième élément que nous devons connaître est le type de la donnée ou de l'information concernée par la tâche. Nous en avons identifié 5 catégories.

L'observation, la sélection et la manipulation portent sur les mêmes types de données :

Zone de l'espace Cette zone peut être à 1, 2 ou 3 dimensions et de dimension variable (d'un détail à la scène globale) ;

Objet topologique Ce sont des objets ou des parties d'objets géométriques, pouvant généralement être structurés par des arbres de composition, et approximés par des "meshs" de polygones ;

Événement Cette catégorie représente les faits ayant lieu dans la scène ou dans l'application. Ils peuvent être liés par exemple aux relations spatiales des objets entre eux (e.g. contacts entre des objets, entre le pointeur de l'utilisateur et un objet, entrée d'un objet dans une zone 3d, etc.) ou à l'état des données (e.g. transformation). D'autres types d'événements de haut niveau nécessitent un rendu adapté : les alertes demandent une réaction rapide, et donc un rendu effectué et compréhensible immédiatement ; d'autres messages complexes ne sont pas décomposables en entités simples et demandent donc un rendu le plus souvent en langage humain (texte, synthèse vocale) ;

Outil Il s'agit de certaines fonctionnalités de l'application pouvant être manipulées ou observées (menu, effecteur virtuel, fenêtre d'affichage, plan de coupe, etc.) ;

Objet numérique Ce sont des données sous forme de variables ou de vecteurs, caractérisés respectivement par un type (entier, flottant, booléen, etc.) ou une norme, une position 3d et une direction.

La navigation est un peu à part des trois autres types de tâche, car l'objet sur lequel elle porte est la scène globale. Une navigation spatiale va modifier le référentiel de l'utilisateur par rapport à tous les objets virtuels. Une navigation temporelle va modifier toute la scène de manière plus ou moins continue, en conservant le même point de vue. Les deux navigations peuvent éventuellement être simultanées.

En dehors de la catégorie à laquelle une donnée appartient, il est également nécessaire de la caractériser par sa dynamicité et sa quantité. Une donnée peut être statique ou dynamique, dans le temps et l'espace. Une tâche peut également manipuler un ensemble ou un champ de données de même type (dans nos travaux, nous avons considéré qu'une tâche ne pouvait pas porter sur plusieurs données de types différents). Ces caractéristiques augmentent encore la variété des données modélisables.

4.2.3 Modèle du rendu multimodal

Les modalités sont les différents moyens de transmettre ou acquérir une information à travers les médias (section 1.3, p.13). Pour faire du rendu multimodal en RV, il faut bien évidemment connaître les canaux sensori-moteurs, les médias permettant d'interagir selon ces canaux, les modalités de transmission d'information selon ces médias, et les différentes manières de présenter ces modalités (les travaux en IHM distinguent classiquement "l'allocation", i.e. le choix de la modalité, et "l'instanciation", i.e. le choix et le rendu des attributs de la modalité [Rousseau, 2006]). Comme nous l'avons dit dans la section précédente, nous avons choisi de n'étudier que les possibilités de rendus des médias visuel, audio et haptique (i.e. tactiles et kinesthésiques) car les dispositifs de stimuli olfactifs ou gustatifs sont encore anecdotiques et les interfaces vestibulaires très spécifiques à la navigation. Nous avons identifié des modalités concrètement réalisables qui sont disponibles dans la plupart des installations de RV actuelles ou qui sont recommandées par les ergonomes dans le cadre des analyses des applications. Nous les présentons par canaux et par catégories et les mettons en rapport avec les tâches, les paramètres ou les données pour lesquelles elles semblent correspondre. Les appellations des catégories seront réutilisées dans la suite de ce manuscrit et sont également employées dans notre implémentation. Nous étudions également les différents attributs d'instanciation et les combinaisons possibles de ces modalités.

4.2.3.1 Modalités visuelles

Les dispositifs et les moteurs de rendu visuel permettent d'ajouter ou de modifier des modalités pour les différents données que nous avons identifiées. Nous avons regroupé ces modalités en 9 catégories :

Topologie Ces modalités sont adaptées à des données qui peuvent avoir un rendu visuel sous forme de topologie (pièce mécanique, protéine, isosurface, etc.). Le rendu effectif dépend de l'application, et est géré par le moteur graphique (par exemple en OpenGL).

Texte alphanumérique Même si l'immersion peut pâtir de cette représentation

textuelle, elle est particulièrement utile pour les informations numériques ou les messages complexes.

Ajout interactif C'est l'ajout d'éléments graphiques du paradigme WIMP (menu, icône, bouton, etc.) avec lesquels l'utilisateur peut éventuellement interagir. On peut également imaginer d'autres widgets, comme une fenêtre d'affichage 2d insérée dans la scène, une boussole visuelle, etc. Tout comme le texte, les applications de RV tentent de les limiter pour ne pas pénaliser l'immersion.

Ajout réaliste Il existe de nombreux effets visuels permettant d'améliorer le rendu des éléments de la scène : effet de lumière, bump mapping, texture, etc.

Ajout sémantique Cette catégorie d'effets visuels n'a pas pour objectif d'améliorer l'esthétisme ou le réalisme mais de mettre en valeur certains éléments (e.g. reverse video, boîte englobante) ou de communiquer des informations de plus haut niveau. L'ajout de couleurs ayant un sens (compris intuitivement, ou par convention, ou par une légende) sur un rendu visuel existant en est un cas répandu : par exemple l'aplat de rouge sur des objets en collision, ou la coloration en bleu des zones hydrophobes d'une protéine.

Niveau de détail Cet effet graphique particulier consiste à modifier la précision du rendu visuel dans une zone 3d de la scène. Grâce à des algorithmes, cette représentation permet principalement d'améliorer les performances de rendu en évitant de lourds calculs lorsque ce n'est pas nécessaire (typiquement, lorsque l'utilisateur porte son attention ailleurs, ou lorsque l'échelle de la scène le permet). On peut donc imaginer jouer sur ce niveau de détail en fonction de la localisation et de la précision de la tâche.

Caméra C'est la modification du point de vue sur la scène, permettant d'optimiser la visualisation d'informations ou l'interaction. Par exemple, pour la manipulation d'un élément masqué par d'autre, le système peut décider de proposer un autre angle de vue libre [Plemenos and Benayada, 1996].

Guide C'est l'ensemble des modalités visuelles qui permettent d'apporter une assistance pour atteindre un objectif : affichage d'une trajectoire optimale pour atteindre une cible, attraction visuelle de l'acteur virtuel vers la cible, coloration des zones 3d à éviter, etc.

Contrainte Il s'agit d'appliquer visuellement le paramètre "contrainte" de la tâche. Par exemple, il est possible de découpler les objets visuels de la manipulation par l'utilisateur lorsque leur position devient incompatible avec une contrainte spatiale, ou encore de masquer certaines parties de l'environnement pour ne pas pouvoir y interagir. Pour les contraintes physiques, il s'agit de faire appel au moteur physique de l'application et se conformer visuellement à ses calculs (e.g. non interpénétration visuelle).

4.2.3.2 Modalités auditives

En ce qui concerne l'audio, Conversy présente différentes techniques sonores employées dans les interfaces [Conversy, 2000]. Cela nous permet de classer les modalités sonores en 7 catégories :

Icône auditive Dans le domaine de l'IHM, les icônes auditives désignent des sons enregistrés et utilisés pour accompagner des événements. Le but est d'utiliser des sons connus par l'humain (par exemple, des sons de la vie de tous les jours) et qui ont une correspondance symbolique ou métaphorique immédiate avec l'information à transmettre. Par exemple, Gaver [Gaver, 1989] propose d'accompagner le déplacement d'une icône graphique par un son de frottement, ou bien de signaler une activité de compilation par un son de machine.

Earcon Les *earcons* sont des petites mélodies, des motifs musicaux qui servent à donner des informations sur des objets ou des interactions. Ces motifs abstraits sont reconnaissables par apprentissage de leur rythme, de leur hauteur et de leur volume. Même si cet apprentissage n'est pas forcément difficile, il nécessite un effort de la part de l'utilisateur. Elles peuvent être combinées entre elles ou transformées.

Auralisation L'auralisation consiste à projeter des informations à une ou plusieurs dimensions sur une ou plusieurs dimensions du son, comme le volume ou la fréquence. C'est le pendant sonore de la visualisation³. En CAO, l'auralisation pourrait associer la variation de la longueur d'une extrusion d'une pièce à la fréquence d'un son. Nous pouvons ajouter à cette catégorie les techniques de traitement sonore qui accentuent ou atténuent un son, comme la distorsion ou l'assourdissement.

Synthèse non parlée C'est la manipulation de signaux sonores par des fonctions et des combinaisons mathématiques. Conversy en présente différentes classes, parmi lesquelles la synthèse physique [Conversy and Beaudouin-Lafon, 1996]. Celle-ci permet grâce à des calculs sur des modélisations physiques d'obtenir des sons réalistes lors des interactions avec les objets virtuels, qui correspondent par exemple aux forces de collision ou de frottement, aux densités et aux matériaux des objets, etc.

Parole C'est la diffusion de messages vocaux soit à partir d'enregistrements, soit par génération informatique (synthèse vocale). Comme pour l'affichage de texte, cette modalité est utile pour communiquer des informations linguistiques et/ou sémantiquement complexes à représenter via une autre modalité.

Musique L'état de l'art a montré que certains travaux donnaient un aspect musical à la sonification de données : soit en utilisant des sons issus d'instruments de

³Par exemple, sur une carte de France, on peut représenter les villes par des cercles et projeter le nombre d'habitants sur le rayon

musique, soit en post-traitant la sonification. Il semble que la musique apporte un certain confort par rapport à la sonification "brute", mais le risque est d'en masquer la dimension sémantique : par exemple, introduire des harmonies sans fondement sur les données peut distraire l'utilisateur.

4.2.3.3 Modalités haptiques

Nous distinguons 4 catégories de modalités réalisables à l'aide des dispositifs haptiques :

Epistémique C'est le rendu principal pour lequel les interfaces kinesthésiques sont apparues. Il s'agit de faire ressentir à l'utilisateur les efforts physiques entre l'effecteur virtuel qu'il manipule et les objets de la scène (collisions, frottements, déformations, textures, etc.). Comme pour le rendu physique visuel, il fait appel à un moteur physique, en particulier pour la détection de collisions ;

Guide Là encore, c'est un terme qui concerne principalement l'haptique (d'ailleurs, le mot "guide" en RV est souvent implicitement utilisé pour "guide haptique"). Il s'agit d'attirer, d'aider l'utilisateur à effectuer une action, se déplacer vers une cible, etc. Plusieurs métaphores existent, comme l'aimantation ou le ressort ;

Contrainte Au contraire du guide, il s'agit d'empêcher l'utilisateur de faire certaines actions ou certains déplacements (par exemple en interdisant d'aller dans une zone de l'espace, ou en bloquant certains degrés de liberté de l'interface). La contrainte ultime est le blocage total de l'interface ;

Haptisation Tout comme la visualisation et l'auralisation, il s'agit de projeter les dimensions des données sur les dimensions de la force (intensité, direction, fonction) ou de la vibration (intensité, disposition, fréquence, profil). On peut par exemple attirer ou repousser plus ou moins l'effecteur en fonction d'une valeur numérique signée, ou ajouter de la viscosité au mouvement lors du parcours d'un champ de vecteurs. La vibration est la modalité privilégiée des interfaces vibro-tactiles (par exemple un gant de données muni de vibreurs disposés sur les doigts). Le rendu n'est pas réaliste mais permet de transmettre des informations. On peut également imaginer des vibrations sur une interface haptique classique (même si ce n'est pas recommandé en raison des contraintes mécaniques).

4.2.3.4 Instanciation des modalités

Chacune de ces modalités peut être utilisée dans divers paradigmes, et paramétrée dans l'espace et le temps. Comme nous l'avons défini, un paradigme est l'emploi particulier d'une ou plusieurs modalités pour une information donnée. La plupart

du temps, le choix d'un paradigme va consister à choisir chacun des attributs de chaque modalité ou la façon dont ces attributs vont varier en fonction de la donnée représentée. Pour chaque canal, nous avons identifié plusieurs attributs de modalités :

- pour les modalités visuelles : dimensions, intensité lumineuse, couleur ;
- pour les modalités sonore : volume, fréquence, timbre ;
- pour les modalités haptiques : fonction, intensité et direction de la force.

Par ailleurs, certains attributs des modalités se retrouvent sur les trois canaux. En ce qui concerne l'espace, les modalités visuelles ou audio peuvent être présentées à l'utilisateur :

- en un point fixe de l'environnement réel : par exemple "dans le coin en haut à droite de l'écran") ;
- en un point fixe dans le champ de vision de l'opérateur ;
- dans une position liée à la scène.

Ce choix est guidé par les positions relatives de la scène, du dispositif et de l'utilisateur, par les charges des médias dans chaque position, par le niveau de l'information, etc.

Les modalités haptiques sont spécifiques car le dispositif est relié physiquement à l'utilisateur.

Ensuite, au niveau temporel, les modalités visuelles, audio et haptiques peuvent être :

- transmises immédiatement ou après un retard ;
- continues ou discontinues ;
- ponctuelles, répétées ou se dérouler pendant un certain temps.

Ce choix dépend bien sûr de la temporalité de l'information et de son niveau d'importance, mais également des latences potentielles au niveau des périphériques.

La figure 4.3 résume ces différentes possibilités de rendu en fonction des canaux sensori-moteurs.

4.2.3.5 Combinaisons des modalités

Les paradigmes possibles d'une modalité sur un canal sont nombreux, mais la combinaison de modalités sur un ou plusieurs canaux offre une diversité de rendu encore plus vaste.

Le principe du rendu multimodal est d'associer à une information une ou plusieurs modalités, réparties dans l'espace et le temps. Comme une application transmet plusieurs informations, l'utilisateur va être soumis à une combinaison spatiale,

Canaux	Modalités	Attributs
Visuel 	Topologie Texte Ajout interactif Ajout réaliste Ajout sémantique Niveau de détail/Caméra Guide Contrainte Physique	Espace Temps Dimension(s) Intensité Couleur
Audio 	Icône auditive <i>Earcon</i> Auralisation Synthèse non parlée Parole Musique	Espace Temps Volume Fréquence Timbre
Haptique 	Epistémique Guide Contrainte Haptisation	Temps Fonction Intensité Direction

FIG. 4.3 – Modèle du rendu multimodal en RV.

temporelle et sémantique de plusieurs modalités, soit représentant une même information, soit plusieurs informations différentes. A un instant et en un endroit donné, il peut y avoir plusieurs modalités et plusieurs informations différentes.

Les enjeux sont donc multiples. Tout d'abord, nous avons comme objectif que le système (de façon automatique) ou bien le concepteur de l'application (par un guide de programmation) puissent déterminer le meilleur rendu multimodal pour chaque information, c'est-à-dire des modalités adéquates et compatibles entre elles. Ensuite, nous souhaitons, en présence des rendus de différentes informations, que ceux-ci soient également compatibles entre eux, i.e. ne se perturbent pas les uns les autres. Enfin, le nombre de rendus est limité par les capacités perceptives et intégratrices humaines, donc le nombre d'informations communicables l'est d'autant : bien que l'une des justifications de la multimodalité est que l'ajout de canaux de communication augmente le nombre et la qualité des informations transmises, le nombre de rendus disponibles n'est pas illimité.

4.2.4 Discussion et exemples

Pour répondre au problème de la correspondance entre les tâches de RV et le rendu multimodal optimal pour leur réalisation, nous venons de proposer deux modèles. D'une part, le modèle des tâches est constitué de trois éléments (type de tâche, paramètre, donnée) qui peuvent représenter la variété des activités possibles dans une application de RV. Les composants de ce modèle des tâches renferment des indices sur le processus de raisonnement, sur les objectifs de l'utilisateur. D'autre part, nous proposons un modèle du rendu multimodal qui élabore une catégorisation des modalités et montre la diversité des représentations possibles.

Chaque élément de ces deux modèles vont intervenir dans la décision du rendu multimodal, chacun ayant des modalités privilégiées pour transmettre les informations qui lui correspondent. Par exemple, si l'on veut garder le caractère naturel de l'interaction, il est pertinent d'utiliser les modalités haptiques pour des tâches de manipulation et de sélection. A l'inverse, l'observation et la navigation sont dans la vie réelle plutôt guidées par le canal visuel, mais peuvent également se baser sur des indices audio. De même, en ce qui concerne les paramètres, les tâches réclamant de l'aide ou au contraire de la contrainte feront probablement appel à des modalités haptiques. Enfin, le rendu dépendra des données concernées. Par exemple, certaines ont un rendu visuel par défaut, tandis que d'autres ne peuvent pas être visualisées.

Le problème qui se pose immédiatement est donc la reconnaissance au sein de l'application du type de la tâche, des paramètres et de la données pour les communiquer au superviseur lors de la requête. Cette reconnaissance est triviale lorsque l'interaction est explicitement démarrée par une commande de l'utilisateur, comme par exemple la commande vocale "extrusion" dans VRAD (section 3.2.3.1, p.78), qui sera suivie du déplacement d'une face de l'objet selon un axe, correspond à une tâche de manipulation d'un objet topologique connu dans l'application. Mais déterminer automatiquement les trois constituants définissant une certaine tâche s'avère beaucoup plus difficile lorsque l'interaction est naturelle et ne correspond à aucune commande ponctuelle, ou encore lorsque plusieurs interactions sont effectuées simultanément. Dans ce cas, le système doit séparer les actions, reconnaître si le comportement dynamique est propre à une tâche, déduire ce que l'utilisateur va faire ensuite, etc. Ce problème ne sera d'ailleurs pas traité dans le cadre de cette thèse. Dans la suite de l'étude, nous considérerons que les composants des interactions sont reconnus par les commandes de l'utilisateur.

Cette modélisation des interactions peut être appliquée aux trois applications analysées dans le chapitre précédent. Nous avons choisi d'étudier plus précisément quelques interactions concrètes voulues par les utilisateurs ou imaginées par les ergonomes et les concepteurs. Ces interactions sont représentées dans la figure 4.4 et illustrées dans les figures 4.5, 4.6 et 4.7 .

Appl.	N°	Interactions désirées par l'utilisateur	Types de tâche	Données	Paramètres	Modalités réalisables
MF	1	Ne pas pouvoir traverser une isosurface	Manipulation	Contact	Contrainte	Contrainte haptique ou visuelle
	2	Etre totalement contraint sur une isosurface		Proxy		
	3	Etre attiré sur une isosurface		Variable	Objectif	Haptisation, attraction visuelle
	4	Ressentir la topologie d'une isosurface	Observation	Topologie (Isosurf.)	Physique & Local	Epistémique haptique, texture haptique
	5	Connaître la valeur exacte de la température en une position d'une simulation de liquide		Variable	Quantitatif	Texte, synthèse vocale
	6	Savoir si la température en une position d'une simulation de liquide est supérieure ou inférieure à un seuil		Variable	Etat	Texte, couleur, icône, auralisation, vibrations
BI	1	Etudier l'hydrophobicité d'une protéine	Observation	Champ de variables	Etat & Global	Couleur, topologie, icônes
	2	Ressentir les forces électrostatiques entre deux protéines		Variable	Variation	Haptisation, auralisation, <i>errors</i> , vibrations
	3	Ressentir la topologie d'une protéine		Proxy	Physique & Local	Epistémique haptique, texture haptique
CAO	1	Ressentir la collision entre la pièce qu'il manipule et un plan	Manipulation	Contact	Physique	Epistémique haptique ou visuel, coloration du contact, son de collision
	2	Avoir des informations sur le futur contact entre la pièce qu'il manipule et un plan		Pré-contact		Vecteur, informations textuelles, couleur
	3	Ressentir la topologie d'une pièce		Proxy	Physique & Local	Epistémique haptique, texture haptique
	4	Trouver la meilleure trajectoire pour emboîter une pièce dans une autre	Sélection	Topologie (Pièce)		Guide haptique ou visuel vers la position/orientation de la cible
	5	Extruder une pièce pour atteindre la cote d'une autre		Objectif		Guide haptique vers la cote, auralisation
	6	Sélectionner un élément topologique d'une pièce		Topologie (Elément)		Guide haptique vers l'élément topologique le plus proche

FIG. 4.4 – Exemples de modélisations d'interactions concrètes dans les applications de Mécanique des Fluides (MF), Bio-Informatique (BI) et Conception Assistée par Ordinateur (CAO).

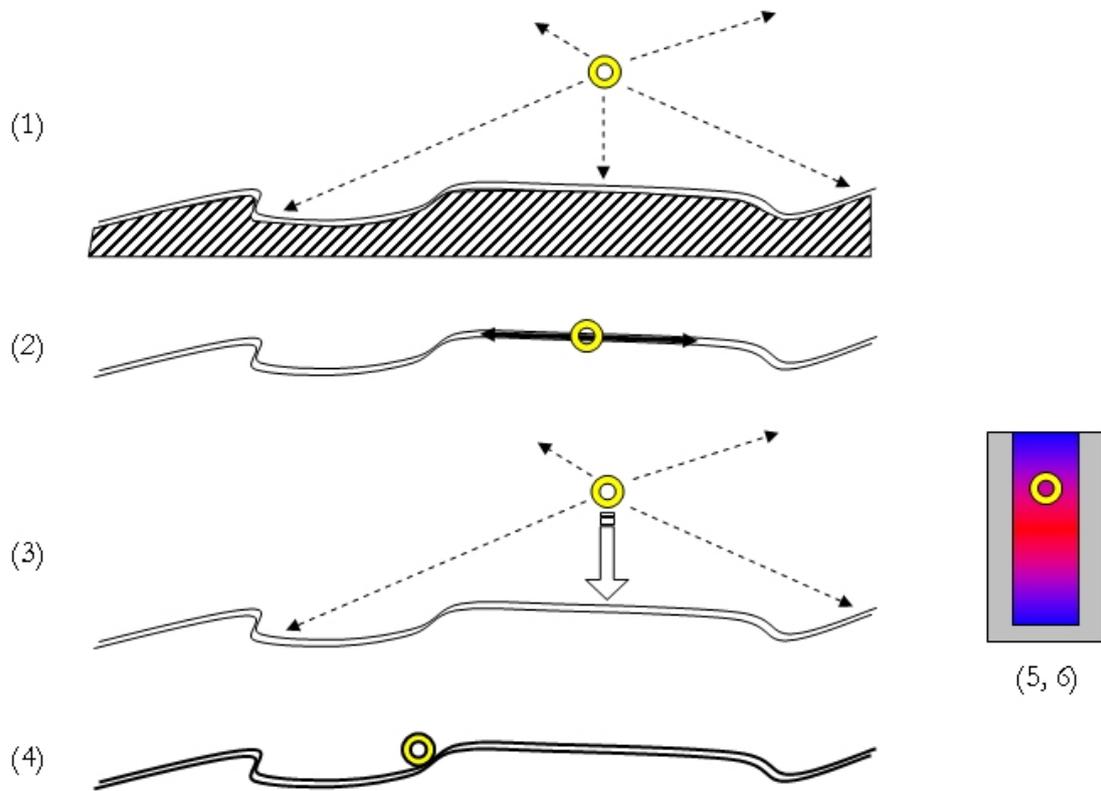


FIG. 4.5 – Exemples d'interactions concrètes en Mécanique des Fluides.

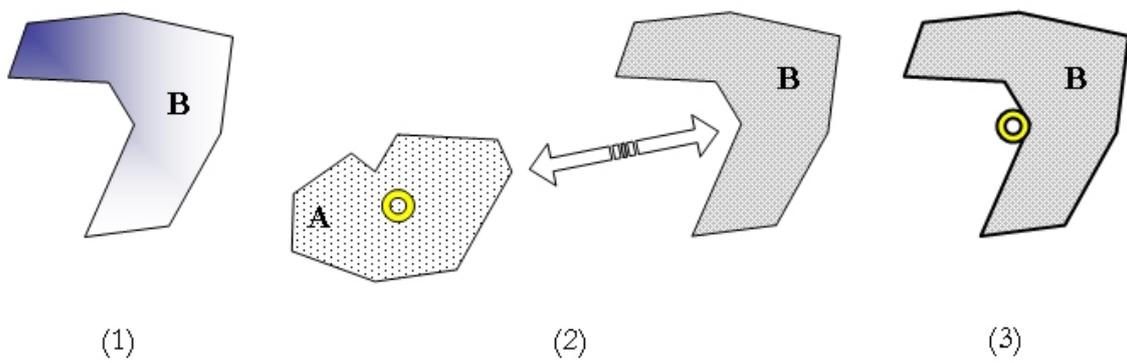


FIG. 4.6 – Exemples d'interactions concrètes en Bio-Informatique.

L'intérêt de ces exemples est d'abord de montrer que notre modélisation permet

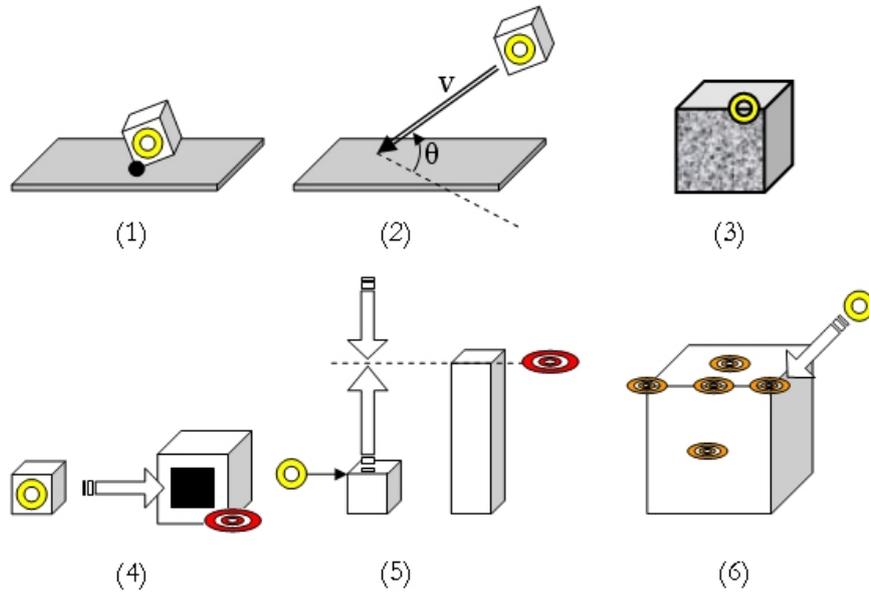


FIG. 4.7 – Exemples d’interactions concrètes en CAO.

de représenter précisément et simplement des tâches variées de RV. Ensuite, ces exemples mettent en évidence la généralité du modèle et donc la possibilité de développer un superviseur multimodal commun à plusieurs applications de RV. D’une part, ils montrent que des tâches sur des données ou dans des applications différentes correspondent en réalité à des activités semblables, qu’elles peuvent donc être modélisées de la même manière pour mener à un traitement multimodal identique. Par exemple, le ressenti de la topologie d’une isosurface (MF 4), d’une protéine (BI 3) ou d’une pièce mécanique (CAO 3), sont trois activités semblables sur trois données ayant des propriétés communes. Elles sont donc modélisées par un type de tâche (manipulation), une donnée (topologie) et des paramètres (physique et local) identiques. De même, l’aide à la manipulation par attraction soit vers la surface d’un objet, soit vers une cible, peut être appliquée aux trois domaines, sur trois objets différents, mais fera l’objet d’une seule modélisation. Le rendu sera donc le même dans des conditions identiques : l’attraction sera traduite par un guide haptique actif (entraînement), un guide visuel actif (attraction visuelle de l’effecteur virtuel) ou encore un guide visuel passif (affichage de la trajectoire).

On voit d’autre part que les paramètres sont essentiels pour permettre de différencier des tâches de même type et de même donnée, mais qui doivent mener à un rendu différent. Ainsi, dans les exemples 5 et 6 de Mécanique des Fluides, seul le paramètre est différent. Il influence sur le rôle que va devoir jouer le système pour assister l’utilisateur dans sa tâche.

Ces quelques exemples montrent que la prise en compte d'un nombre limité de tâches, paramètres et données permet de représenter un spectre large des interactions de RV.

Le tableau présente également, pour chaque interaction, des modalités de rendu possibles. Les règles de distribution multimodale vont consister à associer les composants du modèle des tâches avec ceux du modèle du rendu. Cela aboutit à une catégorisation des modalités selon leurs capacités de communication sémantique, c'est-à-dire selon leurs affinités avec les tâches, les données et les paramètres. A terme, les modalités de RV pourraient donc être classées dans des catégories sémantiques, du type : modalités d'assistance à un objectif, modalités quantitatives, modalités de manipulation, modalités d'alerte, etc.

La section suivante détaille le rôle du superviseur, les éléments qui le composent ainsi que l'architecture de la supervision des applications. Les règles de distribution sont également présentées, ainsi que le processus de décision et les choix d'implémentation.

4.3 Processus de supervision

Dans notre approche, nous laissons au concepteur de l'application la possibilité de définir les rendus multimodaux pour chaque interaction qu'il veut intégrer. Cependant, comme nous l'avons illustré en introduction de ce chapitre (section 4.1) toutes les situations d'interaction ne peuvent pas être prévues à l'avance (cf. dynamique du contexte de l'interaction en RV), c'est pourquoi une supervision des rendus est nécessaire. Cette section présente en détail le processus et les acteurs de la supervision multimodale.

4.3.1 Principe

L'architecture de la supervision multimodale repose sur quatre acteurs principaux (figure 4.8) :

Le monde réel L'utilisateur humain interagit avec le système à l'aide des interfaces. Ces interactions sont composées des commandes multimodales (entrée) de l'utilisateur vers le système, des rendus multimodaux (sortie) présentés par le système à l'utilisateur, et des différents suivis d'éléments du monde réel par le système, via des capteurs (entrée).

L'application Nous modélisons l'application comme un trio d'éléments. L'Environnement Virtuel contient une représentation virtuelle de l'utilisateur, ainsi que l'ensemble des données qui vont faire l'objet des interactions : scène, objets

topologiques, variables numériques, etc. L'application est également soutenue par une architecture logicielle et matérielle. Cette architecture gère l'EV et fait le lien entre lui et le monde réel. L'EV peut par exemple être décrit par un graphe de scène, ou les données peuvent faire l'objet de calculs par un moteur physique. Les interfaces d'entrée et de sortie sont dirigées par des pilotes et les rendus qui y sont présentés sont générés par des moteurs de rendu (graphique, sonore et haptique). Enfin, le gestionnaire d'interaction interprète les commandes de l'utilisateur, modélise les interactions (tâches et modalités), demande au superviseur de contrôler le rendu et commande à l'architecture de présenter le rendu décidé.

Le superviseur Le superviseur est en charge du contrôle du rendu multimodal. Pour cela, il détient une base de connaissances renfermant les règles de distribution des modalités, une base de règles de décision qui déterminent son comportement et une base de contexte qui contient les éléments susceptibles d'influer sur le rendu.

L'observateur/interpréteur Ce module permet la communication entre l'application et le superviseur, en faisant office de traducteur des informations échangées. En effet, le superviseur se voulant générique (cf. spécification n°1 p.94), il doit pouvoir s'adapter à différentes applications, celles-ci pouvant être programmées dans des langages propres. Cette traduction entre le langage de l'application et celui du superviseur est basée sur le modèle des tâches et le modèle du rendu. Son second rôle est d'observer en temps réel l'exécution de l'application afin d'en isoler les éléments du contexte et de les mettre dynamiquement à disposition du superviseur. Les éléments du contexte peuvent être issus du monde réel (comme les données de suivi) et informer par exemple de la position de l'utilisateur ou de ses mouvements, ou bien de l'EV et informer de l'organisation spatiale de la scène. La connaissance des capacités de rendu de l'application sont particulièrement utiles : les médias et les modalités disponibles dans l'application, les charges dynamiques des canaux sensoriels, etc.

Le processus de supervision (figure 4.9) consiste en un échange successif de messages entre ces quatre acteurs :

1. Dans le cas le plus simple, une interaction débute par une commande de l'utilisateur au système. Elle indique que l'utilisateur désire réaliser une tâche sur les données de l'EV. L'application, sur la base de critères donnés a priori et de fonctions de reconnaissance automatique, peut également être en mesure de lancer une interaction (par exemple une observation de donnée numérique, suite au dépassement d'un seuil, ou une navigation suite à la détection de changement d'état d'un objet critique).

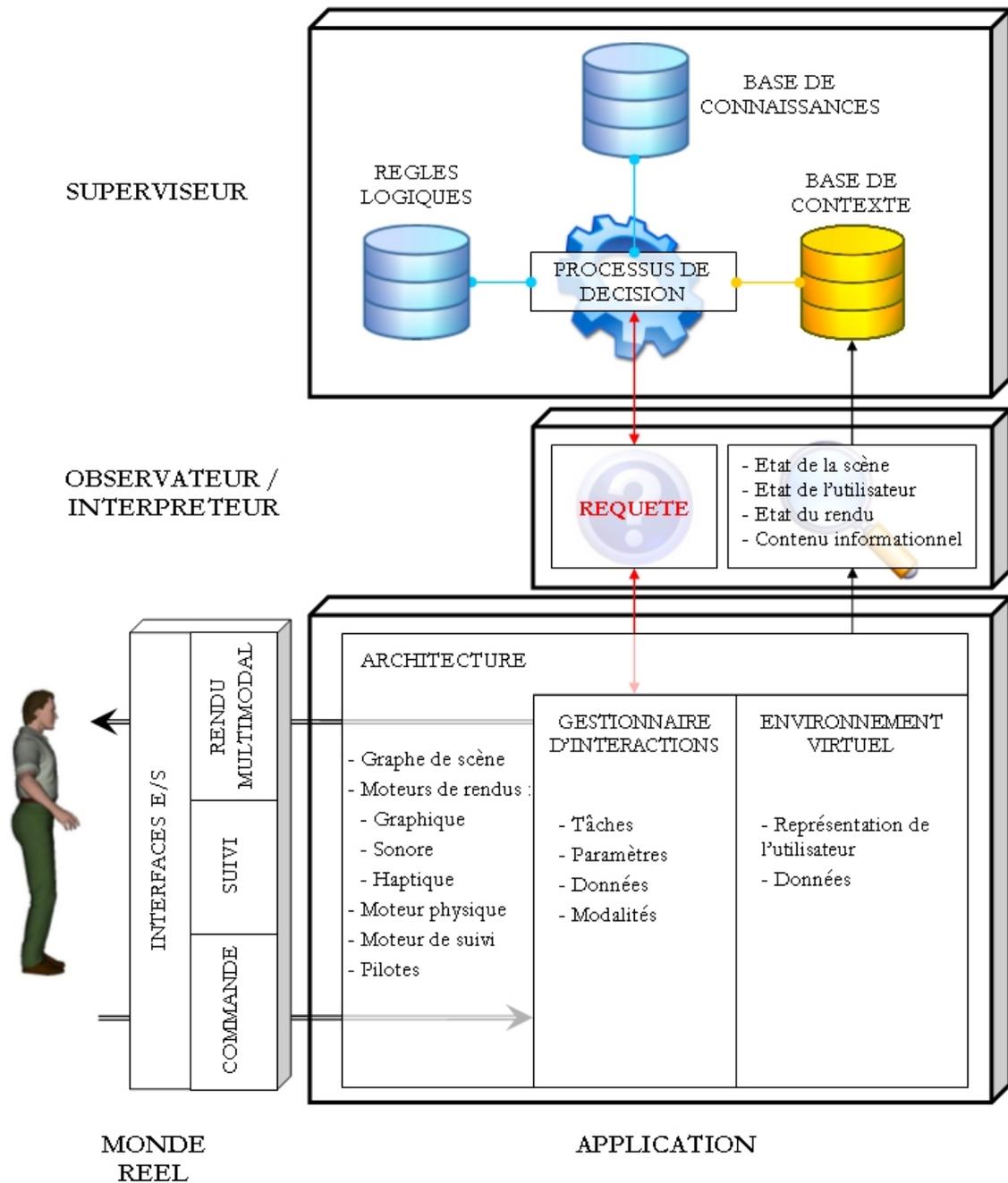


FIG. 4.8 – Architecture générale de la supervision multimodale.

2. Le gestionnaire d'interaction interprète cette commande, et modélise la tâche à effectuer par un type, un ou des paramètres éventuels et une donnée (sec-

tion 4.2). Puis il effectue une requête au superviseur, afin que celui-ci détermine les modalités les plus adéquates. Si le concepteur a prévu un rendu précis, ou si l'utilisateur en a demandé explicitement un lors de sa commande, ces propositions sont ajoutées à la requête.

3. La requête est traduite et propagée par l'interpréteur vers le superviseur. Dans le cas où aucun rendu n'est proposé, le superviseur devra le déterminer complètement. Si des propositions accompagnent la requête, le superviseur devra les valider, les refuser ou les préciser. La décision du superviseur est détaillée dans la suite de ce chapitre. Elle suit une démarche logique, à base de connaissances statiques, de connaissances dynamiques et de règles qui s'appliquent à ces connaissances. D'abord, les connaissances statiques renferment la sémantique du rendu multimodal. Elles sont composées des choix de distribution des informations sur les modalités. Ces choix sont établis à partir d'évaluations ergonomiques et psychophysiques, des analyses de besoins, des expertises des utilisateurs, etc.⁴ Ensuite, les connaissances dynamiques sont les éléments du contexte fournis par l'observateur. Enfin, un ensemble de règles logiques régit le comportement du superviseur : leur finalité est d'aboutir à une décision de rendu multimodal à partir des prédicats que forment les connaissances.

4. Le résultat de la requête (une liste d'identifiants issus du modèle du rendu et représentant des modalités) est traité par le gestionnaire d'interactions, qui effectue les correspondances entre termes et modalités concrètement disponibles dans l'application. Puis le rendu est généré par les moteurs de rendu appropriés, et transmis à l'utilisateur par l'intermédiaire des médias. Toutes les informations contextuelles sont mises à jour, tant au niveau de l'application que dans la base de contexte du superviseur.

⁴Ce sont ces connaissances qui pourraient être répertoriées dans un cahier des charges générique pour la conception d'applications de RV multimodales en sortie.

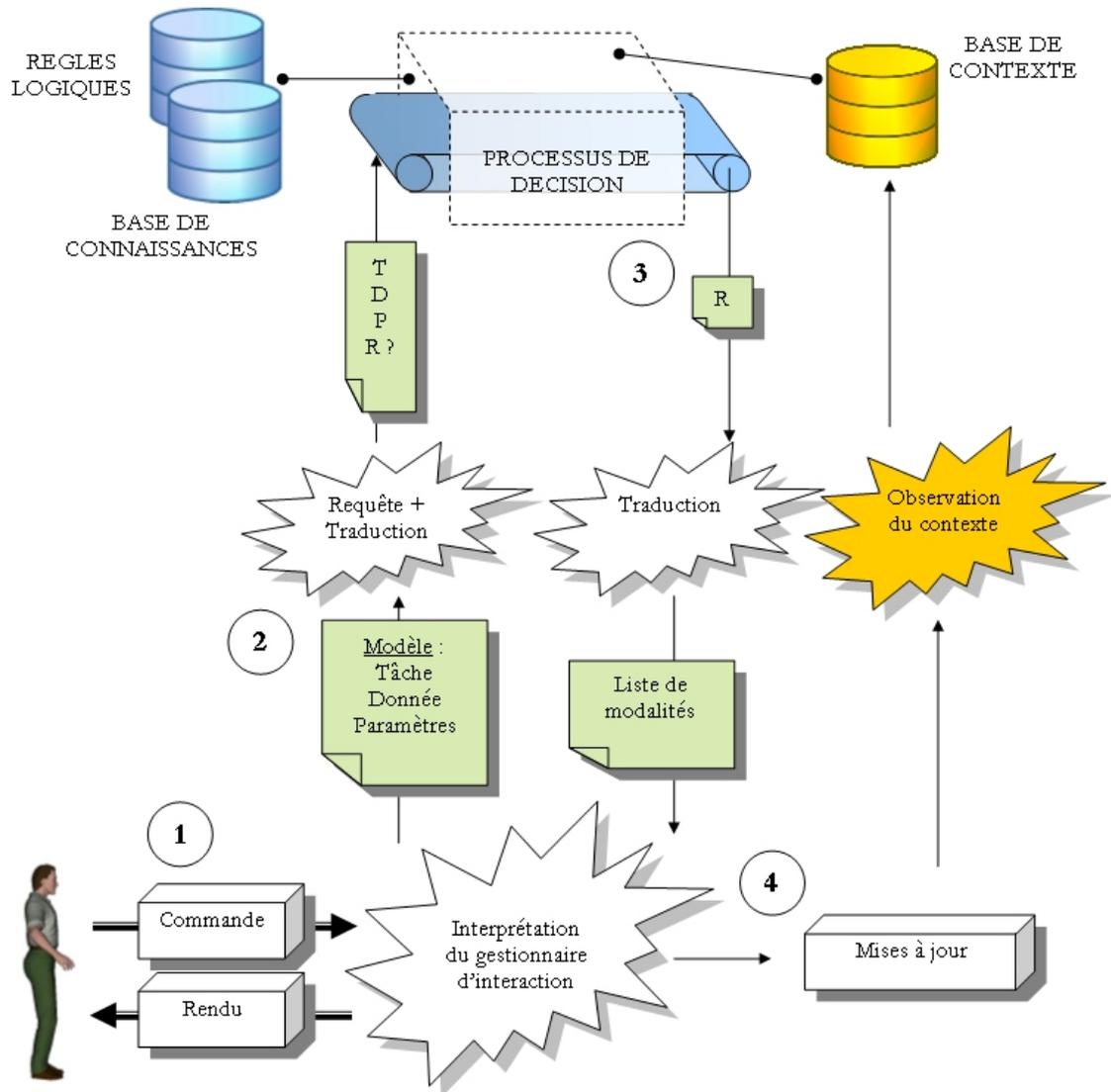


FIG. 4.9 – Principales étapes du processus de supervision entre la commande de l'utilisateur et le rendu multimodal.

La suite du chapitre présente plus en détail la mise en place des éléments de la supervision et des communications entre les acteurs.

4.3.2 Choix d'implémentation

Nous avons choisi de bâtir notre superviseur autour d'un système de décision en logique du premier ordre. Le langage d'implémentation choisi est Prolog, qui est l'un des principaux langages de programmation logique⁵. Un programme Prolog est constitué d'un ensemble de "clauses"; une clause est une affirmation portant sur des "atomes logiques"; un atome logique exprime une relation entre des "termes"; les termes sont les objets de l'univers.

Tout d'abord, les termes peuvent être :

- des variables, qui représentent les objets inconnus de l'univers, et qui sont identifiées par des chaînes alpha-numériques commençant par une majuscule;
- des termes élémentaires (ou termes atomiques) qui représentent les objets simples connus de l'univers. On en distingue trois sortes : les nombres, les atomes (chaîne alpha-numérique commençant par une minuscule), les chaînes de caractères (entre guillemets);
- des termes composés qui représentent les objets structurés de l'univers. Syntaxiquement, un terme composé est de la forme `foncteur(t1, ..., tn)` où `foncteur` est une chaîne alpha-numérique commençant par une minuscule, et les `ti` sont des termes. Le nombre d'arguments `n` est appelé arité du terme. Par exemple, `adresse(18,"rue des lilas",Ville)` est un terme composé du foncteur `adresse` d'arité 3, dont les deux premiers arguments sont les termes élémentaires `18` et `rue des lilas` et le troisième argument est la variable `Ville`.

Les atomes logiques expriment des relations entre des termes. Un atome se présente sous la forme `expression(t1, t2, ..., tn)` où `expression` est une chaîne alphanumérique commençant par une minuscule, et les `ti` sont des termes.

Une clause est une affirmation inconditionnelle (appelée "fait") ou conditionnelle (appelée "règle"). Elle se termine par un point. Un fait est de la forme `A.` où `A` est un atome logique. Il signifie que la relation définie par `A` est vraie. Par exemple, le fait `chien(pluto, mickey).` indique que la relation "pluto est le chien de mickey" est vraie. Une règle est de la forme `P :- P1, P2, ..., Pn.` où `P` et les `Pi` sont des atomes logiques. Une telle règle signifie que la relation `P` est vraie si toutes les relations `Pi` sont vraies.

Exécuter un programme Prolog consiste à poser une question à un interpréteur. Une question est une suite d'atomes logiques séparés par des virgules. La réponse de Prolog est `yes` si la question est une conséquence logique du programme, ou `no` si la question n'est pas une conséquence logique du programme. Une question peut

⁵Prolog est d'ailleurs un acronyme de PROgrammation LOGique.

comporter des variables. La réponse de Prolog est alors l'ensemble des valeurs des variables pour lesquelles la question est une conséquence logique du programme.

Le choix de ce langage est influencé par plusieurs critères. Le premier découle de la nature même de notre processus de supervision. En effet, l'objectif est de répondre à une "question" : "quelle est le rendu le plus adéquat pour une tâche donnée?". La réponse à cette question doit être élaborée à partir de "faits" (les "connaissances") et de "relations" connues entre des informations et des modalités. Le processus correspond donc totalement au fonctionnement d'une programme Prolog. Le second critère est l'évolutivité à moyen terme de notre superviseur : dans l'optique future d'intégrer de l'apprentissage artificiel pour la base de connaissances, Prolog est un langage approprié. Enfin, deux critères liés à l'implémentation concrète viennent s'ajouter. Le premier est la contrainte de notre architecture de faire communiquer une application de RV généralement codée en langage objet (C++ ou Java) et le superviseur. Or plusieurs librairies Prolog permettent cette communication entre langages différents et sont disponibles gratuitement⁶. Le second est la nécessité que le superviseur soit relativement simple à faire évoluer dans le futur, par l'ajout de connaissances et de règles de comportement. Le langage Prolog a cette qualité d'accessibilité.

La base de connaissances, la base de contexte ainsi que les informations qui accompagneront la requête seront définies par des faits. Les règles de comportement seront définies par des règles. La requête sera la question à l'interpréteur. Nous allons détailler maintenant l'implémentation de tous ces éléments.

4.3.3 Implémentation du modèle des interactions

Nous avons vu que les différents acteurs de la supervision, et en particulier l'application et le superviseur, devaient communiquer des informations sur les interactions à l'aide d'un modèle commun. L'implémentation du modèle des interactions permet de représenter les utilisateurs, les tâches et les rendus en cause.

4.3.3.1 Utilisateur

Un utilisateur est caractérisé par un identifiant, une liste de ses capacités sensori-motrices (correspondant aux trois canaux visuel, audio et haptique) et un niveau d'expertise. La liste des capacités (et donc les handicaps) va potentiellement restreindre dès le départ les canaux (et donc les modalités) utilisables par le système. Le niveau d'expertise sera utilisé au terme de la décision des modalités de rendu,

⁶Nous avons choisi la librairie Amzi! Prolog Logic Server : www.amzi.com

notamment pour décider si la redondance est nécessaire ou non. Le prédicat est de la forme :

```
user(idUser, [canal1, ..., canalN], niveau).
```

Ces trois composants sont fixés au lancement de l'application et ne sont pas supposés changer pendant son déroulement. Un changement d'opérateur demandera donc la création d'un nouvel utilisateur dans la base.

4.3.3.2 Tâche

Nous représentons la tâche de l'utilisateur par un ensemble de faits. Le premier est obligatoire et sert à créer un identifiant unique qui relie les faits entre eux. Les autres faits définissent le type de la tâche (facultatif, plusieurs possibles), la donnée (obligatoire, unique) et les paramètres (facultatif, plusieurs possibles).

Une tâche est d'abord définie par un identifiant unique et deux attributs, l'un indiquant par un booléen si la tâche est liée à l'utilisateur ou non. Par exemple, une tâche de manipulation d'un objet par l'utilisateur sera considérée comme liée, tandis qu'une tâche d'observation d'une variable qui fluctue indépendamment de l'utilisateur sera non liée. L'autre attribut indique par un entier entre 1 et 3 le niveau d'importance de la tâche. Le fait est de la forme :

```
inter(idTache, linkUser, niveau).
```

Types de tâche Les types de tâches modélisables sont ceux identifiés lors de l'analyse des applications. Il y en a donc quatre : Navigation, Observation, Sélection et Manipulation. En Prolog, un type de tâche est représenté par un fait de la forme :

```
task(idTache, typeTache).
```

où `typeTache` peut prendre les trois valeurs possibles : `nav`, `obs`, `sel` et `man`.

Nous l'avons vu lors de la modélisation, chaque type de tâche possède des modalités de rendu favorites. En principe, celles-ci devront être déterminées à partir d'évaluations ergonomiques. On peut néanmoins supposer que les tâches de manipulation profitent plus du rendu haptique, car elles font naturellement appel à des actions manuelles le plus souvent locales et qui requièrent de la précision et/ou de l'assistance de la part du système. De même, dans une tâche d'observation, le rendu visuel sera privilégié, surtout lorsque l'on voudra connaître des informations précises, comme des valeurs numériques.

Paramètres Les paramètres servent à préciser l'interaction réalisée par l'utilisateur, à décrire la manière dont la tâche est réalisée. Ils vont jouer une part très importante lors de la décision du rendu.

Les paramètres sont identifiés dans le moteur par un type et sont associés à une interaction.

`param(idTache, typeParam)`.

Les paramètres ont été définis de façon détaillée dans la section 4.2.2.2 (p.99). Pour l'implémentation et les premières validations, nous avons choisi de concentrer nos recherches sur un nombre restreint de paramètres. Ceux que nous avons intégrés dans le moteur de supervision sont les suivants : **objectif**, **contrainte** liée à la tâche, **contrainte physique** (réaliste), **global**, **local**, **quantitatif**, **état**.

Concrètement, lorsque le paramètre **objectif** est présent, le système doit aider l'utilisateur à réaliser un objectif, par exemple en guidant son outil de manipulation vers une cible. Les modalités privilégiées pour ce paramètre sont les guides, qu'ils soient haptiques (attraction, blocage, champ de force, mur invisible) ou visuels (lignes à suivre, déplacement automatique du pointeur).

Pour la **contrainte**, la tâche de l'utilisateur est soumise à des limitations, soit liées à la tâche elle-même (ex : déplacement d'une pièce entravée par d'autres pièces) soit aux limites que s'impose l'utilisateur (ex : déplacements limités pour éviter des erreurs, actions limitées par un formateur pour entraîner l'apprenant sur d'autres actions).

physique indique le respect des contraintes physiques de la scène, notamment la non interpénétration des objets. Ce paramètre sera vraisemblablement fixé pour toute une période de l'activité. Deux moyens sont possibles pour respecter la physique : soit contraindre les actions de l'utilisateur par l'intermédiaire de l'haptique, soit contraindre les déplacements des objets de la scène (visuellement notamment).

Nous avons considéré qu'une tâche locale allait se traduire par une augmentation du zoom visuel pour se rapprocher des données, éventuellement un changement d'échelle pour les déplacements des interfaces d'entrée, et une prédisposition pour les modalités haptiques. Une tâche globale se traduira par une diminution du zoom visuel et une favorisation des modalités visuelles et sonores par rapport aux modalités haptiques. Inversement, on peut également prévoir qu'une commande de changement de zoom par l'utilisateur ferait passer le système d'un mode **local** à un mode **global** et réciproquement.

Donnée La donnée concernée par une tâche est représentée par un type et la liste éventuelle des objets de l'application qui sont en jeu. Ce deuxième attribut pourrait servir à écarter certains rendus de la décision si les objets de l'EV ne sont pas techniquement compatibles. Le fait est de la forme :

```
data(idTache, typeData, [objet1, ..., objetN]).
```

De même que pour les paramètres, notre première implémentation n'a concerné qu'un nombre restreint de données. Celles que nous avons intégrées dans le moteur de supervision sont les suivantes : **topologie**, **contact** (événement dans la scène), **variable**, **champ** de variable et **proxy** (effecteur virtuel lié à l'utilisateur).

4.3.3.3 Canaux, médias et modalités

Là encore, il s'agit d'implémenter le modèle du rendu que nous avons élaboré précédemment. Les canaux sensori-moteurs et les médias sont confondus dans le modèle. Ils sont au nombre de trois : **visuel**, **audio** et **haptique**. Ces termes sont utilisés dans la liste des capacités de l'utilisateur. Pour chacun d'entre eux, nous avons identifié un certain nombre de modalités (section 4.2.3). En Prolog, toutes les modalités disponibles dans l'application sont représentées par des faits sous la forme :

```
mod(modalité, canal).
```

Notre implémentation actuelle comporte :

- 7 modalités visuelles : **topologie**, **texte**, **widget**, **couleur**, **guide**, **contrainte** et **physique** ;
- 4 modalités auditives : **icone**, **mapping** (correspond à l'auralisation), **vocal** et **synthèse** ;
- 4 modalités haptiques : **mapping** (correspond à l'haptisation), **guide**, **contrainte** et **physique**.

4.3.4 Processus de décision

Le principe du processus de décision (figure 4.10) est d'appliquer des règles de comportement sur des connaissances, afin de répondre à la question : "quel est le rendu multimodal à effectuer pour une tâche donnée ?". Plus précisément, les règles élaborent des scores pour chaque modalité disponible, en fonction de la tâche, des connaissances et du contexte et aboutissent à un résultat sous la forme d'une liste de modalités à rendre. Nous détaillons ce processus en présentant les clauses que nous avons implémentées.

4.3.4.1 Requête

Lorsque l'utilisateur ou l'application demande à effectuer une interaction, le gestionnaire d'interactions commence par rassembler les différents éléments nécessaires à la requête (tâche, donnée, paramètres), puis crée une instance de superviseur.

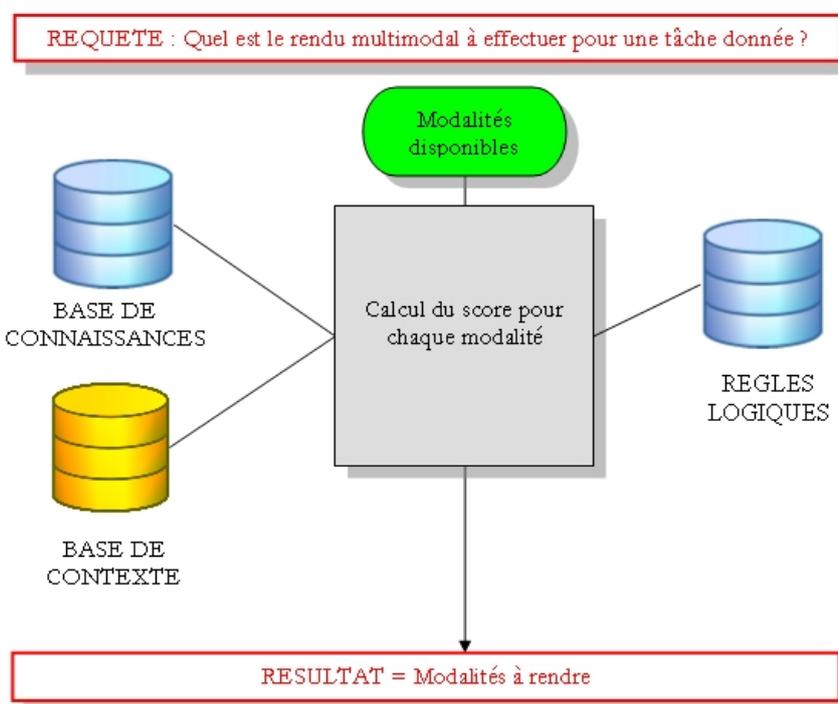


FIG. 4.10 – Principe du processus de décision.

Muni de tous ces éléments, celui-ci exécute la requête auprès du moteur de règles. La requête est de la forme :

?- supervision(idUser, idTache, Resultat).

où idUser et idTache sont des termes atomiques (connus) et Resultat une variable.

La réponse obtenue est la valeur de Resultat qui respecte l'ensemble des faits connus par le superviseur (base de connaissances et contexte). Elle prend la forme d'une liste, contenant des termes élémentaires qui se rapportent à des modalités de rendu :

Resultat = [mod1, ..., modN]

Si l'utilisateur ou le concepteur ont un rendu préférentiel pour la tâche demandée, cette information est transmise au superviseur avant la requête par un fait de la forme :

appli(idTache, rendu).

où idTache et rendu sont des termes atomiques (connus), rendu étant une liste de modalités.

Le résultat de la requête est communiqué au gestionnaire d'interactions qui peut alors commander sa génération et sa présentation concrète à l'aide des médias.

4.3.4.2 Base de connaissances

Le composant central de la base de connaissances, qui renferme la sémantique humaine du rendu multimodal, est un ensemble d'atomes qui définissent des relations entre des éléments de la tâche à effectuer et des modalités favorites, préconisées par le cahier des charges. Ces faits associent à un terme (désignant un type d'élément) une liste de modalités pondérées par un score. Le score final de chaque modalité disponible sera influencé par le score de ces modalités dans chacun des faits élémentaires.

Une interaction étant composée d'une tâche, d'une donnée et de paramètres, il y a donc une famille de prédicats pour chacun de ces éléments. Tout d'abord, nous avons considéré qu'un type de tâche donné pouvait favoriser un ou plusieurs canaux sensori-moteurs particuliers, donc toutes les modalités associées à ces canaux. Les faits que nous avons choisis sont :

```
task_chanfav(obs, [(visual,1), (audio,1)]).
task_chanfav(nav, [(visual,1)]).
task_chanfav(man, [(visual,1), (haptic,1)]).
task_chanfav(sel, [(visual,1), (audio,1), (haptic,1)]).
```

Comme nous l'avons précisé lors de l'analyse, nous avons jugé que toutes les tâches étaient propices à un rendu visuel, tandis que l'observation favorisait également les modalités audio, et que les modalités haptiques étaient préconisées pour des tâches de manipulation et de sélection. Les scores présentés ici sont tous égaux à 1, afin que les canaux qui composent les listes aient tous le même poids. Nous verrons par la suite comment les évaluations futures vont nous permettre d'affiner ces scores et d'éventuellement hiérarchiser les canaux.

Viennent ensuite les listes de modalités favorites selon le type des données et des paramètres. Les faits se présentent sous la forme :

```
data_modfav(typeData, [(mod1, p1), ..., (modN, pN)]).
```

et

```
param_modfav(typeParam, [(mod1, p1), ..., (modN, pN)]).
```

Là encore, les modalités sont initialisées avec des pondérations égales à 1.

Ensuite, nous avons intégré des clauses qui permettent d'indiquer au superviseur les éventuels choix de rendu des utilisateurs en fonction des types de données rencontrés. A chaque utilisateur peuvent donc être associés des faits de la forme :

```
user_modfav(idUser, typeData, [(mod1, p1), ..., (modN, pN)]).
```

L'ensemble de ces faits permet de définir des profils utilisateurs, modifiables et personnalisables en fonction des applications.

Cette base de connaissances statiques n'est pas la seule à servir aux règles pour déterminer le résultat. Il convient également d'observer les éléments du contexte qui peuvent intervenir pour modifier le score final des modalités.

4.3.4.3 Contexte

Les critères essentiels du contexte que nous prenons en compte sont les canaux disponibles (nous les confondons avec les médias dans l'implémentation) et les rendus en cours. La disponibilité d'un canal est tenue à jour dynamiquement par un fait, qui contient également sa charge, i.e. le nombre de modalités qui sont en cours de rendu sur ce canal à un instant t :

```
channel(canal, charge).
```

La charge est modifiée à chaque nouveau rendu concret ou en fin d'interaction par l'observateur de contexte. Pour chacun des canaux, nous avons défini arbitrairement une charge maximale à ne pas dépasser. Cette limite a été fixée arbitrairement et devra être soumise à évaluation :

```
chargeMax(visual, 6).
chargeMax(audio, 3).
chargeMax(haptic, 2).
```

Ceci signifie donc qu'il peut y avoir à tout instant de nombreuses modalités visuelles (mais limitées volontairement ici afin d'activer plus rapidement les autres canaux lors des tests), mais peu de modalités auditives et seulement deux modalités haptiques (afin de pouvoir ressentir au moins un rendu épistémique et une autre fonction, par exemple de type guide).

Cette mise à jour régulière des charges s'accompagne de l'enregistrement dans la base prolog de toutes les interactions courantes (i.e. tâche avec leur rendu correspondant, sous la forme d'une liste de modalités) et des interactions passées :

```
current(idUser, idTache, [mod1, ..., modN]).
done(idUser, idTache, [mod1, ..., modN]).
```

4.3.4.4 Elaboration du score

A partir des connaissances statiques et dynamiques que nous venons de décrire, les règles de décision opèrent une succession de filtres (figure 4.11). Partant de

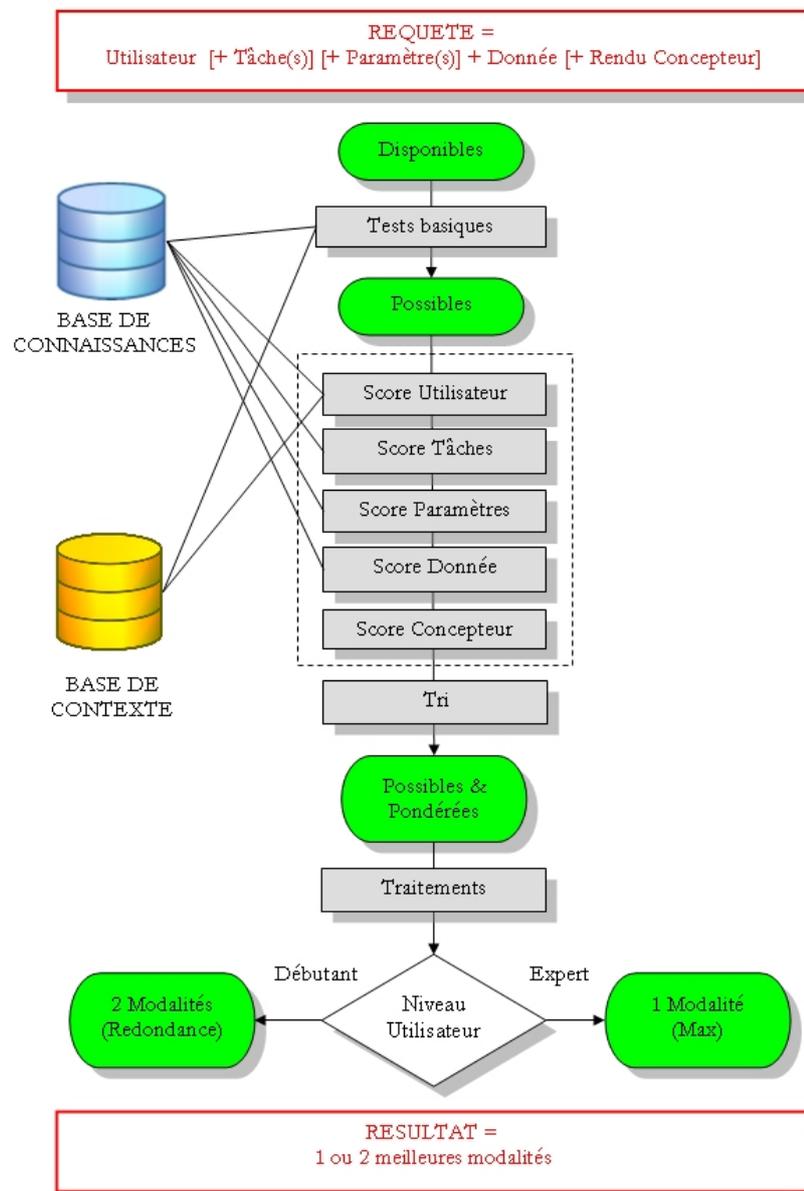


FIG. 4.11 – Processus de décision du superviseur détaillé. L'ordre des filtres de pondération n'est pas significatif.

l'ensemble des modalités disponibles (i.e. programmées) dans l'application, ces filtres enlèvent ou ajoutent des modalités, ou encore incrémentent ou décrémentent leurs pondérations, afin d'aboutir à une liste hiérarchique finale de modalités.

Le premier filtre est simple, puisqu'il s'agit d'éliminer toutes les modalités im-

possibles parmi celles qui sont disponibles dans l'application. Ce filtrage dépend tout d'abord des handicaps de l'utilisateur : par exemple s'il est sourd, la règle élimine les modalités audio. Il dépend ensuite de l'état des médias. Ceux-ci peuvent en effet d'une part être éteints, ou défectueux au moment de l'interaction, et d'autre part ils peuvent être en surcharge, c'est-à-dire qu'un grand nombre de modalités les occupe déjà. Si le nombre maximal admis de charge d'un média est atteint, les modalités selon ce média seront supprimées de la liste.

Une fois que la liste des modalités disponibles est restreinte aux seules modalités possibles ou utiles, l'objectif des filtres suivants est d'aboutir à la ou les meilleures modalités pour l'interaction demandée. Pour cela, les filtres vont faire évoluer les scores de chaque modalité. Les principaux éléments qui vont influencer ce score sont les prédicats de listes favorites définis précédemment dans la base de connaissances. Il existe une règle de filtrage pour chaque élément : tâche, donnée, paramètre, utilisateur. Ces règles fonctionnent toutes selon le même principe : si une modalité est dans la liste des favorites pour l'agent concerné, elle est créditée du poids donné dans le prédicat. Interviennent également les éventuelles préférences de l'utilisateur, qui a pu indiquer ses propres modalités favorites pour certains éléments, et la possible proposition de l'application lors de la requête : le concepteur a pu associer des modalités particulières aux interactions de son programme. Puis une troisième série de filtres va transformer certaines modalités en fonction du contexte. La charge courante des canaux intervient (une seconde fois après le filtrage binaire des tests basiques) de façon à défavoriser les modalités qui vont trop approcher les canaux de leur limite de charge.

Une fois tous ces filtrages effectués, le moteur de décision doit choisir parmi une liste de modalités pondérées. L'objectif est d'obtenir la ou les modalités de meilleur(s) score(s), selon que l'on désire une assignation ou une redondance. Pour ce faire, la liste finale est triée par score croissant. Ensuite, le choix dépend du niveau d'expertise de l'utilisateur. En effet, il nous a semblé utile d'appliquer la redondance des modalités à des utilisateurs débutants ou handicapés. C'est pourquoi, si le niveau de l'utilisateur est égal à 3, seule la meilleure modalité est retenue. Dans le cas contraire, toutes les modalités de score maximum sont choisies.

4.3.4.5 Résultat

La ou les modalités obtenues à l'issue du processus de décision correspondent à un rendu concret codé dans l'application. Le superviseur se veut générique et indépendant des plate-formes matérielles. Il n'est donc pas en charge du rendu concret, et ne choisit pas le média de transmission des modalités choisies : l'ensemble des acteurs du rendu concret est géré par l'application. Le gestionnaire d'application va recevoir

le résultat et l'interpréter en mettant à jour les paramètres de rendu de l'application. Il doit également faire les mises à jour nécessaires et commander à l'observateur de contexte de les répercuter dans les bases Prolog. Puis, lors des boucles de rendu visuel, sonore et haptique, les paramètres de rendu seront concrétisés sur les médias associés.

4.3.5 Validation sur un exemple

Pour illustrer le mécanisme de la supervision, nous détaillons le processus pour l'exemple de Mécanique des Fluides n°5 présenté dans la section 4.2.4 (figure 4.5) : "l'utilisateur désire connaître la valeur exacte de la température en une position d'une simulation de liquide". Dans un souci de clarté de la démarche, nous utiliserons une base de connaissances simplifiée. La construction détaillée de cette base fera l'objet du chapitre suivant.

L'interaction est déclenchée par une commande explicite de l'utilisateur, par exemple un ordre vocal du type "température, ici" accompagné de la désignation gestuelle de l'emplacement. La modélisation de la tâche est établie par le gestionnaire d'interactions à partir de cette commande :

```
Tâche = observation
Donnée = variable
Paramètre = quantitatif
```

Les informations que doit connaître le superviseur par l'intermédiaire de l'observateur / interpréteur de contexte sont les médias disponibles dans le dispositif accompagnés de leur charge, les modalités programmées dans l'application ainsi que le profil de l'utilisateur (handicap et expertise). Les deux dernières informations ont été fixées au démarrage de l'application et peuvent être considérées comme statiques dans la base de contexte Prolog. Ici nous considérerons que tous les médias et toutes les modalités sont disponibles, que la tâche est la première demandée (donc que les charges des médias sont nulles) et que l'utilisateur est un expert sans handicap.

```
user(expert, [visual, audio, haptic], 3).
```

```
channel(visual, 0).
channel(audio, 0).
channel(haptic, 0).
```

```
mod(vcolor, visual).
mod(vtext, visual).
...
mod(hguide, haptic).
```

```
...
mod(earcon, audio).
...
```

La requête est effectuée par l'application. Elle est composée de l'ajout des clauses de modélisation, puis de la demande de supervision proprement dite :

```
inter(#, true, 3).
task(#, obs).
data(#, variable, [temperature]).
param(#, quanti).
? supervision(expert, #, Resultat).
```

où # est un identifiant généré automatiquement par le gestionnaire d'interactions.

La base de connaissances qui va être utilisée dans ce cas est composée des listes de modalités favorites pour les éléments de la modélisation :

```
task_medfav(obs, [(visual,1), (audio,1)]).
data_modfav(variable, [(amapping,1), (hmapping,1), (vtext,1)]).
param_modfav(quanti, [(vtext,1), (avocal,1)]).
```

Puis le superviseur enclenche le processus de décision pour aboutir au résultat. La liste des modalités disponibles dans l'application (ici, toutes) n'est pas modifiée par le premier filtre, car l'utilisateur n'a pas de handicap, les médias sont tous disponibles et non chargés. Ensuite, les règles modifient les pondérations de chaque modalité de la liste en fonction des différents éléments (nous considérons ici que ni l'utilisateur, ni le concepteur n'ont précisé de préférence pour ce type de tâche) :

- Score Utilisateur : ne modifie rien ;
- Score Tâches : incrémente de 1 les scores des modalités visuelles et sonores ;
- Score Paramètres : incrémente de 1 les scores de `vtext` et `avocal` ;
- Score Donnée : incrémente de 1 les scores de `vtext`, `amapping` et `hmapping` ;
- Score Application : ne modifie rien ;

Après tri par ordre décroissant, la liste des modalités pondérées est :

```
[(vtext,3), (avocal,2), (amapping,2), (hmapping,1),
(aicon,1), (vtopo,1), (vconstraint,1), (vgoal,1),
(vphys,1), (vicon,1), (vchange,1), (widget,1),
(vreverse,1), (vbox,1), (vcolor,1), (hattract,0),
(hphys,0), (hblock,0), (hguide,0)]
```

L'utilisateur étant un expert, le superviseur décide de ne conserver que la modalité de meilleur score. La liste résultat est donc :

```
Résultat = [vtext]
```

Le superviseur communique ensuite ce résultat au gestionnaire d'interactions, via l'interpréteur. Le gestionnaire commande le rendu textuel au moteur visuel pour la variable concernée (ici la température). Le moteur affiche la valeur numérique de la variable.

Pour cette tâche élémentaire, les éléments qui ont influencé le rendu sont les composants de la modalisation de la tâche. Nous avons en effet choisi un contexte non restrictif. Dans un cas réel, nous pouvons supposer que les limitations contextuelles dépendront des charges des différents médias dues aux tâches précédentes et des choix préférentiels de l'utilisateur et du concepteur. La présence d'un utilisateur aveugle et l'absence de média visuel sont peu probable dans l'application visée. La limite maximale de charge visuelle étant relativement élevée, le rendu textuel de la tâche que nous venons d'étudier sera donc le plus fréquent (avec la base de connaissances considérée).

Après avoir détaillé le coeur de la supervision, nous abordons dans la suite de ce chapitre d'autres fonctionnalités qui nous paraissent pertinentes pour effectuer du rendu multimodal.

4.4 Extension de la supervision

Sur la base du processus de supervision décrit précédemment, nous avons développé des fonctionnalités supplémentaires, afin de gérer tant la temporalité que la spatialité du rendu. Par exemple, le processus précédent ne permet pas de paramétrer la date de déclenchement, la durée, la position ou l'orientation dans l'espace du rendu : la date de déclenchement était la date de réception du résultat de la requête ; sa durée et sa position étaient celles de l'interaction. Il est important d'intégrer ces deux dimensions évoquées dans le modèle des interactions. De plus, le contexte que nous avons observé peut être étendu pour prendre en compte des situations plus diverses. Enfin, nous avons identifié des cas particuliers qui nécessitent des comportements de supervision adaptés, et nous abordons le problème des retraitements des interactions.

4.4.1 Gestion de l'espace

La gestion de l'espace est différente selon le canal en charge de la perception de la modalité. Nous avons identifié trois spatialisations possibles pour les modalités visuelles, suivant le repère par rapport auquel elles sont positionnées :

1. Repère de la scène : la position de la modalité dépend de l'environnement virtuel, par exemple de la position des objets ou des informations à rendre.

Ainsi, on peut spatialiser un earcon pour qu'il indique la position exacte d'une collision entre deux pièces.

2. Repère du dispositif : la position de la modalité dépend de l'environnement réel, des médias utilisés. On peut par exemple afficher des informations textuelles dans un coin de l'écran de projection. Bien que pouvant perturber l'immersion, ce genre d'affichage peut être requis si le canal visuel est déjà chargé, mais que l'information doit perdurer (e.g. message d'alerte à péréniser).
3. Repère de l'utilisateur : la position de la modalité dépend de celle de l'opérateur humain, de son focus visuel et de la position de sa main. Par exemple, il est possible d'afficher une valeur numérique de vitesse ou encore un message d'alerte dans son champ de vision plutôt qu'en une position donnée de la scène ou de l'écran. Cela permet de maximiser le degré d'information, mais au prix d'une surcharge cognitive.

La figure 4.12 présente les trois cas possibles pour un exemple simple de CAO. Ici, l'utilisateur veut connaître la hauteur d'un cylindre rouge. Si le superviseur valide l'affichage de cette cote par la modalité textuelle, celle-ci peut être présentée : sur la pièce concernée, sur l'écran ou dans le champ de vision de l'utilisateur. De plus, selon ses trois solutions, les possibilités sont multiples. Si l'affichage sur la pièce respecte la norme CAO, il aura lieu le long de la hauteur du cylindre. Sur l'écran, il peut y avoir une ou plusieurs zones 2d réservées à ce type d'affichage. Ou encore cela peut être le coin ou le côté le plus proche de l'objet, etc. De même, dans le champ de vision, il reste encore des incertitudes. Comme pour le choix de la modalité, le choix de cet attribut spatial va dépendre de l'état du rendu (charges visuelles dans les différentes positions possibles), de l'état de la scène (positions respectives des objets, des dispositifs et de l'utilisateur) et de l'interaction courante de l'utilisateur. Ainsi, dans l'exemple, le rendu sera différent si l'utilisateur est en train de manipuler le cube bleu en fonction de la cote observée, ou bien s'il veut conserver celle-ci pour une tâche future.

Un autre exemple un peu plus complexe est la comparaison de vitesses en deux positions d'une simulation d'écoulement (figure 4.13). L'utilisateur expert peut tout d'abord effectuer une première tâche d'observation de la vitesse V_1 à la position P_1 . Celle-ci va être affichée en P_1 . Ensuite, une seconde commande d'observation va entraîner l'affichage de la vitesse V_2 en P_2 . Mais s'il veut la comparer à la précédente alors qu'elle n'est plus dans son champ de vision, plusieurs choix sont possibles : le rappel de la vitesse V_1 dans le nouveau champ de vision, l'affichage de la différence entre V_1 et V_2 , l'affichage des deux vitesses dans une zone dédiée, etc.

Concernant les modalités audio, seules deux types de spatialisation sont possibles : l'une dans le repère de la scène, l'autre dans le repère du dispositif. En effet, le concept de champ auditif n'est pas applicable puisque la perception sonore est omnidirectionnelle.

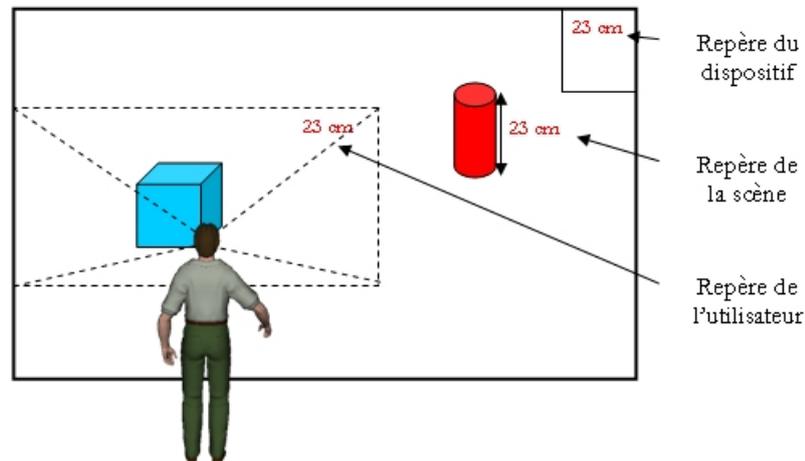


FIG. 4.12 – Exemple de choix de présentation spatiale pour une tâche d'observation de cote en CAO.

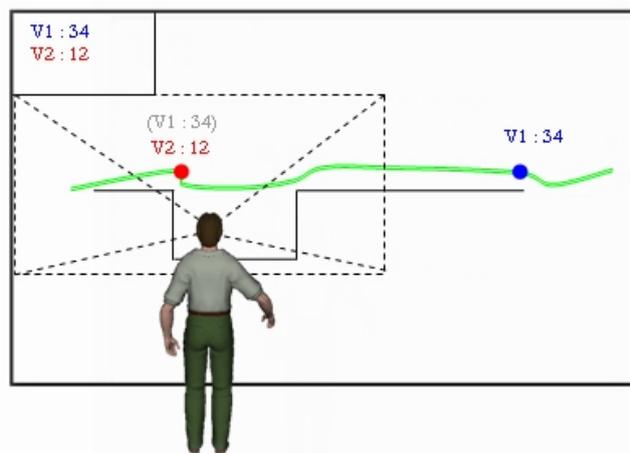


FIG. 4.13 – Exemple de choix de présentation spatiale pour une tâche de comparaison de vitesses en Mécanique des Fluides.

Enfin, les modalités haptiques, ne sont pour des raisons physiques que perceptible dans le repère de l'utilisateur : les informations sont communiquées aux parties du corps en contact avec l'interface concernée.

Afin que le moteur de décision gère ces différents cas possibles pour chaque rendu, nous avons choisi d'ajouter cet attribut spatial à chaque modalité manipulée par le moteur de décision. Cet attribut prend la forme de trois constantes : **scene**, **real** et **human**. Il apparaît donc dans les listes de modalités favorites de la base de connaissances. Des règles sont également ajoutées pour pondérer cet attribut si la

modalité apparaît plusieurs fois au cours du processus de score. Lors de la réception du résultat, l'application concrétise cet attribut en fonction des données concernées.

4.4.2 Gestion du temps

Jusqu'ici, nous avons observé des cas dans lesquels, lorsque la liste des modalités était décidée, tout était rendu au même moment, et pour une durée égale à celle de la tâche ou celle de la modalité choisie. La gestion temporelle du rendu consiste à choisir la date de départ, la durée et la répétition de la modalité. Lorsque plusieurs modalités sont combinées pour une même tâche, ces attributs temporels permettent de gérer la synchronisation des rendus, que ce soit au déclenchement, à la terminaison, pour compenser des latences ou faire des redondances temporelles (figure 4.14).

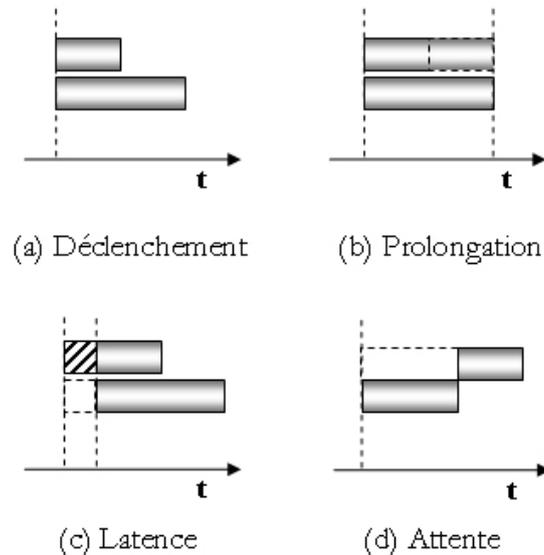


FIG. 4.14 – Exemples de synchronisation temporelle de deux modalités pour une tâche. (a) Synchronisation du déclenchement. (b) Synchronisation de déclenchement et de terminaison par prolongation d'une modalité. (c) Retard forcé d'une modalité pour synchronisation avec la latence d'une autre modalité. (d) Retard forcé d'une modalité pour synchronisation avec la terminaison d'une autre modalité.

Nous avons donc intégré cette dimension dans notre modèle et dans notre moteur de décision, en ajoutant à chaque modalité (et par extension à chaque liste de modalités) une date de départ et une durée de rendu. Comme pour les listes de modalités favorites, le concepteur va devoir indiquer pour chaque modalité les paramètres favoris. Par défaut la date de début de rendu va être la date de réception du résultat de la requête, et sa durée va être égale à la durée de la tâche.

Par la suite, plusieurs effets sont permis. Le premier est l'application de règles de combinaison particulières, comme le fait de demander une redondance temporelle d'information. Celle-ci pourrait par exemple intervenir si l'utilisateur ne réagit pas au premier rendu. Le second effet est la gestion en amont des latences, quand celles-ci sont connues ou prévues, comme par exemple lorsque le concepteur de l'application sait que le calcul du rendu haptique est plus long que le rendu visuel. Les paramètres vont permettre de retarder le rendu le plus rapide pour que les deux soient synchronisés.

Une toute autre fonctionnalité temporelle à prévoir est la mémorisation d'une interaction (tâche et rendu) ainsi que de son contexte, lorsque cette tâche va devoir être effectuée plusieurs fois par la suite. Par exemple, si l'utilisateur prévoit de réaliser plusieurs extrusions de solides à la suite, ou si le superviseur détecte plusieurs requêtes identiques successives, il est intéressant de ne conserver le premier résultat et d'économiser les ressources de décision. Ainsi, lorsque la tâche se présente à nouveau, le superviseur la compare à celles déjà supervisées, vérifie si le contexte est identique, et réactualise le précédent résultat. Néanmoins, en cas de changement important du contexte, il faudra réitérer la décision.

4.4.3 Extension du contexte

Une première extension simple du contexte concerne les canaux sensoriel. Jusqu'ici, nous nous limitons à prendre en compte les charges sonores, visuelles et haptiques dues aux interactions avec l'EV en cours de rendu. Cependant, il est également nécessaire de considérer les charges sensorielles causées par l'environnement réel, hors interactions avec l'EV. Par exemple, l'activité de l'utilisateur peut avoir lieu dans un environnement bruyant, qui complique la compréhension des informations audio. Pour que le moteur inclut cet élément dans la décision, il suffit d'augmenter artificiellement le nombre de charges du canal audio (on augmente ainsi le "bruit"), pour que toutes les tâches à effectuer aient tendance à être rendues sur les autres canaux. Par contre, si l'utilisateur se met à parler (par exemple pour expliquer son activité à un auditoire), il faut plutôt concevoir une règle de retraitement qui diminue temporairement le volume des rendus audio courants, puis les remet à jour lorsque l'utilisateur a fini ses explications. Le premier comportement est également valable pour les rendus visuels, en cas de forte luminosité par exemple.

Un deuxième élément du contexte qui a son importance est l'attention spatiale de l'utilisateur, c'est-à-dire les zones de l'espace où sont concentrées ses perceptions. Nous ne retenons ici que l'attention visuelle (champ visuel) et l'attention manuelle (champ d'action de la main) car elles sont toutes deux directionnelles. Le concept d'attention spatiale auditive n'a pas vraiment de sens ici car la perception auditive est omnidirectionnelle et toujours en action. En général, lors d'une manipulation, l'utilisateur regarde ce qu'il fait, c'est-à-dire que l'attention visuelle est colocalisée

à l'attention manuelle. Une situation pouvant mener à un rendu particulier serait la délocalisation entre les deux. Par exemple un chercheur de Mécanique des Fluides pourrait se tourner vers son public lors de la manipulation précise et haptique d'un outil scientifique. Ou encore, l'opérateur peut vouloir manipuler un objet en ayant son attention visuelle sur une autre donnée. L'observateur de contexte devrait suivre cette colocalisation, et en informer le superviseur, qui pourrait alors décider dans le premier exemple de bloquer l'interface haptique pour que le chercheur ne perde pas sa position actuelle (et de la débloquent dès que les deux champs seraient à nouveau colocalisés), et dans le deuxième exemple d'ajuster le point de vue afin de pouvoir percevoir à la fois l'objet manipulé et la donnée observée.

4.4.4 Gestion des retraitements

Comme nous venons de le voir dans la section précédente, le moteur permet de créer des rendus qui durent pendant une certaine période. Cependant, ces rendus n'étaient jusqu'alors pas modifiés au cours du temps. Or, l'évolution du contexte des interactions peut entraîner qu'un rendu préalablement décidé par le superviseur ne soit plus adapté à son interaction. Par exemple, si l'utilisateur observe les variations d'une variable numérique par une variation fréquentielle sonore, et que l'environnement devient bruyant, alors le rendu n'est plus adapté. Le moteur doit donc prendre en compte l'évolution temporelle des rendus, ce qui se traduit par la possibilité de retraiter une interaction. Nous avons identifié plusieurs situations nécessitant un retraitement :

1. la commande explicite de l'utilisateur de réévaluer une interaction, ou encore le refus d'un rendu proposé ;
2. une caractéristique de l'interaction qui demande un réexamen périodique (associé à une minuterie) ;
3. modification d'un élément précis du contexte duquel dépend une interaction ;
4. l'apparition d'une nouvelle interaction de niveau supérieur à une interaction existante et qui réclame le même rendu : l'interaction moins importante laisse sa place mais doit perdurer.

Cette fonctionnalité est permise dans notre moteur grâce à l'enregistrement des interactions en cours et de leur rendu. A chaque cause de retraitement est associée une règle qui indique le comportement à suivre. C'est le gestionnaire d'interaction qui reconnaît ces causes et relance une requête semblable à celle d'un traitement normal.

La difficulté principale des retraitements est la gestion des transitions entre les modalités de rendu pre- et post-retraitement. En effet, les effets cognitifs du passage d'une modalité à une autre pour communiquer une même information n'ont pas été

beaucoup étudiés. Dans l'état actuel du superviseur, lorsque l'utilisateur n'est pas à l'origine du retraitement, il est averti explicitement du changement de rendu.

Il serait intéressant d'étudier la pertinence d'utiliser la redondance multimodale pour gérer le problème des transitions. Il faudrait pour cela créer des procédures de "retraitement fictif", dont le rôle serait d'effectuer un processus de décision de rendu, sans appliquer concrètement le résultat, mais en le comparant au rendu courant. Dans certains cas, par exemple si le score de la modalité courante décroît de manière perceptible au cours des retraitements fictifs, il serait possible de prévoir l'imminence d'un retraitement et d'un changement de rendu, et de préparer cette transition par l'apparition d'une redondance (modalité courante et probable prochaine modalité). Nous pourrions ainsi créer un "fondu multimodal", permettant à l'utilisateur d'être moins déstabilisé que lors d'une transition brutale entre rendus.

4.4.5 **Traitement de cas particuliers**

Nous avons vu précédemment que pour limiter la surcharge cognitive des canaux sensori-moteur, nous limitons le nombre de modalités simultanées sur chacun des canaux. Il existe cependant d'autres limites au choix d'une modalité que la charge maximale du canal : des limites techniques, des limites perceptives et des limites de compréhension. Les limites techniques concernent certaines modalités (notamment haptiques) dont le rendu sur un dispositif interdit tout autre rendu sur le même dispositif. C'est parfois le cas des contraintes haptiques, qui, en bloquant des degrés de liberté de l'interface empêchent l'activation d'autres guides ou rendus épistémiques. Les limites perceptives concernent les incompatibilités sensorielles entre certaines modalités, qui peuvent par exemple être mises en évidence dans la littérature sur les illusions ou sur la perception humaine. L'utilisation de ces modalités peut entraîner des artefacts sensoriels non désirés, qui compromettent l'information transmise. Enfin, les limites de compréhension sont proches des précédentes, mais sont plus liées à la difficulté pour l'utilisateur de distinguer des informations différentes transmises par des modalités qui se brouillent l'une l'autre. Par exemple, deux haptisations simultanées créent un risque de confusion d'informations (alors qu'une haptisation est compatible avec un rendu haptique épistémique, par exemple). Ou encore, un message vocal d'alerte doit pouvoir être perçu clairement, sans être masqué par d'autres modalités sonores. Les moyens mis en oeuvre pour gérer ces incompatibilités sont l'ajout de filtrages spécifiques et l'ajout de connaissances dans la base. Les connaissances sont des clauses particulières concernant des modalités (modalité sans doublon, modalité unique sur le canal) ou des couples de modalités (incompatible, risqué). Puis, à la fin de la procédure de pondération, des filtres peuvent diminuer le score des modalités sensibles, ou les supprimer, ou les remplacer.

D'autre part, certains rendus trop contraignants ne doivent pas être imposés à l'utilisateur mais seulement proposés, sous forme visuelle ou sonore, et accompagnés

d'un moyen d'accepter ou de refuser le rendu (e.g boîte de dialogue, commande vocale, bouton de l'interface de manipulation, etc.). Le superviseur doit donc indiquer à l'application lors de la requête d'une part le rendu proposé, d'autre part qu'il faut le soumettre au choix de l'utilisateur.

Enfin, il existe des interactions nécessitant des actions immédiates de la part du système. Lorsque de telles interactions sont reconnues (typiquement des alertes système), celles-ci font l'objet d'une requête et d'un traitement par le superviseur, en respect de la cohérence qui doit être maintenue dans l'application. Cependant, la décision du rendu est accélérée par des règles spécifiques, qui suppriment certains filtres. Par exemple, pour la communication d'une donnée d'alerte de niveau élevé, nous avons choisi d'appliquer automatiquement une redondance de modalités prédéfinies : `couleur`, `earcon` et `texte`.

4.4.6 Enregistrement pour l'évaluation

La première évaluation nécessaire est celle du superviseur. Il convient de vérifier la validité des décisions prises, qui sont issues directement de la base de connaissances et de la base de règles. De plus, nous avons spécifié que l'un des rôles du superviseur est de permettre à moyen terme l'évaluation du bien-fondé, de l'intérêt de la multimodalité en sortie. En particulier, il s'agit de savoir quelles sont les interactions effectuées par les utilisateurs, quels sont les rendus décidés par le superviseur (à partir des règles établies par les concepteurs) et si ces rendus correspondent aux attentes et aux besoins des utilisateurs.

L'enregistrement des interactions et de leurs rendus en cours est déjà réalisé pour des raisons de connaissance dynamique du contexte, pour des besoins de retraitement et des éventuels problèmes de transitions entre modalités. L'enregistrement à fin d'évaluation consiste à conserver dans un fichier ou une base de données : les interactions ayant fait l'objet d'une requête au superviseur, les décisions prises à la suite du processus de score et enfin le rendu concrètement effectué. Il faut ensuite vérifier l'adéquation de la décision avec les éventuelles modalités favorites de l'utilisateur (choix faits a priori) qui avaient été enregistrées dans la base de connaissances, ainsi que la réaction a posteriori de l'utilisateur au rendu.

4.5 Conclusion

Dans ce chapitre, nous avons présenté en détail notre conception d'un superviseur de rendu multimodal en RV. Pour cela, nous avons élaboré un modèle des interactions de RV, composé d'un modèle des tâches et d'un modèle du rendu multimodal. L'architecture du superviseur et de l'application manipule ces représentations par un processus de décision logique. Celui-ci utilise à la fois des connaissances statiques

sur la projection modalités-données ainsi que des connaissances dynamiques sur le contexte d'interaction. Nous avons implémenté le superviseur en Prolog et nous avons réalisé une première validation de cette architecture sur des exemples concrets tirés de l'analyse des applications. Le chapitre suivant prolonge ces évaluations en proposant auparavant une méthode de construction de la base de connaissances, et en présentant le simulateur d'application de RV que nous avons développé afin de vérifier l'opérationnalité logicielle du superviseur multimodal.

Chapitre 5

Evaluation, résultats et discussion

Résumé. *Dans ce chapitre, nous implémentons notre superviseur et nous réalisons un simulateur d'application de RV afin de tester le processus de supervision. Puis, nous élaborons une méthode de construction de la base de connaissances. Cela nous permet alors d'évaluer notre supervision multimodale sur des tâches concrètes issues des applications du chapitre 3, et notamment sur un scénario interactif complet de Bio-Informatique. Les résultats sont satisfaisants, tant au niveau de la fiabilité et de la qualité des rendus décidés, que de l'adaptabilité du superviseur à des contextes différents et de la rapidité de la décision, compatible avec des interactions en temps réel. Les résultats sont analysés, puis nous discutons de problèmes multimodaux spécifiques soulevés au cours des recherches.*

5.1 Introduction

L'objectif de ce chapitre est de présenter les évaluations que nous avons menées sur notre supervision multimodale et les résultats que nous avons obtenus.

Pour cela, nous effectuons d'abord des tests afin de valider notre implémentation du moteur de décision ainsi que l'architecture de supervision que nous avons conçue. Puis, nous expliquons la méthode que nous avons élaborée afin d'ajouter des connaissances au superviseur. Celui-ci a ensuite été évalué sur un scénario interactif complet. La dernière section nous permet de discuter les résultats obtenus et de mettre en évidence des problématiques relatives à la multimodalité en sortie.

5.2 Evaluation de l'architecture

Après avoir implémenté chacun des éléments de la supervision (section 4.3.3, p. 118), il est nécessaire de valider le fonctionnement interne du superviseur ainsi que l'architecture globale de la supervision. Pour cela, nous avons dans un premier

temps effectué des tests au sein de l'environnement Prolog. Puis, une fois que le fonctionnement du moteur de décision a été confirmé, nous l'avons englobé dans une boîte noire et nous avons testé ses différentes entrées-sorties depuis un simulateur d'applications de RV, réalisé en langage objet. Ces tests ont été effectués sur des interactions fictives et à l'aide d'une base de connaissances limitée. La construction de cette dernière fait l'objet de section 5.3, tandis que la section 5.4 présente des évaluations plus poussées.

5.2.1 Editeur Prolog

Tout d'abord, le coeur du moteur étant développé en Prolog, nous avons effectué des tests sous l'éditeur correspondant (voir Annexe B). Celui-ci dispose d'un interpréteur de commandes, qui permet notamment de tester séparément chaque règle, mais aussi de créer divers protocoles d'exécution ou encore de poser et de retirer dynamiquement des prédicats (par exemple pour créer des tâches à la volée, modifier en cours d'exécution des éléments du contexte comme la disponibilité et la charge des médias, etc.). Nous avons en particulier utilisé un protocole de supervision qui, à partir de la création d'un certain nombre de tâches, exécute le processus complet de décision de rendu, puis récupère le résultat, enregistre l'interaction courante, met à jour le contexte (les charges des médias) et enfin affiche les informations désirées. Ce protocole a été testé sur des tâches élémentaires et dans des contextes divers.

Un résultat notable est que, sous Prolog, l'ensemble du protocole de supervision pour ces tâches est réalisé en temps réel, de l'ordre de 5 ms entre le lancement de la requête et la fin des mises à jour nécessaires en fonction du résultat. En réalité, une part importante du temps est occupée par la présentation des informations dans l'interpréteur, affichage qui n'aura plus lieu dans le superviseur final.

5.2.2 Simulateur d'application

Après avoir testé le bon fonctionnement de la base de règles en Prolog (validité des règles, informations nécessaires à leur application, etc.) dans différents cas, nous avons voulu tester la robustesse de l'architecture globale de supervision multimodale. Dans cette optique, nous avons réalisé un simulateur d'application de RV. Celui-ci tient compte des contraintes inhérentes aux applications de RV et des données qu'elles manipulent, mais s'affranchit aussi de certaines implémentations requises dans la réalité, comme par exemple la programmation du gestionnaire d'interaction et la mise en place des moteurs de rendu de toutes les modalités prises en compte par le superviseur.

Ce simulateur se présente sous la forme d'une application fenêtrée (programmée

en C++/MFC¹), composée de différentes sections qui permettent de :

1. Choisir un profil utilisateur préchargé dans la base de connaissances ;
2. Créer artificiellement une tâche suivant notre modélisation et lancer une requête au superviseur ;
3. Afficher le résultat de la requête (les modalités décidées par le superviseur) sous forme textuelle ;
4. Choisir les médias disponibles et connaître leur charge courante ;
5. Connaître les interactions en cours et pouvoir les stopper ou les retraiter.

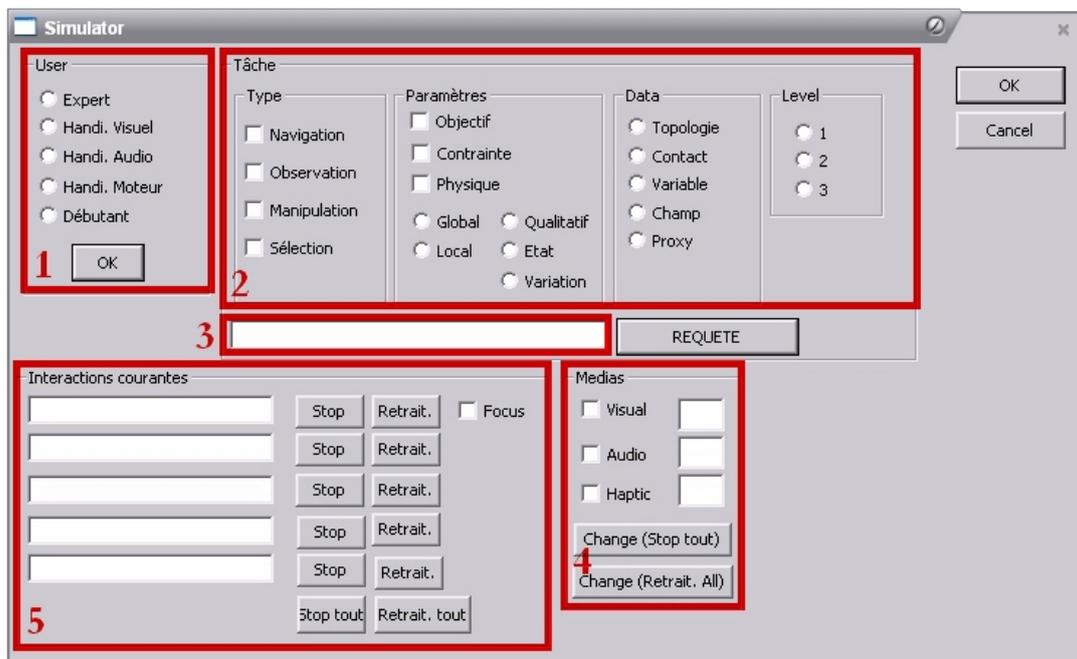


FIG. 5.1 – Simulateur de test de supervision multimodale.

D'autres fonctionnalités internes du simulateur ont été implémentées pour s'approcher d'une application réelle. En particulier un système de "threads" permet d'exécuter des requêtes de supervision en parallèle, lorsque des situations d'interactions intensives le demandent. Ces situations ne peuvent cependant pas être déclenchées par l'interface de la figure 5.1, car les décisions sont prises assez rapidement pour que deux requêtes ne se chevauchent pas. Une procédure automatisée permet de créer plusieurs tâches et d'effectuer les supervisions simultanément.

¹Microsoft Foundation Class : librairie de programmation sous Microsoft Windows permettant de créer des applications C++ de type WIMP.

L'architecture interne respecte les spécifications du superviseur générique, tout en simplifiant les contraintes réelles des applications de RV. Des classes représentent les tâches (types, paramètres, données), les modalités, les médias et les utilisateurs. Le contexte est géré par un module spécifique, de même que la requête au superviseur et le superviseur lui-même. Le superviseur, programmé en Prolog, est compilé sous la forme d'un fichier unique, chargé au démarrage de l'application. Les différents modules peuvent alors communiquer avec le superviseur par l'intermédiaire de commandes Prolog.

Une nouvelle fois, en réitérant la succession de commandes que nous avons effectuées en Prolog, nous confirmons que le temps entre le lancement de la requête et l'affichage (en mode texte) du rendu décidé, en passant par les mises à jour nécessaires dans la base de connaissances et dans l'application, est extrêmement rapide (moyenne de 90 ms par interaction, affichage compris). Ceci permet de penser que l'architecture et les moyens de communication entre l'application et le superviseur sont compatibles avec une activité de RV en temps réel.

Une fois l'implémentation des outils validée, la construction de la base de connaissances sur des interactions réelles a pu débuter. La section suivante rend compte de ce processus.

5.3 Construction de la base de connaissances

Nous l'avons vu dans le chapitre 4, les décisions du superviseur s'appuient sur une base de connaissances. Celle-ci contient l'ensemble des règles de distribution, qui prennent la forme de listes de modalités favorites pondérées pour chacun des éléments du modèle des tâches. L'élaboration de cette base va s'effectuer en deux étapes principales : la constitution des listes, puis la mise au point des pondérations.

Dans les deux cas nous avons suivi une même méthodologie de construction. Nous nous sommes appuyés sur un corpus d'interactions issues des applications du chapitre 3. Un sous-ensemble de ces interactions a fait l'objet de supervisions successives, entraînant la composition progressive des listes jusqu'à l'obtention de résultats corrects (i.e. contenant tout ou partie des résultats attendus). Ces listes étant composées de modalités initialisées à 1, les scores ont ensuite été affinés par de nouvelles itérations, la condition d'arrêt étant cette fois que la modalité jugée la plus adéquate selon le contexte obtienne le score final le plus élevé.

5.3.1 Corpus d'interactions

A partir de notre modélisation des interactions et de l'identification de certains types de tâches, paramètres et types de données, nous pouvons construire un corpus des interactions qui peuvent être représentées (figure 5.2). Pour simplifier les

figures, nous nous sommes restreints dans cette section aux quatre types de tâches, à six données et à cinq paramètres possibles parmi tous les éléments proposés lors de la modélisation des interactions. Une tâche doit obligatoirement comporter une donnée unique, mais peut avoir plusieurs types et aucun ou plusieurs paramètres. Nous interdisons quelques combinaisons : une sélection ne peut pas porter sur une donnée de contact ou le proxy, une navigation ne peut concerner que la scène et les paramètres quantitatifs, état ou variation n'ont pas de sens pour cette navigation. Ces cas particuliers ne restreignent pourtant pas beaucoup le nombre de tâches de RV modélisables par ces 15 éléments simples.

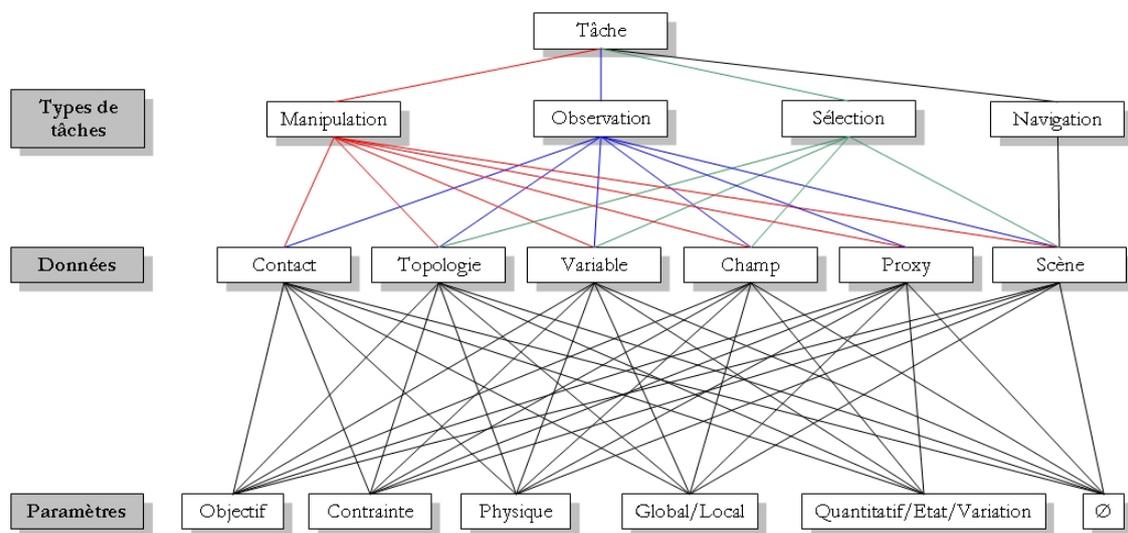


FIG. 5.2 – Graphe du corpus d'interactions possibles à partir de certains types de tâches, paramètres et données.

Parmi toutes les combinaisons possibles du corpus, nous avons choisi un sous-ensemble d'interactions pour servir à la construction de la base de connaissances. Les trois critères de choix des combinaisons étaient de :

- pouvoir leur faire correspondre un exemple de tâche réalisable dans une des applications que nous avons analysées (chapitre 3) ;
- connaître le ou les rendus multimodaux pertinents de ces tâches pour les utilisateurs (par analyse ergonomique, expériences psychophysiques, expertise ou déduction) ;
- permettre la représentation de tous les éléments du modèle mais pas de toutes les combinaisons.

La figure 5.3 contient quelques interactions qui vérifient ces trois critères.

N°	Interactions désirées par l'utilisateur	Types de tâche	Données	Paramètres	Modalités possibles
MF2	Etre totalement contraint sur une isosurface	Manipulation	Proxy	Contrainte	Contrainte haptique ou visuelle
MF3	Etre attiré sur une isosurface		Variable	Objectif	Haptisation, attraction visuelle
MF4	Ressentir la topologie d'une isosurface		Topologie (Isosurf.)	Physique & Local	Epistémique haptique, texture haptique
CAO1	Ressentir la collision entre la pièce qu'il manipule et un plan		Contact	Physique	Epistémique haptique ou visuelle, coloration du contact, son de collision
BI2	Ressentir les forces électrostatiques entre deux protéines		Variable	Variation	Haptisation, auralisation, <i>earcon</i> , vibrations
MF6	Savoir si la température en une position d'une simulation de liquide est supérieure ou inférieure à un seuil	Observation	Variable	Etat	Texte, couleur, icône, auralisation, vibrations
BI1	Etudier l'hydrophobicité d'une protéine		Champ de variables	Etat & Global	Couleur, topologie, icônes

FIG. 5.3 – Tableau des interactions élémentaires ayant servi à la construction de la base de connaissances.

5.3.2 Contexte élémentaire

Le contexte sur lequel nous nous sommes appuyés dans un premier temps est simple. Les éléments statiques (ceux fixés avant le lancement des interactions) sont :

- le handicap utilisateur ;
- la disponibilité des médias ;
- les préférences utilisateur (choix de rendus spécifiques pour certains éléments du modèle).

Les éléments dynamiques sont :

- la charge temporelle des canaux (i.e. le nombre de modalités par canal à un instant donné) ;
- la colocalisation entre l'attention visuelle de l'opérateur et la donnée concernée par la tâche.

5.3.3 Constitution des listes de modalités

La méthode pour initialiser la base de connaissances consiste à exécuter itérativement des requêtes de rendu sur les tâches du sous-ensemble du corpus. Les listes de modalités favorites des différents éléments représentés par les combinaisons choisies sont modifiées jusqu'à ce que le superviseur donne des résultats corrects dans les situations idéales. Puis le même processus est lancé dans des contextes différents : les modalités de substitution ou de redondance doivent être choisies respectivement à la place des modalités impossibles ou en complément des modalités insuffisantes.

A la suite de ce processus, les listes que nous avons obtenues peuvent être re-

		Modalités favorites														
		Visuel							Audio				Haptique			
		Topologie	Texte	Widget	Couleur	Guide	Contrainte	Physique	Icone	Mapping	Vocal	Synthèse	Mapping	Guide	Contrainte	Physique
Eléments du modèle des interactions	Types de tâche	Navigation	•	•	•	•	•	•								
		Observation	•	•	•	•	•	•	•	•	•	•				
		Manipulation	•	•	•	•	•	•	•				•	•	•	•
		Sélection	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	Paramètres	Objectif			•		•			•			•	•		
		Contrainte						•		•			•		•	
		Physique						•				•				•
		Global	•			•										
		Local														•
		Quantitatif		•							•					
		Etat		•	•	•										
		Variation								•			•			
	Données	Topologie	•			•										
		Contact				•		•	•							•
		Variable		•						•			•			
Champ					•								•			
Proxy							•								•	

FIG. 5.4 – Tableau des modalités favorites par éléments du modèle des interactions.

présentées dans un tableau (figure 5.4). Cette représentation permet par ailleurs de déterminer visuellement le résultat de la supervision d'une tâche dans un contexte idéal. Par exemple (figure 5.5), pour une tâche de **manipulation** d'une **topologie** avec **contrainte**, les modalités qui ont le score maximum de 2 sont : **topologie**, **couleur** et **contrainte** visuelles, **mapping** et **contrainte** haptiques.

Cet exemple montre bien la nécessité d'affiner les scores, de manière à hiérarchiser les modalités au sein des listes, et ainsi restreindre le nombre de modalités dans la liste résultat.

5.3.4 Affinage des pondérations

Le travail d'affinage des pondérations s'avère plus délicat que la constitution des listes. Il s'agit d'effectuer de nouvelles supervisions itératives sur le sous-ensemble

		Modalités favorites														
		Visuel						Audio				Haptique				
		Topologie	Texte	Widget	Couleur	Guide	Contrainte	Physique	Icone	Mapping	Vocal	Synthèse	Mapping	Guide	Contrainte	Physique
Eléments du modèle des interactions	Types de tâche	Navigation	•	•	•	•	•	•								
		Observation	•	•	•	•	•	•	•	•	•	•				
		Manipulation	•	•	•	•	•	•					•	•	•	•
		Sélection	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	Paramètres	Objectif			•		•			•			•	•		
		Contrainte					•			•			•	•		
		Physique						•				•				•
		Global	•			•										
		Local														•
		Quantitatif		•							•					
		Etat		•	•	•										
	Variation								•			•				
	Données	Topologie	•			•										
Contact					•		•	•							•	
Variable			•						•			•				
Champ					•								•			
Proxy							•								•	

FIG. 5.5 – Tableau des modalités favorites par éléments du modèle des interactions : exemple.

du corpus, mais cette fois pour qu'une seule modalité attendue obtienne le score final le plus élevé. Il faut également s'assurer que les rendus sont corrects dans des contextes non idéaux (e.g. absence d'un dispositif de rendu, handicap de l'utilisateur, surcharge due aux interactions précédentes, etc.).

Pour l'heure, les scores ont été peu modifiés. Une des perspectives de recherche ouverte par ces travaux est l'utilisation de procédures automatisées (type réseaux de neurones) à la place des réglages manuels.

Les listes obtenues à la suite de ce processus de pondération (figure 5.6) fournissent également des règles de distribution génériques, pouvant être réutilisées en dehors du superviseur multimodal. Elles devront bien sûr faire l'objet d'évaluations ergonomiques.

		Modalités favorites													
		Visuel						Audio				Haptique			
		Topologie	Texte	Widget	Couleur	Guide	Contrainte	Physique	Icone	Mapping	Vocal	Synthèse	Mapping	Guide	Contrainte
Types de tâche	Navigation	1													
	Observation	1						1,1							
	Manipulation	1										1,1			
	Sélection	1						0,9				1			
Paramètres	Objectif			0,9		1			1			1	1,1		
	Contrainte					1		1				1		1,1	
	Physique						1			0,9					1
	Global	1			1										
	Local														1
	Quantitatif		1							1					
	Etat		1	1	1										
	Variation								1,1			1			
Données	Topologie	1			1										
	Contact				1		1	1							1
	Variable		1						1			1			
	Champ				1								1		
	Proxy						1								1

FIG. 5.6 – Tableau des modalités favorites pondérées par éléments du modèle des interactions.

5.4 Evaluation de la supervision

Une fois la base de connaissances construite sur un sous-ensemble du corpus des interactions, nous avons évalué la supervision sur d'autres combinaisons vérifiant les mêmes critères (existence d'un exemple concret et connaissance d'un rendu pertinent), puis sur des scénarii interactifs plus complexes, mettant en jeu plusieurs interactions successives ou simultanées, ainsi que des perturbations du contexte.

5.4.1 Interactions élémentaires

Afin d'illustrer les premières évaluations que nous avons effectuées, considérons trois exemples de tâches concrètes (figure 5.7), modélisées par des combinaisons

d'éléments qui n'ont pas servi à la construction de la base de connaissances.

N°	Interactions désirées par l'utilisateur	Types de tâche	Données	Paramètres
1	Déplacer un objet sans pouvoir aller dans une zone de l'espace	Manipulation	Topologie	Contrainte
2	Etudier la géométrie d'un objet	Observation		Global
3	Etre informé d'un contact entre deux objets de la simulation		Contact	Physique

FIG. 5.7 – Interactions élémentaires pour l'évaluation : exemples.

La supervision de ces tâches dans un contexte idéal donnent les résultats de la figure 5.8.

```

Start : normal 1
[(hconstraint , 2.2), (hmapping , 2.1), (vconstraint , 2), (vcolor , 2),
(vtopo , 2), (hphys , 1.1), (hguide , 1.1), (amapping , 1), (vphys , 1),
(vguide , 1), (vwidget , 1), (vtext , 1), (asynth , 0), (avocal , 0),
(aicon , 0)]
-> [hconstraint] 2.2
Stop : normal 1

Start : normal 2
[(vcolor , 3), (vtopo , 3), (asynth , 1.1), (avocal , 1.1), (amapping ,
1.1), (aicon , 1.1), (vphys , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1), (hphys , 0), (hconstraint , 0), (hguide , 0),
(hmapping , 0)]
-> [vcolor, vtopo] 3
Stop : normal 2

Start : normal 3
[(vphys , 3), (aicon , 2.1), (hphys , 2), (asynth , 2),
(vcolor , 2), (avocal , 1.1), (amapping , 1.1), (vconstraint , 1),
(vguide , 1), (vwidget , 1), (vtext , 1), (vtopo , 1), (hconstraint , 0),
(hguide , 0), (hmapping , 0)]
-> [vphys] 3
Stop : normal 3

```

FIG. 5.8 – Résultats de l'évaluation sur des interactions élémentaires.

1. La tâche n°1 demande un rendu par contrainte haptique. Ce choix est influencé par la manipulation et le paramètre contrainte.
2. Pour la tâche n°2, le superviseur obtient deux modalités de même score : couleur et topologie. Dans notre cas, l'étude de la géométrie demanderait plutôt

un rendu topologique que l'utilisation de couleur. Il serait donc intéressant de supprimer la couleur des modalités favorites pour la donnée `topologie`, de la conserver pour le paramètre `global`, et de modéliser la tâche n°2 sans aucun paramètre.

3. La tâche n°3 n'est pas liée à l'utilisateur. Il semble donc cohérent de ne pas utiliser le canal haptique pour communiquer cet événement de la scène. Le superviseur a choisi le rendu physique visuel, qui informe du contact mais sans le mettre en valeur (par de la couleur ou un son). Nous pourrions imaginer de créer une règle spécifique qui favorise les modalités de mise en valeur lorsque l'utilisateur accorde beaucoup d'importance à une tâche (ce qui se traduit pas un niveau élevé de la tâche dans la clause de création).

Le superviseur multimodal se montre donc cohérent et capable de contrôler le rendu de tâches variées, à partir d'une base de connaissances générique. Il est intéressant d'évaluer maintenant s'il est utilisable dans des conditions réalistes, c'est-à-dire pour gérer le rendu d'interactions successives et simultanées et des contextes non idéaux. Nous l'évaluons dans la section suivante sur un scénario interactif de Bio-Informatique.

5.4.2 Scénario de Bio-Informatique

L'application de Bio-Informatique en cours de développement dans le cadre du projet CoRSAIRE a pour vocation d'assister les biologistes dans l'étude du problème du docking de protéines (cf. chapitre 3). Dans son état actuel, l'application (figure 5.9) permet, dans un environnement immersif, de manipuler deux protéines à l'aide de deux interfaces d'entrée : un bras haptique² et une souris sans fil munie d'un dispositif de suivi infrarouge. Le dispositif de projection permet la stéréoscopie active. Plusieurs représentations visuelles des protéines sont possibles : atomique, surfacique, bâton et structurelle. Au niveau biologique, des fonctionnalités de calculs de forces électrostatiques et de complémentarité hydrophobe sont disponibles. Une librairie de détection de collisions a également été implémentée. Concernant l'audio, un module sous Max/MSP permet de présenter des modalités sonores spatialisées à l'aide de haut-parleurs.

5.4.2.1 Protocole

Le protocole de test de l'activité de docking que nous avons choisi crée plusieurs tâches successives ou simultanées, effectue la supervision de celles-ci, affiche les résultats et met à jour les éléments nécessaires (base de contexte et enregistrement des interactions courantes).

²Virtuose de chez Haption, à 6 degrés de liberté et 3 degrés de retour d'effort



FIG. 5.9 – Application de Docking de Protéines utilisée pour le scénario d'évaluation

Utilisateur L'utilisateur est un biologiste, expert de l'application et sans handicap :

```
user(expert, [visual, audio, haptic], 3).
```

Contexte Dans un premier temps, le contexte est idéal. Tous les médias/canaux sont disponibles et sont libres au démarrage de l'application (ils ont donc une charge nulle) :

```
channel(visual, 0).
channel(audio, 0).
channel(haptic, 0).
```

Dans un deuxième temps, le même protocole sera réalisé sans interface haptique.

Objets Les deux objets présents dans l'application sont deux protéines que nous appellerons *x* et *y* (*Barnase* et *Barstar* dans la figure 5.9), manipulables en position et en orientation, non déformables. Elles sont compatibles avec les trois canaux de rendu : elles disposent des fonctions de rendu de toutes les modalités que le superviseur pourrait choisir.

```
obj(x, [visual, audio, haptic]).
obj(y, [visual, audio, haptic]).
```

Tâches Voici la succession des tâches que l'utilisateur veut réaliser :

1. Observation de la topologie des 2 protéines. Un rendu par défaut est prévu par le concepteur et doit être favorisé par le superviseur : il s'agit de la modalité d'affichage surfacique ou **topologie visuelle**. La tâche est modélisée, puis enregistrée dans le moteur (en Prolog) par les faits correspondants :

```
inter(affx, false, 3).
task(affx, obs).
data(affx, topo, [x]).
appli(affx, [vtopo]).
```

```
inter(affy, false, 3).
task(affy, obs).
data(affy, topo, [y]).
appli(affy, [vtopo]).
```

où `inter()` crée la tâche par un identifiant, un lien avec l'utilisateur et un niveau, `task()` précise son type, `data()` le type de donnée et les objets concernés, et `appli()` indique quel est le rendu proposé par le concepteur de l'application.

2. Commande de rendu de la complémentarité topologique entre les 2 protéines. Un algorithme permet de calculer les distances minimales entre les protéines et de constituer un score global qui permet d'avoir des informations sur la complémentarité géométrique.

```
inter(comptopo, true, 3).
task(comptopo, man).
data(comptopo, variable, [x, y]).
param(comptopo, state).
```

où `param` indique le paramètre de la tâche.

3. Commande de rendu de la complémentarité électrostatique entre les 2 protéines. Un algorithme permet de calculer la force électrostatique résultante entre les protéines.

```
inter(copelec, true, 3).
task(copelec, man).
data(copelec, variable, [x, y]).
param(copelec, variation).
```

4. Arrêt de la tâche de complémentarité électrostatique.

5. Arrêt de la tâche de complémentarité topologique.
6. Commande de rendu physique du contact géométrique.

```
inter(puz, true, 3).
task(puz, man).
data(puz, topo, [x, y]).
param(puz, phys).
```

7. Arrêt du rendu physique.
8. Arrêt de l’affichage visuel.

5.4.2.2 Résultats : contexte idéal

Dans un contexte idéal, les résultats sont présentés dans la figure 5.10.

Les rendus choisis par le superviseur sont les rendus attendus :

1. Les observations topologiques sont réalisées à l’aide du rendu par défaut proposé par le concepteur : ce choix a eu une forte influence sur le choix du superviseur.
2. La complémentarité topologique est transmise par haptisation : ce choix a été guidé par la prédominance de l’haptique pour la tâche de **manipulation**, car la donnée **variable** et le paramètre **variation** privilégient autant l’auralisation que l’haptisation.
3. la complémentarité électrostatique est ensuite rendue par auralisation. Si cette tâche avait été unique, les résultats auraient été les mêmes que précédemment. Cependant, un filtrage spécifique a supprimé l’haptisation de la liste finale, pour cause d’incompatibilité avec une haptisation existante.
6. Le rendu physique des contacts est effectué par le dispositif haptique : ce choix est influencé par le type de tâche (**manipulation**), le paramètre (**physique**) et la donnée (**contact**).

5.4.2.3 Résultats : sans interface haptique

En l’absence d’interface haptique, les choix du superviseur sont cohérents, même si évidemment les rendus semblent moins adaptés que dans un contexte idéal (figure 5.11) :

1. Le rendu des observations topologiques n’est pas modifié.
2. La complémentarité topologique est transmise par auralisation : en l’absence d’haptisation, cette modalité auditive est le choix de substitution (2ème modalité de la liste dans le contexte idéal).

```

Start : normal affx
[(vtopo , 4), (vcolor , 2), (asynth , 1.1), (avocal , 1.1), (amapping ,
1.1), (aicon , 1.1), (vphys , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1), (hphys , 0), (hconstraint , 0), (hguide , 0),
(hmapping , 0)]
-> [vtopo] 4

Start : normal affy
[(vtopo , 3.5), (vcolor , 2), (asynth , 1.1), (avocal , 1.1), (amapping ,
1.1), (aicon , 1.1), (vphys , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1), (hphys , 0), (hconstraint , 0), (hguide , 0),
(hmapping , 0)]
-> [vtopo] 3.5

Start : normal comptopo
[(hmapping , 3.1), (amapping , 2.1), (vtext , 2), (hphys , 1.1),
(hconstraint , 1.1), (hguide , 1.1), (vphys , 1), (vconstraint , 1),
(vguide , 1), (vcolor , 1), (vwidget , 1), (vtopo , 0.5), (asynth , 0),
(avocal , 0), (aicon , 0)]
-> [hmapping] 3.1

Start : normal compelec
[(amapping , 2.1), (vtext , 2), (vphys , 1), (vconstraint , 1), (vguide ,
1), (vcolor , 1), (vwidget , 1), (vtopo , 0.5), (asynth , 0), (avocal ,
0), (aicon , 0)]
-> [amapping] 2.1

Stop : normal compelec
Stop : normal comptopo

Start : normal puz
[(hphys , 2.1), (vphys , 2), (vcolor , 2), (vtopo , 1.5), (hconstraint ,
1.1), (hguide , 1.1), (hmapping , 1.1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1), (asynth , 0.9), (avocal , 0), (amapping , 0),
(aicon , 0)]
-> [hphys] 2.1

Stop : normal puz
Stop : normal affx
Stop : normal affy

```

FIG. 5.10 – Résultats de l'évaluation sur le scénario de Bio-Informatique (contexte idéal).

3. La complémentarité électrostatique est ensuite affichée textuellement. Comme dans le cas précédent, c'est un filtrage spécifique qui a détecté l'existence d'une auralisation déjà en cours de rendu, et qui a diminué le score de cette modalité auditive pour la nouvelle tâche (-0,5).
6. Le rendu physique des contacts ne peut pas être effectué par le dispositif haptique. Les modalités de substitution sont la physique visuelle (non inter-pénétration) et une colorisation des contacts.

```

Start : normal affx
[(vtopo , 4), (vcolor , 2), (asynth , 1.1), (avocal , 1.1), (amapping ,
1.1), (aicon , 1.1), (vphys , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1)]
-> [vtopo] 4

Start : normal affy
[(vtopo , 3.5), (vcolor , 2), (asynth , 1.1), (avocal , 1.1), (amapping ,
1.1), (aicon , 1.1), (vphys , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1)]
-> [vtopo] 3.5

Start : normal comptopo
[(amapping , 2.1), (vtext , 2), (vphys , 1), (vconstraint , 1), (vguide ,
1), (vcolor , 1), (vwidget , 1), (vtopo , 0.5), (asynth , 0), (avocal ,
0), (aicon , 0)]
-> [amapping] 2.1

Start : normal compelec
[(vtext , 2), (amapping , 1.6), (vphys , 1), (vconstraint , 1), (vguide ,
1), (vcolor , 1), (vwidget , 1), (vtopo , 0.5), (asynth , 0), (avocal ,
0), (aicon , 0)]
-> [vtext] 2

Stop : normal compelec
Stop : normal comptopo

Start : normal puz
[(vphys , 2), (vcolor , 2), (vtopo , 1.5), (vconstraint , 1), (vguide ,
1), (vwidget , 1), (vtext , 1), (asynth , 0.9), (avocal , 0), (amapping ,
0), (aicon , 0)]
-> [vphys, vcolor] 2

Stop : normal puz
Stop : normal affx
Stop : normal affy

```

FIG. 5.11 – Résultats de l'évaluation sur le scénario de Bio-Informatique (contexte non idéal : sans haptique).

5.5 Synthèse et perspectives

Dans cette section, nous faisons la synthèse de nos principaux résultats et rappelons les avantages d'un langage de programmation logique pour l'implémentation de notre superviseur multimodal. Puis nous discutons certains points qui n'ont pas (ou peu) été abordés auparavant et qui soulignent la complexité et la diversité des problèmes liés à un rendu multimodal intelligent : la dépendance entre multimodalité en entrée et multimodalité en sortie, les effets du rendu sur les utilisateurs, l'instanciation des rendus et les interactions multi-utilisateurs. Nous examinons les apports potentiels de notre modèle à ces problèmes, ce qui donne un aperçu des perspectives à court et long terme de nos travaux.

5.5.1 Principaux résultats

Plusieurs résultats sont à retenir de ces premières implémentations et évaluations. Le premier résultat important est que notre modèle et notre moteur de supervision permettent de décider d'un rendu multimodal pour des tâches de RV. Nous avons pu tester le processus de supervision pour des tâches précises tirées des applications du chapitre 3, et dans des contextes d'interactions courants : les rendus décidés s'avèrent cohérents et respectueux des attentes des utilisateurs. Plus globalement, le comportement du superviseur est conforme aux besoins que nous avons identifiés dans ces applications : gérer le contrôle global des interactions afin de fournir à l'utilisateur les rendus les plus adéquats en fonction de l'information à transmettre et de la situation.

Le second résultat, qui concerne directement la future implémentation du superviseur dans des dispositifs de RV, est que les décisions de rendu sont prises en temps réel. Cela montre que, dans les cas que nous avons évalués, le recours à un moteur intelligent ne pénalise pas de manière significative le rendu multimodal, par rapport à des décisions codées au sein de l'application.

Nos évaluations montrent enfin qu'une décomposition en éléments simples des tâches, ainsi qu'un nombre limité de règles et de connaissances sur chacun des éléments séparés peut suffire à garantir un résultat correct sur un large panel de tâches (voire, peut prédire le rendu de tâches nouvelles).

Néanmoins, un vaste travail reste à effectuer pour consolider la base de connaissances et les règles, afin qu'elles s'adaptent à de nombreuses autres situations.

5.5.2 Choix de réalisation

Comme nous l'avons vu, le choix d'un langage de programmation logique tel que Prolog pour la réalisation du superviseur multimodal a été animé par des raisons d'ingénierie et dans un objectif d'Intelligence Artificielle. La plupart des moteurs de rendu multimodal existants fonctionnent à partir de règles de distribution établies en langage XML. Choisir Prolog nous prive des capacités de XML à échanger simplement des informations balisées. En contrepartie, nous nous évitons une description extensive de l'ensemble des interactions multimodales possibles, puisque notre supervision est basée sur des règles (représentation en intension) qui prennent en compte des éléments génériques : les composants de la tâche, les possibilités de rendu et le contexte.

De plus, les outils logiques de preuve et d'apprentissage de règles vont être d'une grande utilité pour l'évaluation et le perfectionnement futur de notre moteur de décision. La manière dont le moteur est conçu permet notamment la mise à jour simple de la base de connaissances (modification des listes de modalités favorites et des pondérations) ainsi que l'ajout de règles de décision ou de prédicats statiques

(modalités, médias, utilisateurs, etc.) pour s'adapter à de nouvelles interactions. Il est également possible de satisfaire les besoins spécifiques de certains domaines, applications, utilisateurs ou contextes d'utilisation, en ajoutant dynamiquement des règles ou des connaissances particulières à l'ensemble des règles et des connaissances génériques.

Le choix du développement d'un simulateur a été lui influencé d'une part par les contraintes matérielles et temporelles du développement, d'autre part par notre objectif de généricité du moteur de supervision. En effet, la validation du processus de supervision au sein d'une architecture globale nécessitait que les applications visées (Mécanique des Fluides, Bio-Informatique et CAO) soient suffisamment avancées, et que les communications entre les acteurs de la supervision soient implémentées sous les contraintes de chaque application. Ces contraintes ont été simplifiées en simulant chacun des aspects des applications réelles : dispositifs, moteurs de rendu, commandes et suivi de l'utilisateur, scène, etc. Seuls persistent les flux d'informations entre l'application et le superviseur : requête de rendu pour une tâche, mises à jour d'éléments du contexte, décision de rendu multimodal, etc.

Ce développement intermédiaire a également permis de clarifier les spécifications sur les entrées attendues et les sorties possibles du superviseur, en vue de réaliser une véritable "boîte noire" susceptible d'être utilisée par les différentes applications que l'on voudrait doter de multimodalité en sortie. Ces spécifications liées au caractère générique du superviseur vont contraindre certains aspects des applications cibles. En particulier, celles-ci devront implémenter le modèle des interactions, ainsi que la reconnaissance des résultats du superviseur (listes de modalités) pour les transformer en rendus concrets.

5.5.3 Dépendance entre entrée et sortie

Les recherches en multimodalité séparent souvent l'entrée de la sortie, la fusion de la fission, la commande du rendu. Pourtant, nous avons fondé notre supervision des rendus multimodaux sur une modélisation des tâches désirées par l'utilisateur. Or, ces tâches sont directement liées aux entrées du système, soit car elles sont données explicitement par des commandes (vocales, gestuelles, etc.), soit parce qu'elles sont inférées automatiquement par le système, à partir des périphériques de suivi à sa disposition (suivi de tête, de main, microphone, etc.). La multimodalité en sortie est donc dépendante des entrées, qui représentent la tâche, les objectifs de l'utilisateur, le contexte de l'interaction, etc.

Même si nos travaux se sont concentrés sur la partie aval de cette reconnaissance des tâches, nous pensons que les recherches sur la multimodalité en sortie ne peuvent ignorer celles sur la multimodalité en entrée.

5.5.4 Respect des choix de l'utilisateur

L'une des spécifications que respecte notre superviseur (section 4.1, p.94) est qu'il soit "assujetti aux commandes utilisateur et à ses choix, lorsque c'est possible et sans danger". Cela suppose d'abord que les préférences de l'utilisateur concernant des modalités de rendu pour des tâches ou des éléments de tâches soient prises en compte dans la pondération de la décision. Ensuite, le processus de supervision doit également permettre à l'utilisateur de refuser des rendus qui lui sont présentés auquel cas le superviseur devra retraiter la requête en modifiant le résultat. Dans certains cas, par exemple si le superviseur aboutit à une modalité résultat de score faible ou très différente des rendus habituels, il serait intéressant que cette décision soit soumise à l'utilisateur avant tout rendu concret. Ce mode de dialogue non systématique peut être identifié par une modalité spécifique **proposition**, qui prend le pas sur la modalité résultat. Celle-ci est mise en attente. Si l'utilisateur accepte la proposition, la modalité de rendu est réactivée et présentée. Le cas échéant, le superviseur retraite la requête en supprimant la modalité refusée.

De plus, cet outil de validation des rendus permettrait de participer à l'automatisation de l'affinage des connaissances : les modalités les plus plébiscitées ou les moins validées seraient respectivement privilégiées ou retirées des listes de modalités favorites.

5.5.5 Latences

Un être humain est très sensible aux incohérences sensorielles (a fortiori sensorimotrices) qui peuvent être induites par les interfaces humain-environnement. La latence (i.e. le temps qui s'écoule entre le moment où le rendu est décidé par l'application et le moment où les stimuli sont ressentis par l'humain) fait partie des incohérences à traiter lorsque l'on développe une application de RV. L'utilisateur va pouvoir accepter dans une certaine mesure des retards dans les rendus, mais uniquement si ces retards sont faibles et cohérents, c'est-à-dire s'ils restent constants pour les mêmes interactions.

Notre modèle de supervision des rendus autorise deux gestions différentes des latences. Grâce à la maîtrise de la temporalité des rendus, le moteur permet une gestion en amont des latences. Lorsque des modalités ont des latences prévisibles (par exemple intrinsèques aux dispositifs matériels ou à la communication réseau) ou connues par l'expérience (tests de performance), celles-ci peuvent être intégrées à la base de connaissances. Ainsi, lors de la redondance de modalités, si l'une possède des latences, le superviseur peut les compenser en ajoutant des retards aux autres modalités.

Une gestion des latences par la substitution est également possible : lorsque le superviseur découvre que la modalité qui découle du processus de décision possède

une latence trop élevée pour l'interaction (par exemple si celle-ci a un fort niveau de priorité, ou si les données manipulées sont critiques, ou encore si l'utilisateur est débutant), il peut alors la substituer par une des modalités suivantes de la liste résultat, soit définitivement, soit en attendant que la première choisie soit disponible.

Dans tous les cas, les latences ne peuvent être prises en compte par le superviseur que si elles sont connues à l'avance, et intégrées dans la base de connaissances. L'identification concrète des latences est cependant complexe, et constitue un problème en soi.

5.5.6 Transitions entre rendus

Un autre problème lié à l'effet des rendus sur les utilisateurs et que le superviseur doit gérer à moyen terme est le problème des transitions entre modalités. Cette question peut se poser dans différents cas :

- le retraitement d'une tâche donne lieu à une décision différente de la décision précédente ;
- une tâche laisse la place à une nouvelle tâche de même type (par exemple mêmes paramètres et même type de donnée mais sur des objets différents de l'EV) qui donne lieu à une décision différente de la décision précédente ;
- une tâche laisse la place à une nouvelle tâche différente, mais dont le rendu décidé est le même.

Il faut donc vérifier la compatibilité matérielle et sémantique entre l'ancien et le nouveau rendu, et gérer la transition effective entre les deux.

Un comportement simple adapté à tous les cas est d'utiliser le système de dialogue abordé précédemment, afin de laisser l'utilisateur valider ou non le nouveau rendu.

Les effets du premier cas peuvent aussi être limités en ajoutant dans la base de connaissances des couples de modalités successives à éviter lorsqu'elles concernent la même tâche (sur le modèle des incompatibilités de modalités simultanées). Ainsi, le nouveau rendu ne pourra pas être incompatible avec l'ancien. Cependant, même si les rendus sont compatibles, la transition peut rester difficile. Lorsqu'elle concerne le même canal, il est possible d'utiliser des procédés de fondus visuels, audio ou haptiques (par analogie avec les fondus musicaux dans les logiciels de musique). Lorsque les deux rendus sont sur des canaux différents, une approche possible est l'utilisation de redondance, évoquée dans la section 4.4.4 (p. 134). Elle suppose des "retraitements fictifs" périodiques afin de prévoir un futur changement de rendu, et d'adoucir la transition en superposant le prochain rendu à l'ancien avant le remplacement complet.

Le second cas est délicat car le rôle du superviseur est de décider du rendu le plus adéquat pour une interaction donnée. Or, si le rendu décidé pour la seconde interaction est différent de la première, c'est qu'un élément du processus de pondération a donné lieu à un changement : objet concerné ou contexte. Donc, le rendu

précédent n'est plus le mieux adapté, voire est impossible. S'il reste possible, seule une commande de l'utilisateur peut explicitement demander le recours à cet ancien rendu.

Le troisième cas ne pose pas de difficulté sensorielle, mais sémantique. Le superviseur ne peut pas gérer cette situation. Il faut donc veiller à ce que l'utilisateur ait conscience que le nouveau rendu s'applique à une nouvelle interaction, ce qui est évident si la commande est explicite, et doit lui être signalé lorsque l'interaction a été déclenchée par l'application.

5.5.7 Modules de rendu externes

L'abstraction des modalités lors de l'élaboration du modèle du rendu fait que certaines d'entre elles ne peuvent pas être concrétisées directement par l'application à partir de leur seule dénomination. Dans ce cas, la réception du résultat de la supervision commande l'appel à des modules dédiés à des fonctionnalités particulières.

Par exemple, les modalités visuelles **niveau de détail** ou **caméra** font appel à des algorithmes respectivement de calcul du niveau de détail requis, et de recherche du meilleur point de vue à partir des positions de l'effecteur virtuel, de l'utilisateur et de la donnée. De même, les modalités d'ordre **physique**, que ce soit haptique ou visuel, ne peuvent être rendues qu'après appel à un moteur physique de détection des collisions, qui est en charge de calculer les forces de répulsion des objets. Un autre exemple est le guide virtuel : qu'il soit visuel ou haptique, il nécessite des calculs de trajectoire qui dépendent de la cible et de la position de l'utilisateur.

5.5.8 Interactions multi-utilisateurs

Un aspect des interactions de RV que nous n'avons pas encore abordé est la présence de plusieurs utilisateurs qui coopèrent ou du moins interagissent sur le même environnement virtuel, soit dans le même dispositif physique, soit dans des dispositifs distants. La multiplicité des utilisateurs complique d'autant les problèmes de multimodalité en entrée et en sortie. D'une part il faut identifier les interactions propres à chacun, et contrôler la cohabitation (compatibilité, surposition spatiale et temporelle) des rendus associés. D'autre part, pour les interactions de coopération sur une donnée commune, il faut garder la possibilité d'avoir des rendus différenciés pour chacun.

Dans son état actuel, notre modèle permet de créer simplement de nouveaux utilisateurs, ayant chacun leurs caractéristiques (handicap, expertise). Ensuite, les interactions faisant l'objet d'une requête sont personnalisées (i.e. liées à un utilisateur). Dans le cas où les utilisateurs effectuent des interactions dissociées, le processus de décision de rendu reste identique au mode mono-utilisateur : les règles de compatibilité ou de surcharge temporelle et spatiale modifient les rendus qui sont

problématiques en fonction des rendus courants, quelque soit l'utilisateur concerné. Le cas des interactions de coopération est plus délicat et le moteur de supervision n'a pas été conçu dans cette optique. En effet, il ne s'agit pas d'effectuer une requête pour une interaction commune à plusieurs utilisateurs, puisqu'ils peuvent agir de manière différente mais sur une donnée commune. Il s'agit plutôt de créer deux interactions distinctes dont le prédicat `data` porte sur le même type de donnée et les mêmes objets, puis de gérer les priorités entre les interactions, et le partage des rendus de chaque interaction. Ceci ne peut être fait en ajoutant simplement quelques règles au processus existant.

5.6 Conclusion

Les premières évaluations de notre superviseur multimodal ont apporté des résultats satisfaisants, que ce soit concernant la qualité des rendus décidés pour des tâches variées, que la rapidité de la décision et l'adaptabilité à des contextes différents. Elles ont également mis à jour de nouvelles problématiques en rendu multimodal, parmi lesquelles la gestion des latences, la gestion des transitions et la coopération entre utilisateurs. Ces trois questions font partie des perspectives de recherche que nous voulons aborder.

Conclusion générale et perspectives

Les travaux présentés dans ce mémoire ont été motivés par deux constatations. La première est qu'un grand nombre d'applications de RV n'exploite pas la vaste bande passante sensorimotrice humaine. Cependant, il apparaît aussi que d'autres applications expérimentent avec succès la conception de nouvelles interactions multimodales, parfois non réalistes, qui améliorent les performances, le confort ou la compréhension des utilisateurs dans les environnements virtuels.

L'ambition de nos recherches résidait dans une approche globale, générique et contextuelle du rendu multimodal en RV, c'est-à-dire en nous éloignant des solutions adhoc et figées. Pour assister l'utilisateur dans ses activités, nous avons voulu fonder la détermination du rendu le plus adéquat sur trois piliers : la sémantique des tâches, les caractéristiques des canaux et des modalités et le contexte des interactions.

Nous avons vu que la principale caractéristique de la RV est d'être le siège d'interactions entre un ou des humains et un environnement numérique, et que l'immersion de l'humain dans son activité pose des problèmes particuliers, qui nécessitent des études psychophysiques et des évaluations ergonomiques. Nous avons montré en quoi les recherches sur la multimodalité en sortie nous semblent utiles, et nous avons détaillé le cadre et l'objectif de cette thèse : concevoir un système de gestion du rendu multimodal prenant en compte la sémantique des informations et le contexte de la tâche.

L'état de l'art a montré que la combinaison intelligente de modalités selon différents canaux sensori-moteurs peut procurer des bénéfices aux applications de RV et aux utilisateurs. Les études se multiplient tant dans les domaines applicatifs (apprentissage, exploration de données scientifiques, assistance à l'opérateur) que dans les champs de la modélisation et de la psychophysique. Cependant, ces études se montrent limitées par le caractère statique de la multimodalité. C'est pourquoi nous avons eu pour objectif de concevoir un moteur de rendu multimodal générique permettant : 1) de réaliser un rendu adapté à diverses informations et tâches en environnement virtuel et 2) de tenir compte du contexte dynamique de l'exécution, lié en particulier à la succession et la simultanéité des interactions.

Afin de développer un système utile et utilisé, nous avons pris pour socle de nos travaux de modélisation et d'implémentation trois applications de RV : l'exploration d'écoulements en Mécanique des Fluides, l'appariement de protéines en Bio-Informatique et la conception de formes assistée par la RV. Après une analyse ergonomique des besoins, nous avons présenté des recommandations en terme de multimodalité. Celles-ci ont permis de constituer un cahier des charges et un corpus de test pour la réalisation et l'évaluation du superviseur de rendu multimodal.

Puis nous avons présenté en détail notre concept d'un superviseur de rendu multimodal en RV. Un processus de supervision a été élaboré. Ce dialogue entre le superviseur et l'application se doit de suivre un langage commun. Pour cela, nous avons proposé un modèle des interactions de RV adapté au problème, composé d'un modèle des tâches et d'un modèle du rendu multimodal. Le superviseur décide du rendu multimodal adéquat par un processus logique, qui utilise à la fois des connaissances statiques (projection entre informations et modalités) et des connaissances dynamiques sur le contexte d'interaction.

Le processus de décision ainsi que l'architecture de supervision ont été implémentés et validés sur des exemples concrets tirés de l'analyse des applications. Puis nous avons proposé une méthode de construction de la base de connaissances et, à l'aide d'un simulateur d'application de RV, nous avons évalué les rendus décidés par le superviseur pour des tâches variées. La décision s'est avérée fiable, rapide et adaptative à des contextes différents. Nous avons également mis à jour de nouvelles problématiques en rendu multimodal.

Contributions majeures

Cette thèse apporte plusieurs contributions aux recherches en Réalité Virtuelle, mais également dans les domaines applicatifs visés.

Tout d'abord, notre approche se veut globale, générique, indépendante des dispositifs et non spécifique à des applications. Ainsi, le processus de supervision et le moteur de décision sont conçus pour s'adapter à différentes applications et domaines d'activité. De plus, nous avons pris en compte les spécificités de la RV par rapport aux IHM non immersives, à savoir : un environnement 3d dynamique, des dispositifs de rendu riches, des tâches et un contexte d'interaction complexes et difficilement prévisibles (liés entre autres au fait que l'utilisateur est mobile) et un utilisateur humain présent à l'intérieur de l'EV, dont les caractéristiques psychophysiques doivent être prises en compte.

Ensuite, nous avons développé un modèle des interactions de RV, composé d'un modèle des tâches et d'un modèle du rendu multimodal. Le premier permet de décrire

la sémantique des tâches à l'aide d'éléments simples et génériques. Le second sert à représenter les capacités de communication d'information des systèmes de RV.

Le coeur du processus de supervision est le moteur de décision. Il est composé de trois éléments principaux : des règles de comportement ; une base de connaissances renfermant des règles de projection entre composants du modèle des tâches et composants du modèle du rendu ; une base du contexte contenant des informations qui peuvent influencer le rendu. Les connaissances internes à ces trois outils, et notamment les règles de projection, pourraient être utilisées indépendamment du moteur, dans d'autres travaux sur le rendu multimodal.

Enfin, ce travail était à l'origine et au coeur du projet CoRSAIRE (Combinaison des Rendus Sensori-moteurs pour l'Analyse Immersive de Résultats) labellisé par l'Agence Nationale de la Recherche en décembre 2005. Dans ce contexte, nous avons participé à l'analyse et à l'élaboration des recommandations multimodales pour les applications de Bio-Informatique et de Mécanique des Fluides. Les interactions et les fonctionnalités en cours de développement par les partenaires nous semblent très prometteuses pour améliorer l'activité des chercheurs.

Ces contributions ont donné lieu à des publications scientifiques dans deux conférences internationales [Bouyer et al., 2007] [Bouyer et al., 2005], un chapitre d'ouvrage [Bourdote et al., 2006b] et une communication nationale [Bouyer, 2004]. Dans le cadre du projet CoRSAIRE, le superviseur multimodal et ses liens avec les applications ont également fait l'objet de plusieurs communications, dont une internationale [Bourdote et al., 2006a].

Perspectives

Les travaux que nous avons présentés sont susceptibles d'être complétés à court terme, et ouvrent en outre de nouvelles pistes de recherche.

Dans un premier temps, il nous faut poursuivre l'amélioration du moteur de supervision, en perfectionnant les fonctionnalités présentées au chapitre 4 (section 4.4) et en ajoutant celles discutées au chapitre 5 (section 5.5). En particulier, les gestions temporelles et spatiales des rendus ainsi que le problème des retraitements et des transitions entre modalités nous semblent être des thèmes essentiels à traiter.

La seconde étape est l'intégration du superviseur multimodal dans une des applications réelles du projet CoRSAIRE. Le choix entre l'exploration d'écoulements et le docking de protéines n'est pas encore établi. Les critères de choix sont à la fois matériels (plateforme d'implémentation, dispositifs disponibles) et fonctionnels (variété des interactions disponibles et des contextes possibles). Cette implémentation servira à valider les spécifications du processus de supervision.

Un travail fondamental est ensuite l'évaluation des interactions supervisées dans l'application choisie. La présence d'ergonomes et d'utilisateurs experts (chercheurs)

sont un avantage indéniable pour aboutir à des interactions multimodales pertinentes et utiles. La conception évolutive du superviseur permettra d'adapter la base de connaissances aux résultats des évaluations. Il sera d'ailleurs intéressant de créer des procédures automatisées d'apprentissage.

A long terme, les intégrations réussies du superviseur multimodal dans quelques applications représentatives des interactions de RV permettront, nous l'espérons, de valider la généralité du modèle et du moteur, et donc d'envisager une implémentation dans de nombreuses autres applications.

Néanmoins, la conception du superviseur n'a pas remplacé et a même un besoin naturel d'études particulières sur les nombreuses projections possibles entre informations et modalités. Il existe donc ici un vaste travail de regroupement et d'analyse des résultats apportés par les communautés de recherches en IHM, RV et psychophysique, afin d'en déduire de nouvelles règles génériques de rendu, respectant notre modèle des interactions. Le problème est (et nous en avons fait l'expérience au début de nos travaux) que la synthèse et la généralisation de résultats particuliers est difficile. En effet, les études ne sont pas forcément reproductibles d'une application à l'autre, et sont souvent inconsciemment biaisées pour mettre en valeur certains résultats. Il est donc probable que l'intégration de rendu multimodaux pertinents dans de nouvelles applications passe par des analyses ergonomiques poussées similaires à celles que nos partenaires du projet CoRSAIRE ont menées.

Toutefois, nous considérons justement notre superviseur multimodal comme un outil puissant permettant de regrouper les évaluations multimodales dans les applications de RV, de pouvoir tester et comparer des rendus différents mais dans des conditions identiques et contrôlées.

En dehors des travaux sur les règles de projection, qui constituent le coeur de la supervision, d'autres problématiques de RV multimodale sont apparues et méritent une attention particulière.

Tout d'abord, les latences des dispositifs de rendu restent un point dur des applications de RV. Si nous entrevoyons des solutions pour les gérer en amont, lors de la supervision, cela ne peut fonctionner que si elles sont connues, ou du moins prévisibles. Une piste de recherche utile serait donc la détermination et la comparaison des latences de différentes modalités, sur les trois canaux principaux. Une autre piste est l'étude approfondie des effets psychophysiques et opératifs de la combinaison de modalités ayant des latences différentes.

Ensuite, certaines modalités, et notamment celles qui réclament un module de calcul de rendu externe, sont particulièrement utiles aux opérateurs, mais complexes à mettre en oeuvre. Nous pensons ici aux différents guides visuels, audio et haptiques, ainsi qu'au calcul de meilleur point de vue pour une tâche et un contexte donnés. Leurs études conjointes permettraient de déterminer les éléments nécessaires à leurs

calculs, et peut-être d'identifier des concepts génériques qui pourraient être intégrés dans le moteur de décision. En parallèle, si les algorithmes requis pour leur réalisation apportaient trop de latence au rendu, il faudrait mener des recherches sur la substitution de ces modalités complexes par des modalités ou des combinaisons de modalités conduisant à un résultat similaire.

Deux autres fonctionnalités très utiles pour certaines applications (e.g. CAO, recherche scientifique) et relativement proches dans leur finalité sont l'annotation et la sauvegarde d'état. Dans les deux cas, il s'agit d'enregistrer ou d'ajouter des informations à l'EV, puis de pouvoir les observer à tout instant ou les retrouver. Cela permet notamment de recréer les conditions dans lesquelles un phénomène est apparu, ou les réflexions qui ont mené à une décision. Les études sur le rendu multimodal de ces annotations et de ces sauvegardes paraissent promises à un bel avenir. En particulier, ces deux fonctionnalités nécessitent une représentation des connaissances, qui pourrait être intégrée dans notre superviseur multimodal.

Les recherches dans le domaine de la présentation d'informations multi-utilisateurs nous intéressent également. L'intégration de la gestion de la coopération multimodale dans le superviseur actuel serait un pas de plus vers sa généricité.

Enfin, une information dont la supervision multimodale est totalement dépendante est la caractérisation de la tâche désirée par l'utilisateur. Dans ce contexte, une piste de recherche porte sur la capacité du gestionnaire d'interactions à décomposer, décrypter en temps réel l'activité de l'utilisateur pour identifier les tâches et les représenter automatiquement selon notre modèle. Le paramètre de la tâche semble le plus difficile à reconnaître, car il est souvent lié aux attentes subjectives, implicites de l'humain, et cette sémantique n'est pas aisément formulable à partir d'indices concrets (comme les informations de tracking ou les données manipulées).

Annexe A : Glossaire des applications scientifiques analysées

Docking de protéines

Ce glossaire est adapté de celui de Nelson [[Nelson, 2006a](#)].

Clustering Classement de sorties de docking suivant différents critères mathématiques : stabilité énergétique, composition de l'interface, RMSD, etc.

Complexe Composé chimique formé de l'association de deux (ou plus) protéines.

Energie Désigne la stabilité d'un complexe calculée par une fonction thermodynamique. Plus l'énergie d'un complexe est faible, plus l'interaction entre les deux partenaires moléculaires est forte.

Forces électrostatiques Phénomènes d'attraction et de répulsion entre des objets selon qu'ils sont respectivement de charge électrique opposée et de même signe.

Hot spot Zone de la protéine dont les méthodes expérimentales suggèrent qu'elle participe fortement à l'interaction moléculaire.

Hydrophobicité (ou hydrophathie) Tendance de certaines zones d'une protéine à se replier sur elles-mêmes sous la forme de "poche" ou de "patch" pour minimiser le contact avec des molécules d'eau.

Interface Zone de contact entre deux protéines.

Protéine Biomolécule constituée d'un enchaînement linéaire non aléatoire d'acides aminés, réalisant des fonctions biologiques très variées dans l'organisme. La fonction d'une protéine est étroitement liée à sa structure dans l'espace.

Résidus Eléments de base d'une protéine, dont l'enchaînement détermine sa structure spatiale et sa fonction biologique. Une vingtaine seulement de résidus participent à la construction de toutes les protéines du corps humain.

RMSD (Root Mean Square Deviation) Indice mathématique de la ressemblance topologique entre deux modèles moléculaires, par exemple une cristallographie et une sortie d'algorithme de docking.

Score Indicateur de la vraisemblance relative des différentes sorties de docking, combinant plusieurs critères numériques (cités plus haut) avec des pondérations différentes choisies par le chercheur.

Solvatation Coût énergétique lié à l'hydratation d'une protéine.

Mécanique des Fluides

Ce glossaire est adapté de celui de Nelson [[Nelson, 2006b](#)].

Isosurface Surface reliant tous les points de l'espace où une variable mathématique prend une valeur donnée.

Facteur Q (hélicité) Paramètre mathématique lié à l'intensité des mouvements de translation subis localement par le fluide, permettant de mesurer l'entraînement du fluide par son propre mouvement. Il est maximal au coeur d'une structure tourbillonnaire.

Mode Paramètre mathématique définissant la topologie d'une structure tourbillonnaire

Protocole de simulation Script informatique permettant la production par un supercalculateur des données d'une simulation numérique.

Reconnexion Collision et pénétration entre deux structures tourbillonnaires avec naissance de structures caractéristiques dans l'écoulement (anneaux tourbillonnaires)

Tourbillon (vortex) Région de l'écoulement subissant des mouvements de rotation importants. Dans la simulation, il se présente sous la forme d'un tube à l'intérieur duquel est concentrée la vorticit .

Vorticité Paramètre mathématique lié à l'intensité des mouvements rotation subis localement par le fluide étudié.

Annexe B : Programmation Prolog

1. Environnement de programmation

L'environnement de programmation de la librairie Amzi Prolog est basé sur un éditeur générique Eclipse (qui permet également de programmer en Java). Il se présente sous une forme classique : une fenêtre principale est dédiée au code, une seconde fenêtre à l'arborescence de fichier et une troisième à l'interpréteur de commandes (figure 5.12). Ce dernier permet d'exécuter des commandes textuelles, soit pour effectuer des requêtes auprès du moteur de règles, soit pour ajouter ou retirer dynamiquement des clauses.

2. Protocole d'évaluation

La première évaluation que nous avons réalisée à la suite de l'implémentation du moteur de décision a consisté à simuler l'ensemble du processus de supervision sur des tâches fictives, depuis la réception de la requête à la communication du résultat, puis la mise à jour des éléments du contexte. Ces tests ont été effectués à l'aide d'une base de connaissances limitée (en particulier, les listes de modalités favorites ne sont pas pondérées), et n'ont pour objectif que de vérifier le bon fonctionnement du processus de décision.

La requête a pour format :

```
?- supervision(idUser, idTache, Resultat).
```

où `supervision` est une règle qui va effectuer l'ensemble des actions du processus de décision : 1) filtres simples, 2) filtres de scores, 3) tri des modalités pondérées, 4) choix de la modalité de score maximal et 5) choix éventuel d'autres modalités de scores proches du maximal.

```
supervision(User, Tache, Resultat) :-  
    basics(User, Tache, Lbasics),           %1  
    maj_score_list(Lbasics, Lscore, Tache), %2  
    tri_insertion(Lscore, Ltri),           %3
```

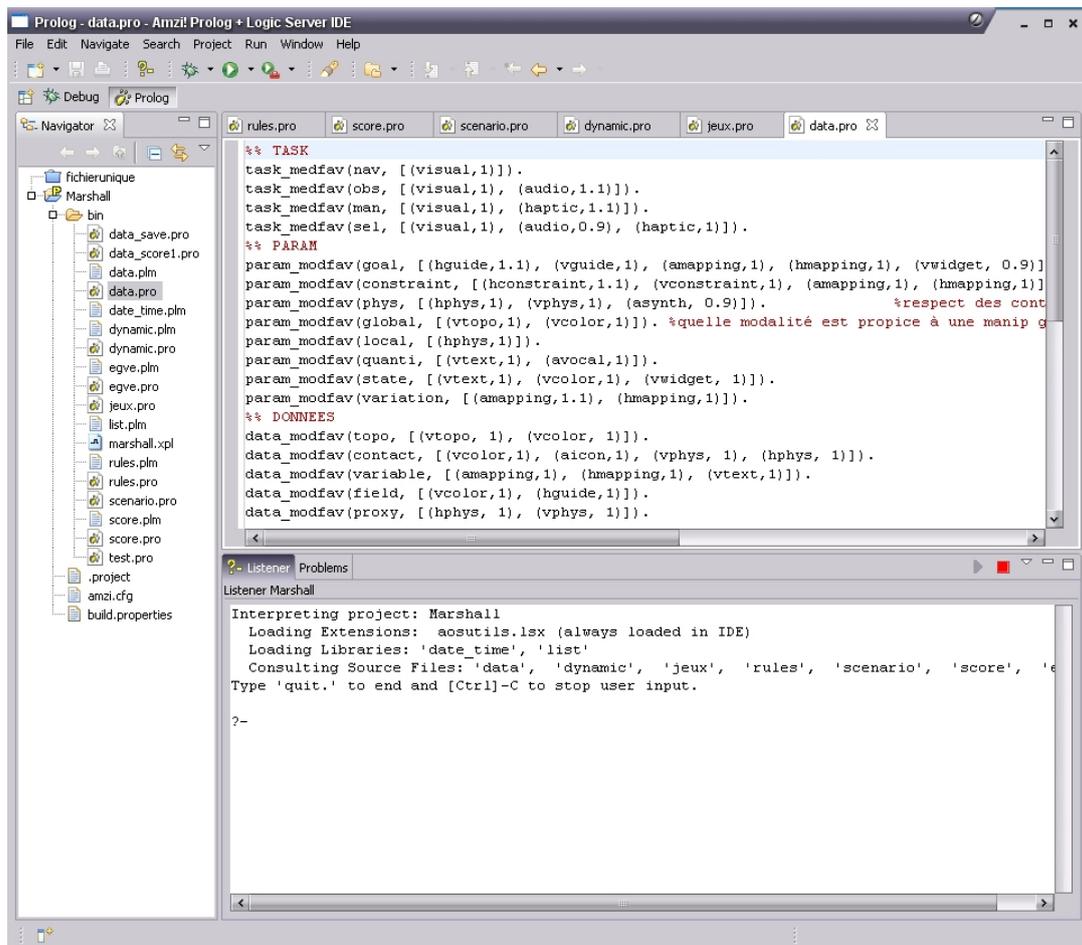


FIG. 5.12 – Editeur Prolog.

```

write(Ltri),nl,
get_max(Ltri, Mod, X),                %4
select_max_list(Ltri, Resultat, X).   %5

```

Cette règle est le coeur du moteur et va être vérifiée chaque fois que l'application effectuera une requête au superviseur.

La suite ne sera pas utilisée dans le processus final de supervision (au sein d'une application) mais a servi pour l'évaluation du moteur de décision. Nous avons simulé les actions de base de l'application : lancement de la requête, réception du résultat, mise à jour de la base de contexte (ajout de la nouvelle interaction et mise à jour des canaux) puis rendu effectif.

```

start(User, Tache) :-
    write('Start : '), tab(2), write(User), tab(2),

```

```

    write(Tache), tab(2),nl,
    supervision(User, Tache, Resultat),
    assert(current(User, Tache, Resultat)),
    maj_channel(Lmod, 1),
    write('-> '), write(Lmod), tab(2), write(X), nl.

```

Une autre règle permet de stopper une interaction si elle était en cours, et de mettre à jour le contexte :

```

stop(User, Tache) :-
    current(User, Tache, Resultat),!,
    retract(current(User, Tache, _)),
    maj_channel(Lmod, -1),
    write('Stop :'), tab(2), write(User),
    tab(2), write(Inter), nl, nl.

```

Enfin des protocoles ont été mis au point afin de tester le moteur de décision dans différents cas. Par exemple, le protocole suivant crée deux utilisateurs, dont un aveugle. Trois médias sont techniquement disponibles. La scène contient deux objets. Les tâches sont fictives. Trois tests sont réalisés : 1) les tâches sont réalisées par un utilisateur expert sans handicap, et tous les médias sont disponibles ; 2) il n'y a pas de dispositif haptique ; 3) l'utilisateur est aveugle et tous les médias sont disponibles.

Le code correspondant à ce protocole est le suivant :

```

%UTILISATEURS
user(normal, [visual, audio, haptic], 3).
user(aveugle, [audio, haptic], 1).

%CANAU
channel(visual, 0).
channel(audio, 0).
channel(haptic, 0).

%OBJETS
obj(x, [visual, audio, haptic]).
obj(y, [visual, audio, haptic]).

%TACHES
inter(1, true, 3).
task(1, man).
data(1, proxy, [x]).

```

```
param(1, constraint).

inter(2, true, 1).
task(2, man).
data(2, topo, [x, y]).
param(2, phys).

inter(3, false, 2).
task(3, obs).
data(3, variable, []).
param(3, goal).

%TEST1 : utilisateur normal et contexte idéal
test1 :-
    start(normal, 1),
    stop(normal, 1),
    start(normal, 2),
    stop(normal, 2),
    start(normal, 3),
    stop(normal, 3).

%TEST2 : utilisateur normal et absence d'haptique
test2 :-
    retract(channel(haptic, _)),
    start(normal, 1),
    stop(normal, 1),
    start(normal, 2),
    stop(normal, 2),
    start(normal, 3),
    stop(normal, 3).

%TEST3 : utilisateur aveugle et contexte idéal
test3 :-
    start(aveugle, 1),
    stop(aveugle, 1),
    start(aveugle, 2),
    stop(aveugle, 2),
    start(aveugle, 3),
    stop(aveugle, 3).
```

Les résultats des tests sont présentés dans les figures suivantes. Un utilisateur

expert dans des conditions idéale a à sa disposition de nombreuses modalités (figure 5.13).

```
?- test1.

Start : normal 1
[(hphys , 2), (hconstraint , 2), (hmapping , 2), (vphys , 2), (vconstraint ,
2), (hguide , 1), (amapping , 1), (vguide , 1), (vcolor , 1), (vwidget , 1),
(vtext , 1), (vtopo , 1), (asynth , 0), (avocal , 0), (aicon , 0)]
-> [hphys, hconstraint, hmapping, vphys, vconstraint] 2
Stop : normal 1

Start : normal 2
[(hphys , 2), (vphys , 2), (vcolor , 2), (vtopo , 2), (hconstraint , 1),
(hguide , 1), (hmapping , 1), (asynth , 1), (vconstraint , 1), (vguide , 1),
(vwidget , 1), (vtext , 1), (avocal , 0), (amapping , 0), (aicon , 0)]
-> [hphys, vphys, vcolor, vtopo] 2
Stop : normal 2

Start : normal 3
[(amapping , 3), (hmapping , 2), (vguide , 2), (vwidget , 2), (vtext , 2),
(hguide , 1), (asynth , 1), (avocal , 1), (aicon , 1), (vphys , 1),
(vconstraint , 1), (vcolor , 1), (vtopo , 1), (hphys , 0), (hconstraint , 0)]
-> [amapping] 3
Stop : normal 3
```

FIG. 5.13 – Résultats du test n°1 : utilisateur expert et contexte idéal.

Dans un contexte sans interface haptique (figure 5.14), les modalités de rendu haptiques décidées dans le contexte idéal ont disparu. C'est le cas des tâches n°1 et n°2 qui n'emploient désormais que des modalités visuelles. La tâche n°3 est inchangée.

```
?- test2.

Start : normal 1
[(vphys , 2), (vconstraint , 2), (amapping , 1), (vguide , 1), (vcolor , 1),
(vwidget , 1), (vtext , 1), (vtopo , 1), (asynth , 0), (avocal , 0), (aicon ,
0)]
-> [vphys, vconstraint] 2
Stop : normal 1

Start : normal 2
[(vphys , 2), (vcolor , 2), (vtopo , 2), (asynth , 1), (vconstraint , 1),
(vguide , 1), (vwidget , 1), (vtext , 1), (avocal , 0), (amapping , 0), (aicon
, 0)]
-> [vphys, vcolor, vtopo] 2
Stop : normal 2

Start : normal 3
[(amapping , 3), (vguide , 2), (vwidget , 2), (vtext , 2), (asynth , 1),
(avocal , 1), (aicon , 1), (vphys , 1), (vconstraint , 1), (vcolor , 1), (vtopo
, 1)]
-> [amapping] 3
Stop : normal 3
```

FIG. 5.14 – Résultats du test n°2 : utilisateur expert et contexte sans haptique.

Si l'utilisateur est aveugle (figure 5.15), les modalités de rendu visuelles décidées lors du test n°1 ont disparu. C'est le cas des tâches n°1 et n°2 qui n'emploient désormais que des modalités haptiques. La tâche n°3 est inchangée.

```
?- test3.  
  
Start : aveugle 1  
[(hphys , 2), (hconstraint , 2), (hmapping , 2), (hguide , 1), (amapping , 1),  
(asynth , 0), (avocal , 0), (aicon , 0)]  
-> [hphys, hconstraint, hmapping] 2  
Stop : aveugle 1  
  
Start : aveugle 2  
[(hphys , 2), (hconstraint , 1), (hguide , 1), (hmapping , 1), (asynth , 1),  
(avocal , 0), (amapping , 0), (aicon , 0)]  
-> [hphys] 2  
Stop : aveugle 2  
  
Start : aveugle 3  
[(amapping , 3), (hmapping , 2), (hguide , 1), (asynth , 1), (avocal , 1),  
(aicon , 1), (hphys , 0), (hconstraint , 0)]  
-> [amapping] 3  
Stop : aveugle 3
```

FIG. 5.15 – Résultats du test n°3 : utilisateur aveugle et contexte idéal.

Table des figures

1.1	La RV : Interaction Humain - Environnement	8
1.2	Approche Orientée Perception : Automobile	9
1.3	Approche Orientée Perception : Jeux Video	10
1.4	Approche Orientée Tâche : 3DVDM	11
1.5	Approche Orientée Tâche : MécaFlux	11
1.6	Continuum Réel-Virtuel de Milgram	12
1.7	Continuum Perception-Tâche	12
1.8	RV : Définitions	14
1.9	Nesbitt : Exemples de modalités	16
1.10	Double boucle de conception	18
2.1	Dispositifs immersifs (CAVE)	24
2.2	Dispositifs immersifs (Workbench)	24
2.3	Visiocasque	25
2.4	HRTF	26
2.5	Dispositifs tactiles	27
2.6	Dispositifs haptiques 1	28
2.7	Dispositifs haptiques 2	28
2.8	Théorie des Spécificités Modales	30
2.9	Théorie des Spécificités Modales modifiée	30
2.10	Multimodalité pour l'apprentissage : ScienceSpace	35
2.11	Multimodalité pour l'apprentissage	36
2.12	Sonification de données	39
2.13	Haptisation de données	42
2.14	Multimodalité pour la Réalité Augmentée	45
2.15	Répartitions des applications multi-sensorielles existantes	47
2.16	L'IHM de MIAMI	50
2.17	Taxonomie des modalités de Bernsen	51
2.18	Espace de conception de la multimodalité en sortie	52
2.19	Schémas et aspects de composition des modalités	53
2.20	Illusions visuelles 2	61

2.21	Illusions visuelles 1	61
2.22	Illusions visuelles 3	62
2.23	Le pseudo-haptique	64
3.1	Logiciels de visualisation moléculaire	71
3.2	Logiciels de visualisation moléculaire immersive	72
3.3	Logiciels de visualisation d'écoulement	74
3.4	Exemples de visualisation en MF	75
3.5	Exemples de visualisation en MF	76
3.6	Exemples d'interaction en MF	77
3.7	Logiciel de CAO (CATIA)	78
3.8	Interaction VRAD 1	79
3.9	Interaction VRAD 2	81
3.10	Exemple de résultat d'ACD	83
3.11	Processus de docking classique vs. paradigme de docking RV	85
3.12	Recommandations de rendu en fonction des caractéristiques des données de Docking	87
3.13	Distributions multimodales possibles de certains critères de docking	88
3.14	Distributions multimodales possibles de certaines données de mécanique	91
4.1	Rôles et retombées de la supervision multimodale	96
4.2	Modèle de la tâche en RV	97
4.3	Modèle du rendu multimodal en RV	107
4.4	Exemples de modélisations d'interactions concrètes	109
4.5	Exemples d'interactions concrètes en Mécanique des Fluides	110
4.6	Exemples d'interactions concrètes en Bio-Informatique	110
4.7	Exemples d'interactions concrètes en CAO	111
4.8	Architecture générale de la supervision multimodale	114
4.9	Processus de supervision	116
4.10	Principe du processus de décision	122
4.11	Processus de décision détaillé	125
4.12	Exemple de choix de présentation spatiale en CAO	131
4.13	Exemple de choix de présentation spatiale en MF	131
4.14	Exemples de synchronisation temporelle	132
5.1	Simulateur de test de supervision multimodale	141
5.2	Graphe du corpus d'interactions possibles	143
5.3	Tableau des interactions élémentaires pour les listes	144
5.4	Tableau des modalités favorites par éléments	145
5.5	Tableau des modalités favorites par éléments : exemple	146
5.6	Tableau des modalités favorites pondérées par éléments	147

5.7 Interactions élémentaires pour l'évaluation : exemples	148
5.8 Résultats de l'évaluation sur des interactions élémentaires	148
5.9 Application de Docking de Protéines	150
5.10 Résultats de l'évaluation sur le scénario de Bio-Informatique	153
5.11 Résultats de l'évaluation sur le scénario de Bio-Informatique (sans haptique)	154
5.12 Editeur Prolog	170
5.13 Résultats du test n°1	173
5.14 Résultats du test n°2	173
5.15 Résultats du test n°3	174

Bibliographie

- [Allen, 1983] Allen, J. F. (1983). *Maintaining Knowledge about Temporal Intervals*, volume 26 of *Communications of the ACM*, chapter 11, pages 832–843.
- [Ammi, 2005] Ammi, M. (2005). *Interface Homme-Machine Multimodale pour la Télé-Micromanipulation*. PhD thesis, Université d'Orléans.
- [André, 2000] André, E. (2000). *Handbook of natural language processing : techniques and applications for the processing of language as text*, chapter The generation of multimedia presentations, pages 305–327. Marcel Dekker.
- [Arens and Hovy, 1995] Arens, Y. and Hovy, E. (1995). The design of a model-based multimedia interaction manager. *Artificial Intelligence Review*, 9(3) :167–188.
- [Barass and Zehner, 2000] Barass, S. and Zehner, B. (2000). Responsive sonification of well-logs. In *International Conference on Auditory Display*, Atlanta, Georgia, USA.
- [Bayle, 2003] Bayle, B. (2003). Perception haptique, séminaire equipe avr, 8 juillet 2003.
- [Bellik, 1995] Bellik, Y. (1995). *Interfaces multimodales : concepts, modèles et architectures*. Thèse de doctorat en informatique, Université de Paris-XI.
- [Bernsen, 1994] Bernsen, N. O. (1994). Foundations of multimodal representations. a taxonomy of representational modalities. *Interacting with Computers*, 6(4) :347–371.
- [Bisseret et al., 1999] Bisseret, A., Sébillotte, S., and Falzon, P. (1999). *Techniques pratiques pour l'étude des activités expertes*. Toulouse : Octarès.
- [Bourdot et al., 2006a] Bourdot, P., Bouyer, G., Fauvet, N., Martin, C., Vézien, J.-M., Ammi, M., and Gherbi, R. (2006a). Corsaire : Combination of sensorimotor renderings for immersive analysis of results. Poster at 3rd INTUITION Workshop VR/VE & Industry - Challenges and opportunities. Schwabenlandhalle, Fellbach/Stuttgart, Germany, 30th of November - 1st of December 2006.
- [Bourdot and Touraine, 2002] Bourdot, P. and Touraine, D. (2002). Polyvalent display framework to control virtual navigation by 6dof tracking. In *IEEE Virtual Reality Conference 2002. Orlando, Florida, USA*.

- [Bourdot et al., 2006b] Bourdot, P., Touraine, D., Vézien, J.-M., and Bouyer, G. (2006b). Les techniques d'intégration multimodale en environnement virtuel. In *Traité de la Réalité Virtuelle, Volume 3 : Outils et Modèles Informatiques des Environnements Virtuels*, chapter 13. Les Presses de l'École des Mines de Paris.
- [Bouyer, 2004] Bouyer, G. (2004). Des feedbacks sensori-moteurs vers le rendu multimodal. Séminaire de l'AS Interface et Information Haptiques, Grenoble, 10 juin 2004.
- [Bouyer et al., 2007] Bouyer, G., Bourdot, P., and Ammi, M. (2007). Supervision of task-oriented multimodal rendering for vr applications. In *Eurographics Symposium on Virtual Environments (EGVE), 15-18th July 2007, Weimar, Germany*.
- [Bouyer et al., 2005] Bouyer, G., Bourdot, P., and Vézien, J.-M. (2005). Totem : An engine for task-oriented multimodal rendering in virtual environments. In *Proceedings of the HCI International Conference (HCII), Vol. 9 : Advances in Virtual Environments Technology : Musings on Design, Evaluation, & Applications. Las Vegas, USA, July 2005, CD-ROM. (MIRA DIGITAL PUBLISHING)*.
- [Bresciani et al., 2004] Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., and Kheddar, A. (2004). Feeling what you hear : auditory signals can modulate tactile taps perception. *Experimental Brain Research (submitted)*.
- [Burkhardt et al., 2003] Burkhardt, J.-M., Bardy, B., and Lourdeaux, D. (2003). Immersion, réalisme et présence dans la conception et l'évaluation des environnements virtuels. *Psychologie française : Recherches en psychologie ergonomique*, 48(2).
- [Burkhardt et al., 2006] Burkhardt, J.-M., Lourdeaux, D., and Mellet-d'Huart, D. (2006). Traité de la réalité virtuelle, volume 4 : Les applications de la réalité virtuelle. In *La Réalité Virtuelle pour l'Apprentissage Humain*. Les Presses de l'École des Mines de Paris.
- [Chaari and Laforest, 2005] Chaari, T. and Laforest, F. (2005). L'adaptation dans les systèmes d'information sensibles au contexte d'utilisation : approche et modèles. Cinquièmes journées scientifiques des jeunes chercheurs en génie électrique et informatique (GEI 2005), du 25 au 27 Mars 2005, Sousse, Tunisie.
- [Chaari et al., 2006] Chaari, T., Laforest, F., and Celantano, A. (2006). Adaptation in context-aware pervasive information systems. *International Journal of Pervasive Computing and Communications*, 3/2006.
- [Childs, 2001] Childs, E. (2001). The sonification of numerical fluid flow simulations. In *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1*.
- [Childs, 2002] Childs, E. (2002). The sonification of numerical fluid flow simulations. partial fulfillment for the degree of master of arts in electro-acoustic music, Dartmouth College.

- [Convard and Bourdot, 2003] Convard, T. and Bourdot, P. (2003). A multimodal approach for computer aided design in immersive environments. In *Virtual Concept. Biarritz (France)*.
- [Conversy, 2000] Conversy, S. (2000). *Conception d'icônes auditives paramétrées pour les interfaces homme-machine*. PhD thesis, Université Paris-Sud XI.
- [Conversy and Beaudouin-Lafon, 1996] Conversy, S. and Beaudouin-Lafon, M. (1996). Le son dans les applications interactives. *Nouvelles Interfaces Homme-Machine, Groupe de travail OFTA*, 18 :65–81.
- [Coutaz et al., 1995] Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., and Young, R. (1995). Four easy pieces for assessing the usability of multimodal interaction : The care properties. In *Interact'95*, pages 115–120, Lillehammer, Norway.
- [Crison et al., 2005] Crison, F., Lécuyer, A., Mellet-D'Huart, D., Burkhardt, J., Michel, G., and Dautin, J. (2005). Virtual technical trainer : Training to use milling machines with multi-sensory feedback in virtual reality. In *Proceedings of the IEEE International Conference on Virtual Reality*.
- [Dalal et al., 1996] Dalal, M., Feiner, S., McKeown, K., Pan, S., Zhou, M., Höllerer, T., Shaw, J., Feng, Y., and Fromer, J. (1996). Negotiation for automated generation of temporal multimedia presentations. In *ACM Multimedia'96, Boston, USA, 18-22 November 1996*, pages 55–64.
- [Day, 1990] Day, R. (1990). The bourdon illusion in haptic space. *Perception and Psychophysics*.
- [de Campo et al., 2005] de Campo, A., Frauenberger, C., and Höldrich, R. (2005). Sonenvir - a progress report. In *ICMC 2005 International Computer Music Conference, Barcelona, September 5-9*.
- [de Cheveigné, 2003] de Cheveigné, A. (2003). Espace et son. Colloque Espaces de l'Homme, Collège de France, 14-15 Octobre 2003.
- [Dede et al., 1999] Dede, C., Salzman, M., Loftin, B., and Sprague, D. (1999). Multisensory immersion as a modeling environment for learning complex scientific concepts. *Nancy Roberts, Wallace Feurzeig, and Beverly Hunter, Computer modeling and simulation in science education*. New York : Springer-Verlag.
- [Dey and Abowd, 2000] Dey, A. and Abowd, G. (2000). Towards a better understanding of context and context-awareness. CHI 2000, Workshop on the What, Who, Where, When, and How of Context-Awareness.
- [Dey et al., 2002] Dey, A., Salber, D., and Abowd, G. (2002). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Context-Aware Computing : A Special Triple Issue of Human-Computer Interaction*. Lawrence-Erlbaum, March 2002.

- [Didier, 2005] Didier, J.-Y. (2005). *Contributions à la dextérité d'un système de réalité augmentée mobile appliqué à la maintenance industrielle*. PhD thesis, Université d'Evry Val d'Essonne, Evry.
- [Dragicevic, 2004] Dragicevic, P. (2004). *Un modèle d'interaction en entrée pour des systèmes interactifs multi-dispositifs hautement configurables*. PhD thesis, Université de Nantes.
- [Drif et al., 2005] Drif, A., Amouri, A., Séguy, N., and Kheddar, A. (2005). The muhaptic, an inclusive haptic interface. In *International Conference on Mechatronics and Automation (ICMA 2005)*. IEEE.
- [Dunn and Clark, 1999] Dunn, J. and Clark, M. A. (1999). "life music" : The sonification of proteins. *Leonardo*, 32(1) :25–32.
- [Dutoit et al., 2001] Dutoit, A., Crighton, O., Klinker, G., Kobylinski, R., Vilsmeier, C., and Bruegge, B. (2001). Architectural issues in mobile augmented reality systems : a prototyping case study. In *ASPEC 2001, Macau*.
- [Eckel, 1998] Eckel, G. (1998). A spatial auditory display for the cyberstage. In *International Conference on Auditory Display*, Glasgow, UK,.
- [Fitch and Kramer, 1994] Fitch, W. T. and Kramer, G. (1994). *Sonifying the Body Electric : Superiority of an Auditory over a Visual Display in a Complex, Multivariate System*, chapter XVIII, pages 307–326. G. Kramer, Addison-Wesley Publishing Company.
- [Foley et al., 1990] Foley, J., Dam, A. V., Feiner, S., and Hughes, J. (1990). *Computer graphics Principles and Practice*. Addison-Wesley, 2e édition.
- [Friedes, 1974] Friedes, D. (1974). Human information processing and sensory modality : Cross-modal functions, information complexity, memory and deficit. *Psychological Bulletin*, 81(5) :284–310.
- [Fritz, 1996] Fritz, J. P. (1996). *Haptic Rendering Techniques for Scientific Visualization*. PhD thesis, University of Delaware (Electrical Engineering).
- [Fuchs et al., 2006a] Fuchs, P., Berthoz, A., and Vercher, J.-L. (2006a). Introduction à la réalité virtuelle. In *Traité de la Réalité Virtuelle, Volume 1 : L'Homme et l'Environnement Virtuel*. Les Presses de l'École des Mines de Paris.
- [Fuchs et al., 2006b] Fuchs, P., Hafez, M., Benali-Koudja, M., Papin, J., Tsingos, N., and Warusfel, O. (2006b). Les sens de l'homme. In *Traité de la Réalité Virtuelle, Volume 1 : L'Homme et l'Environnement Virtuel*. Les Presses de l'École des Mines de Paris.
- [Fuchs and Moreau, 2004] Fuchs, P. and Moreau, G. (2004). *Le Traité de la Réalité Virtuelle, Volume 1 : Fondements et interfaces comportementales*. Les Presses de l'École des Mines de Paris, deuxième édition.

- [Fuchs et al., 2006c] Fuchs, P., Moreau, G., Berthoz, A., and Vercher, J.-L. (2006c). *Traité de la Réalité Virtuelle, Volume 1 : L'Homme et l'Environnement Virtuel*. Les Presses de l'École des Mines de Paris.
- [Garcia-Ruiz et al., 2004] Garcia-Ruiz, M., Bustos-Mendoza, C., Galeana-de la O, L., Andrade-Arechiga, M., Santos-Virgen, M., and Acosta-Diaz, R. (2004). Exploring multimodal virtual environments for learning biochemistry concepts. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 2143–2147. Chesapeake, VA : AACE.
- [Gaver, 1989] Gaver, W. W. (1989). The sonicfinder : An interface that uses auditory icons. In *CHI'89 Conference on Human Factors in Computing Systems*, volume 4, pages 67–94.
- [Gentaz and Hatwell, 2004] Gentaz, E. and Hatwell, Y. (2004). Role of exploratory movements in the müller-lyer, vertical-horizontal and delboeuf haptic perceptual illusions. *Psychonomic Bulletin and Review*, 11(1) :31–40.
- [Gherbi et al., 2006] Gherbi, R., Bourdot, P., Vézien, J.-M., and Hérisson, J. (2006). *Traité de la réalité virtuelle, volume 4 : Les applications de la réalité virtuelle. In Explorations de données scientifiques et expérimentations virtuelles*. Les Presses de l'École des Mines de Paris.
- [Ghiglione et al., 1998] Ghiglione, R., Landré, A., Bromberg, M., and Molette, P. (1998). L'analyse cognitivo-discursive. In *L'analyse automatique des contenus*. Paris : Dunod.
- [Gigante, 1993] Gigante, M. A. (1993). Virtual reality : Definitions, history and applications. In R. A. Earnshaw, M. A. Gigante and H. Jones, editors, *Virtual Reality Systems*, chapter 1. Academic press.
- [Gilfix and Couch, 2000] Gilfix, M. and Couch, A. L. (2000). Peep (the network auralizer) : Monitoring your network with sound. In *LISA '00 : Proceedings of the 14th USENIX conference on System administration*, pages 109–118, Berkeley, CA, USA. USENIX Association.
- [Go and Carroll, 2003] Go, K. and Carroll, J.-M. (2003). Scenario-based task analysis. In D Diaper, N. S., editor, *The handbook of task analysis for human computer interaction*, pages 83–116. Mahwah, NJ : Lawrence Erlbaum Associates.
- [Gosline et al., 2001] Gosline, A. H., Turgay, E., and Brouwer, I. (2001). Haptic illusions : What you feel isn't always what you get. In *Human Interface Technologies 2001*, pages 19–22.
- [Green, 1997] Green, D., F. (1997). Texture sensing and simulation using the phantom : Towards remote sensing of soil properties. In *The Second PHANTOM User's Group Workshop*, Cambridge, Massachusetts, USA, MIT.

- [Green and Salsibury, 1998] Green, D., F. and Salsibury, J. K. (1998). Soil simulation with a phantom. In *The Third PHANToM User's Group Workshop*, Cambridge, Massachusetts, USA, MIT.
- [Guest et al., 2002] Guest, S., Catmur, C., Lloyd, D., and Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, 146 :161–171.
- [Hansen et al., 1999] Hansen, M., Charp, E., Lodha, S., Meads, D., and Pang, A. (1999). Promuse : A system for multi-media data presentation of protein structural alignments. In *Pacific Symposium on Biocomputing*, pages 368–379.
- [Hansen et al., 2000] Hansen, M. D., Lodha, S., , and Pang, A. (2000). Profeel : Low cost visual-haptic perceptualization of protein structure-structure alignments. In *Proceedings of Pacific Symposium on Biocomputing*, volume 5, pages 215–226.
- [Harvey and Gingold, 2000] Harvey, E. and Gingold, C. (2000). Haptic representation of the atom. In *IV '00 : Proceedings of the International Conference on Information Visualisation*, page 232, Washington, DC, USA. IEEE Computer Society.
- [Hayward, 1994] Hayward, C. (1994). *Listening to the Earth Sing*, chapter XVIII, pages 369–404. G. Kramer, Addison-Wesley Publishing Company.
- [Heller, 1982] Heller, M. A. (1982). Visual and tactual texture perception : Inter-sensory cooperation. *Perception & Psychophysics*, 31 :339–344.
- [Hermann et al., 2000] Hermann, T., Nattkemper, T., Ritter, H., and Schubert, W. (2000). Sonification of multi-channel image data.
- [Hérisson et al., 2004] Hérisson, J., Gros, P.-E., Férey, N., Magneau, O., and Gherbi, R. (2004). Dna in virtuo : Visualization and virtual manipulation of 3d genomic structures. In *19th International CODATA Conference, The Information Society : New Horizons for Science, 7-10 november 2004, Berlin, Germany*.
- [Infed et al., 1999] Infed, F., Brown, S. W., Lee, C. D., Lawrence, D. A., Dougherty, A. M., and Pao, L. Y. (1999). Combined visual/haptic rendering modes for scientific visualization. In *Proc. ASME Dynamic Systems and Control Division*, DSC-Vol. 67, pages 93–99, Nashville, TN. Int. Mech. Engr. Cong. & Expo.
- [Jackson and Francioni, 1994] Jackson, J. A. and Francioni, J. M. (1994). *Synchronization of Visual and Aural Parallel Program Performance Data*, chapter XVIII, pages 291–306. G. Kramer, Addison-Wesley Publishing Company.
- [Jameson, 1994] Jameson, D. H. (1994). *Sonnet : Audio-Enhanced Monitoring and Debugging*, chapter XVIII, pages 253–266. G. Kramer, Addison-Wesley Publishing Company.
- [Jones et al., 2004] Jones, M. G., Andre, T., Kubasko, D., Botkinsky, A., Tretter, T., Negishi, A., Taylor, R., and Superfine, R. (2004). Remote atomic force microscopy

- of microscopic organisms : technological innovations for hands-on science with middle and high school students. *Science Education*, 88 (1) :55–71.
- [Kamp et al., 2004] Kamp, J.-F., Poirier, F., and Doignon, P. (2004). Interaction with in-vehicle electronic systems : A complete description of a neural network approach. *Neural Processing Letters*, 19 :109–129.
- [Katz et al., 2007] Katz, B., Tarault, A., Bourdot, P., and Vézien, J.-M. (2007). The use of 3d-audio in a multi-modal teleoperation platform for remote driving/supervision. In *Proceedings of the AES 30th International Conference, Saariselkä, Finland*.
- [Khézami, 2005] Khézami, N. (2005). *Un Système Multi-Agent pour l'analyse, la conception et l'évaluation de la collaboration, appliquée au Télétravail Collaboratif via Internet*. PhD thesis, Université d'Evry Val d'Essonne.
- [Kintsch and Dijk, 1978] Kintsch, W. and Dijk, T. V. (1978). Toward a model of text comprehension and production. *Psychol Rev*, 85 (5) :363–394.
- [Kramer, 1994] Kramer, G. (1994). *An Introduction to Auditory Display*, chapter XVIII, pages 1–78. G. Kramer, Addison-Wesley Publishing Company.
- [Lécuyer et al., 2000] Lécuyer, A., Coquillart, S., Kheddar, A., Richard, P., and Coiffet, P. (2000). Pseudo-haptic feedback : Can isometric input devices simulate force feedback ? IEEE Int. Conf. on Virtual Reality, pages 83-90, New Brunswick, US.
- [Lumsden et al., 2002] Lumsden, J., Brewster, S., Crease, M., and Gray, P. (2002). Guidelines for audio-enhancement of graphical user interface widgets. In BCS, editor, *Proceedings of British HCI 2002 (London, UK)*, volume 2, pages 6–9.
- [Lumsden and Gammell, 2004] Lumsden, J. and Gammell, A. (2004). Mobile note taking : Investigating the efficacy of mobile text entry. In *The 6th International Conference on Human Computer Interaction with Mobile Devices and Services*, Glasgow, Écosse.
- [Lundin et al., 2005] Lundin, K. E., Sillen, M., Cooper, M. D., and Ynnerman, A. (2005). Haptic visualization of computational fluid dynamics data using reactive forces. In Erbacher, R. F., Roberts, J. C., Gröhn, M. T., and Börner, K., editors, *Visualization and Data Analysis 2005. Proceedings of the SPIE, Volume 5669, pp. 31-41 (2005)*., volume 5669 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 31–41.
- [MacDonald and Vince, 1994] MacDonald, L. and Vince, J. (1994). *Interacting with Virtual Environments*. John Wiley & Sons.
- [Maciejewski et al., 2005] Maciejewski, R., Choi, S., Ebert, D. S., and Tan, H. Z. (2005). Multi-modal perceptualization of volumetric data and its application to molecular docking. In *WHC '05 : Proceedings of the First Joint Eurohaptics*

- Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 511–514, Washington, DC, USA. IEEE Computer Society.
- [Malinowski, 2000] Malinowski, S. (2000). Audification/sonification, url <http://www.musanim.com/mam/audifica.htm>.
- [McCabe and Rangwalla, 1994] McCabe, K. and Rangwalla, A. (1994). *Auditory Display of Computational Fluid Dynamics Data*, chapter XVIII, pages 327–340. G. Kramer, Addison-Wesley Publishing Company.
- [McGee et al., 2000] McGee, M. R., Gray, P. D., and Brewster, S. A. (2000). The effective combination of haptic and auditory textural information. In *First Workshop on Haptic Human-Computer Interaction*, pages 33–38.
- [McGee et al., 2001] McGee, M. R., Gray, P. D., and Brewster, S. A. (2001). Haptic perception of virtual roughness. In *ACM CHI 2001*, pages 155–156, Seattle, WA. ACM Press Addison-Wesley.
- [McGurk and MacDonald, 1976] McGurk, H. and MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature*, 264 :746–748.
- [McLaughlin and Orenstein, 1997] McLaughlin, J. P. and Orenstein, B. J. (1997). Haptic rendering of 3d seismic data. In *The Second PHANToM User's Group Workshop*, Cambridge, Massachusetts, USA, MIT.
- [Milgram et al., 1995] Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality : a class of displays on the reality-virtuality continuum. In Das, H., editor, *Proc. SPIE Vol. 2351, p. 282-292, Telemanipulator and Telepresence Technologies, Hari Das ; Ed.*, volume 2351 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 282–292.
- [Minogue et al., 2006] Minogue, J., Jones, M. G., Broadwell, B., and Oppewall, T. (2006). The impact of haptic augmentation on middle school students' conceptions of the animal cell. *Virtual Reality*, 10(3) :293–305.
- [Nelson, 2006a] Nelson, J. (2006a). Analyse de l'activité de docking moléculaire et recommandations pour la conception ergonomique d'un premier démonstrateur corsaire. Technical report, Université René Descartes Paris 5, Unité EA 4070 : Ergonomie, Comportement et Interactions.
- [Nelson, 2006b] Nelson, J. (2006b). Analyse de l'activité d'exploration de simulations numériques d'écoulements en dynamique des fluides et recommandations pour la conception ergonomique d'un démonstrateur corsaire. Technical report, Université René Descartes Paris 5, Unité EA 4070 : Ergonomie, Comportement et Interactions.
- [Nelson, 2006c] Nelson, J. (2006c). Conception centrée utilisateur d'environnements virtuels multimodaux pour l'exploration de données scientifiques. Master's thesis,

- Université Paris 5, Master Sciences de la Vie et de la Santé, Ingénierie Biomédicale, Spécialité Ergonomie.
- [Nesbitt, 2003] Nesbitt, K. V. (2003). *Multi-sensory Display of Abstract Data*. PhD thesis, School of Information Technology, University of Sydney, Australia.
- [Nigay, 2001] Nigay, L. (2001). *Modalité d'interaction et Multimodalité, Thèse Habilitation à Diriger des Recherches*. PhD thesis, Laboratoire de Communication Langagière et Interaction Personne-Système (IMAG), Université Joseph Fourier.
- [Nigay, 2005] Nigay, L. (2005). Espaces de conception. Cours de Master Mathématiques Informatique, Spécialité : 3I Intelligence, Interaction et Information, Module DI : Interaction multimodale.
- [Nigay and Coutaz, 1996] Nigay, L. and Coutaz, J. (1996). Espaces conceptuels pour l'interaction multimédia et multimodale. *TSI, spécial Multimédia et Collecticiel*, 15(9) :1195–1225.
- [Oakley et al., 2002] Oakley, I., Adams, A., Brewster, S., and Gray, P. (2002). Guidelines for the design of haptic widgets. In Springer, editor, *Proceedings of BCS HCI 2002 (London, UK)*, pages 195–212.
- [Ochanine, 1992] Ochanine, D.-A. (1992). The operative image of controlled objects in man-automatic machine systems. In Leplat, J., editor, *L'analyse du travail en psychologie ergonomique*, volume 1, pages 99–106. Toulouse : Octarès.
- [Oviatt, 2006] Oviatt, S. (2006). Human-centered design meets cognitive load theory : designing interfaces that help people think. In *MULTIMEDIA '06 : Proceedings of the 14th annual ACM international conference on Multimedia, Santa Barbara, CA, USA*, pages 871–880, New York, NY, USA. ACM Press.
- [Paljic et al., 2004] Paljic, A., Burkhardt, J.-M., and Coquillart, S. (2004). Evaluation of pseudo-haptic feedback for simulating torque : A comparison between isometric and elastic input devices. In *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2004), Chicago, IL, USA*.
- [Picon et al., 2007] Picon, F., Bourdot, P., and Ammi, M. (2007). Haptic selection for computing aided design. 10th IFF Science Days, June 27-28 2007, Magdeburg, Germany.
- [Plemenos and Benayada, 1996] Plemenos, D. and Benayada, M. (1996). Intelligent display in scene modeling. new techniques to automatically compute good views. In *International Conference GraphiCon'96, St Petersburg (Russia), 1 - 5 of July 1996*.
- [Rabenhorst et al., 1990] Rabenhorst, D., Farrell, E., Jameson, D., Linton, T., and Mandelman, J. (1990). Complementary visualization and sonification of multidimensional data. In E.J. Farrell (Ed.), E. m. f. c. d., editor, *SPIE/SPSE symposium on electronic imaging*, volume 1259, pages 147–153.

- [Reiners et al., 1998] Reiners, D., Stricker, D., Klinker, G., and Muller, S. (1998). Augmented reality for construction task : doorlock assembly. In *Proceedings 1st International Workshop on Augmented Reality (IWAR'98), San Francisco*, pages 31–46.
- [Richard et al., 2006] Richard, E., Tijou, A., Richard, P., and Ferrier, J.-L. (2006). Multi-modal virtual environments for education with haptic and olfactory feedback. *Virtual Reality*, 10(3) :207–225.
- [Robles-De-La-Torre and Hayward, 2000] Robles-De-La-Torre, G. and Hayward, V. (2000). Virtual surfaces and haptic shape perception. In *Haptic interfaces for virtual environments and teleoperator systems symposium. International Mechanical Engineering Congress and Exposition*.
- [Robles-De-La-Torre and Hayward, 2001] Robles-De-La-Torre, G. and Hayward, V. (2001). Force can overcome object geometry in the perception of shape through active touch. *Nature*, 412 :445–448. Reviewed in 'Neurobiology : Feeling bumps and holes' by J.R. Flanagan and S.J. Lederman , News and Views, *Nature*, 2001 Jul 26 ;412(6845) :389-91.
- [Rousseau, 2006] Rousseau, C. (2006). *Présentation multimodale et contextuelle de l'information*. PhD thesis, Université Paris-Sud XI, Orsay, France.
- [Salzman et al., 1996] Salzman, M. C., Dede, C., and Loftin, B. (1996). Sciencespace : Virtual realities for learning complex and abstract scientific concepts. In *IEEE Virtual Reality Annual International Symposium*, pages 246–253. New York : IEEE Press.
- [Schomaker et al., 1995] Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoit, C., Guiard-Marigny, T., Goff, B. L., Robert-Ribes, J., Adjoudani, A., Defee, I., Munch, S., Hartung, K., and Blauert, J. (1995). A taxonomy of multimodal interaction in the human information processing system. A report of Multimodal Integration for Advanced Multimedia Interfaces (MIAMI). ESPRIT III, Basic Research Project 8579.
- [Seeger and Chen, 1997] Seeger, A. and Chen, J. (1997). Controlling force feedback over a network. In *The Second PHANTOM User's Group Workshop*, Cambridge, Massachusetts, USA, MIT.
- [Shi et al., 2007] Shi, X., Cai, Y., and Chan, C. (2007). Electronic music for biomolecules using short music phrases. *Leonardo*, 40 (2) :137–141.
- [Sperandio, 1988] Sperandio, J.-C. (1988). La représentation mentale. In *L'ergonomie du travail mental*, pages 77–90. Paris : Masson.
- [Stephanidis and Savidis, 2001] Stephanidis, C. and Savidis, A. (2001). Universal access in the information society : Methods, tools, and interaction technologies. *Universal Access in the Information Society*, 1 (1) :40–55.

- [Storms and Zyda, 2000] Storms, R. L. and Zyda, M. J. (2000). Interactions in perceived quality of auditory-visual displays. *Presence*, 9(6) :557–580.
- [Sweller and Chandler, 1994] Sweller, J. and Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12 (3) :185–233.
- [Tarault et al., 2005] Tarault, A., Bourdot, P., and Vézien, J.-M. (2005). Sacari : An immersive remote driving interface for autonomous vehicles. In *Workshop on Computer Graphics and Geometric Modelling (CGGM'05)*, in *5th International Conference on Computational Science, (ICCS'05)*, 22-25 May 2005, Atlanta, GA, USA.
- [Tyndiuk and Thomas, 2001] Tyndiuk, F. and Thomas, G. (2001). Modélisation de la navigation en environnement virtuel. In *Quatorzièmes Journées de l'Association Française d'Informatique Graphique (AFIG'01)*, Limoges, France, pages pp 93–100.
- [Vernier, 2001] Vernier, F. (2001). *La multimodalité en sortie et son application à la visualisation de grandes quantités d'information*. PhD thesis, Université Joseph Fourier - Grenoble I.
- [Vernier and Nigay, 2000] Vernier, F. and Nigay, L. (2000). A framework for the combination and characterization of output modalities. In *DSV-IS'00*, pages 32–48, Limerick, Ireland.
- [Väljamäe et al., 1995] Väljamäe, A., Larsson, P., Västfjäll, D., and M.Kleiner (1995). Travelling without moving : Auditory scene cues for translational self-motion. In *Intl. Conf. on Auditory Display (ICAD)*, Limerick Ireland.
- [Wu et al., 1999] Wu, W.-C., Basdogan, C., and Srinivasan, M. A. (1999). Visual, haptic, and bimodal perception of size and stiffness in virtual environments. *ASME Dynamic Systems and Control Division*, 67 :19–26.
- [Yamada et al., 2002] Yamada, T., Ogi, T., Tsubouchi, D., and Hirose, M. (2002). Desk-sized immersive workplace using force feedback grid interface. In *VR '02 : Proceedings of the IEEE Virtual Reality Conference 2002*, page 135, Washington, DC, USA. IEEE Computer Society.