



THESE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

pour obtenir le titre de

DOCTEUR EN INFORMATIQUE

par

Grégory SAUGIS



INTERFACE 3D POUR LE TRAVAIL COOPERATIF SYNCHRONE, UNE PROPOSITION

Thèse soutenue le 30 janvier 1998, devant la commission d'examen :

Président :	J.-M. GEIB	USTL, Villeneuve d'Ascq
Directeur de thèse :	C. CHAILLOU	USTL, Villeneuve d'Ascq
Rapporteurs :	D. THALMANN	EPFL, Lausanne, Suisse
	M. BEAUDOUIN-LAFON	Université Paris Sud, Orsay
Examineurs :	C. KINTZIG	CNET, Issy-Les-Moulineaux
	A. DERYCKE	CUEEP, Villeneuve d'Ascq

DOYENS HONORAIRES DE L'ANCIENNE FACULTE DES SCIENCES

M. H. LEFEBVRE, M. PARREAU

PROFESSEURS HONORAIRES DES ANCIENNES FACULTES DE DROIT
ET SCIENCES ECONOMIQUES, DES SCIENCES ET DES LETTRES

MM. ARNOULT, BONTE, BROCHARD, CHAPPELON, CHAUDRON, CORDONNIER, DECUYPER, DEHEUVELS, DEHORS, DION, FAUVEL, FLEURY, GERMAIN, GLACET, GONTIER, KOURGANOFF, LAMOTTE, LASSERRE, LELONG, LHOMME, LIEBAERT, MARTINOT-LAGARDE, MAZET, MICHEL, PEREZ, ROIG, ROSEAU, ROUELLE, SCHILTZ, SAVARD, ZAMANSKI, Mes BEAUJEU, LELONG.

PROFESSEUR EMERITE

M. A. LEBRUN

ANCIENS PRESIDENTS DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

MM. M. PARREAU, J. LOMBARD, M. MIGEON, J. CORTOIS, A. DUBRULLE

PRESIDENT DE L'UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

M. P. LOUIS

PROFESSEURS - CLASSE EXCEPTIONNELLE

M. CHAMLEY Hervé
M. CONSTANT Eugène
M. ESCAIG Bertrand
M. FOURET René
M. GABILLARD Robert
M. LABLACHE COMBIER Alain
M. LOMBARD Jacques
M. MACKE Bruno

Géotechnique
Electronique
Physique du solide
Physique du solide
Electronique
Chimie
Sociologie
Physique moléculaire et rayonnements atmosphériques

M. MIGEON Michel
M. MONTREUIL Jean
M. PARREAU Michel
M. TRIDOT Gabriel

EUDIL
Biochimie
Analyse
Chimie appliquée

PROFESSEURS - 1ère CLASSE

M. BACCHUS Pierre
M. BIAYS Pierre
M. BILLARD Jean
M. BOILLY Bénoni
M. BONNELLE Jean Pierre
M. BOSCOQ Denis
M. BOUGHON Pierre
M. BOURIQUET Robert
M. BRASSELET Jean Paul
M. BREZINSKI Claude
M. BRIDOUX Michel
M. BRUYELLE Pierre
M. CARREZ Christian
M. CELET Paul
M. COEURE Gérard
M. CORDONNIER Vincent
M. CROSNIER Yves
Mme DACHARRY Monique
M. DAUCHET Max
M. DEBOURSE Jean Pierre
M. DEBRABANT Pierre
M. DECLERCQ Roger
M. DEGAUQUE Pierre
M. DESCHEPPER Joseph
Mme DESSAUX Odile
M. DHAINAUT André
Mme DHAINAUT Nicole
M. DJAFARI Rouhani
M. DORMARD Serge
M. DOUKHAN Jean Claude
M. DUBRULLE Alain
M. DUPOUY Jean Paul
M. DYMENT Arthur
M. FOCT Jacques Jacques
M. FOUQUART Yves
M. FOURNET Bernard
M. FRONTIER Serge
M. GLORIEUX Pierre
M. GOSSELIN Gabriel
M. GOUDMAND Pierre
M. GRANELLE Jean Jacques
M. GRUSON Laurent
M. GUILBAULT Pierre
M. GUILLAUME Jean
M. HECTOR Joseph
M. HENRY Jean Pierre
M. HERMAN Maurice
M. LACOSTE Louis
M. LANGRAND Claude

Astronomie
Géographie
Physique du Solide
Biologie
Chimie-Physique
Probabilités
Algèbre
Biologie Végétale
Géométrie et topologie
Analyse numérique
Chimie Physique
Géographie
Informatique
Géologie générale
Analyse
Informatique
Electronique
Géographie
Informatique
Gestion des entreprises
Géologie appliquée
Sciences de gestion
Electronique
Sciences de gestion
Spectroscopie de la réactivité chimique
Biologie animale
Biologie animale
Physique
Sciences Economiques
Physique du solide
Spectroscopie hertzienne
Biologie
Mécanique
Métallurgie
Optique atmosphérique
Biochimie structurale
Ecologie numérique
Physique moléculaire et rayonnements atmosphériques
Sociologie
Chimie-Physique
Sciences Economiques
Algèbre
Physiologie animale
Microbiologie
Géométrie
Génie mécanique
Physique spatiale
Biologie Végétale
Probabilités et statistiques

M. LATTEUX Michel
M. LA VEINE Jean Pierre
Mme LECLERCQ Ginette
M. LEHMANN Daniel
Mme LENOBLE Jacqueline
M. LEROY Jean Marie
M. LHENAFF René
M. LHOMME Jean
M. LOUAGE François
M. LOUCHEUX Claude
M. LUCQUIN Michel
M. MAILLET Pierre
M. MAROUF Nadir
M. MICHEAU Pierre
M. PAQUET Jacques
M. PASZKOWSKI Stéfan
M. PETIT Francis
M. PORCHET Maurice
M. POUZET Pierre
M. POVY Lucien
M. PROUVOST Jean
M. RACZY Ladislas
M. RAMAN Jean Pierre
M. SALMER Georges
M. SCHAMPS Joël
Mme SCHWARZBACH Yvette
M. SEGUIER Guy
M. SIMON Michel
M. SLIWA Henri
M. SOMME Jean
Melle SPIK Geneviève
M. STANKIEWICZ François
M. THIEBAULT François
M. THOMAS Jean Claude
M. THUMERELLE Pierre
M. TILLIEU Jacques
M. TOULOTTE Jean Marc
M. TREANTON Jean René
M. TURRELL Georges
M. VANEECLOO Nicolas
M. VAST Pierre
M. VERBERT André
M. VERNET Philippe
M. VIDAL Pierre
M. WALLART François
M. WEINSTEIN Olivier
M. ZEYTOUNIAN Radyadour

Informatique
Paléontologie
Catalyse
Géométrie
Physique atomique et moléculaire
Spectrochimie
Géographie
Chimie organique biologique
Electronique
Chimie-Physique
Chimie physique
Sciences Economiques
Sociologie
Mécanique des fluides
Géologie générale
Mathématiques
Chimie organique
Biologie animale
Modélisation - calcul scientifique
Automatique
Minéralogie
Electronique
Sciences de gestion
Electronique
Spectroscopie moléculaire
Géométrie
Electrotechnique
Sociologie
Chimie organique
Géographie
Biochimie
Sciences Economiques
Sciences de la Terre
Géométrie - Topologie
Démographie - Géographie humaine
Physique théorique
Automatique
Sociologie du travail
Spectrochimie infrarouge et raman
Sciences Economiques
Chimie inorganique
Biochimie
Génétique
Automatique
Spectrochimie infrarouge et raman
Analyse économique de la recherche et développement
Mécanique

PROFESSEURS - 2ème CLASSE

M. ABRAHAM Francis	Composants électroniques
M. ALLAMANDO Etienne	Biologie des organismes
M. ANDRIES Jean Claude	Analyse
M. ANTOINE Philippe	Génétique
M. BALL Steven	Biologie animale
M. BART André	Génie des procédés et réactions chimiques
M. BASSERY Louis	Géographie
Mme BATTIAU Yvonne	Systèmes électroniques
M. BAUSIERE Robert	Mécanique
M. BEGUIN Paul	Physique atomique et moléculaire
M. BELLET Jean	Physique atomique, moléculaire et du rayonnement
M. BERNAGE Pascal	Sciences Economiques
M. BERTHOUD Arnaud	Sciences Economiques
M. BERTRAND Hugues	Analyse
M. BERZIN Robert	Physique de l'état condensé et cristallographie
M. BISKUPSKI Gérard	Algèbre
M. BKOUCHE Rudolphe	Biologie végétale
M. BODARD Marcel	Biochimie métabolique et cellulaire
M. BOHIN Jean Pierre	Mécanique
M. BOIS Pierre	Génie civil
M. BOISSIER Daniel	Spectrochimie
M. BOIVIN Jean Claude	Physique
M. BOUCHER Daniel	Biologie appliquée aux enzymes
M. BOUQUELET Stéphane	Gestion
M. BOUQUIN Henri	Chimie
M. BROCARD Jacques	Paléontologie
Mme BROUSMICHE Claudine	Mécanique
M. BUISINE Daniel	Biologie animale
M. CAPURON Alfred	Géographie humaine
M. CARRE François	Chimie organique
M. CATTEAU Jean Pierre	Sciences Economiques
M. CAYATTE Jean Louis	Electronique
M. CHAPOTON Alain	Biochimie structurale
M. CHARET Pierre	Composants électroniques optiques
M. CHIVE Maurice	Informatique théorique
M. COMYN Gérard	Composants électroniques et optiques
Mme CONSTANT Monique	Psychophysiologie
M. COQUERY Jean Marie	Sciences Economiques
M. CORLAT Benjamin	Paléontologie
Mme CORSIN Paule	Physique nucléaire et corpusculaire
M. CORTOIS Jean	Chimie organique
M. COUTURIER Daniel	Tectonique géodynamique
M. CRAMPON Norbert	Biologie
M. CURGY Jean Jacques	Physique théorique
M. DANGOISSE Didier	Analyse
M. DE PARIS Jean Claude	Composants électroniques et optiques
M. DECOSTER Didier	Electrochimie et Cinétique
M. DEJAEGER Roger	Informatique
M. DELAHAYE Jean Paul	Physiologie animale
M. DELORME Pierre	Sciences Economiques
M. DELORME Robert	Sociologie
M. DEMUNTER Paul	Physique atomique, moléculaire et du rayonnement
Mme DEMUYNCK Claire	Informatique
M. DENEL Jacques	Physique du solide - cristallographie
M. DEPREZ Gilbert	

M. DERIEUX Jean Claude	Microbiologie
M. DERYCKE Alain	Informatique
M. DESCAMPS Marc	Physique de l'état condensé et cristallographie
M. DEVRAINNE Pierre	Chimie minérale
M. DEWAILLY Jean Michel	Géographie humaine
M. DHAMELINCOURT Paul	Chimie physique
M. DI PERSIO Jean	Physique de l'état condensé et cristallographie
M. DUBAR Claude	Sociologie démographique
M. DUBOIS Henri	Spectroscopie hertzienne
M. DUBOIS Jean Jacques	Géographie
M. DUBUS Jean Paul	Spectrométrie des solides
M. DUPONT Christophe	Vie de la firme
M. DUTHOIT Bruno	Génie civil
Mme DUVAL Anne	Algèbre
Mme EVRARD Micheline	Génie des procédés et réactions chimiques
M. FAKIR Sabah	Algèbre
M. FARVACQUE Jean Louis	Physique de l'état condensé et cristallographie
M. FAUQUEMBERGUE Renaud	Composants électroniques
M. FELIX Yves	Mathématiques
M. FERRIERE Jacky	Tectonique - Géodynamique
M. FISCHER Jean Claude	Chimie organique, minérale et analytique
M. FONTAINE Hubert	Dynamique des cristaux
M. FORSE Michel	Sociologie
M. GADREY Jean	Sciences économiques
M. GAMBLIN André	Géographie urbaine, industrielle et démographie
M. GOBLOT Rémi	Algèbre
M. GOURIEROUX Christian	Probabilités et statistiques
M. GREGORY Pierre	I.A.E.
M. GREMY Jean Paul	Sociologie
M. GREVET Patrice	Sciences Economiques
M. GRIMBLOT Jean	Chimie organique
M. GUELTON Michel	Chimie physique
M. GUICHAOUA André	Sociologie
M. HAIMAN Georges	Modélisation,calcul scientifique, statistiques
M. HOUDART René	Physique atomique
M. HUEBSCHMANN Johannes	Mathématiques
M. HUTTNER Marc	Algèbre
M. ISAERT Noël	Physique de l'état condensé et cristallographie
M. JACOB Gérard	Informatique
M. JACOB Pierre	Probabilités et statistiques
M. JEAN Raymond	Biologie des populations végétales
M. JOFFRE Patrick	Vie de la firme
M. JOURNAL Gérard	Spectroscopie hertzienne
M. KOENIG Gérard	Sciences de gestion
M. KOSTRUBIEC Benjamin	Géographie
M. KREMBEL Jean	Biochimie
Mme KRIFA Hadjila	Sciences Economiques
M. LANGEVIN Michel	Algèbre
M. LASSALLE Bernard	Embryologie et biologie de la différenciation
M. LE MEHAUTE Alain	Modélisation,calcul scientifique,statistiques
M. LEBFEVRE Yannic	Physique atomique,moléculaire et du rayonnement
M. LECLERCQ Lucien	Chimie physique
M. LEFEBVRE Jacques	Physique
M. LEFEBVRE Marc	Composants électroniques et optiques
M. LEFEVRE Christian	Pétrologie
Melle LEGRAND Denise	Algèbre
M. LEGRAND Michel	Astronomie - Météorologie
M. LEGRAND Pierre	Chimie
Mme LEGRAND Solange	Algèbre
Mme LEHMANN Josiane	Analyse
M. LEMAIRE Jean	Spectroscopie hertzienne

M. LE MAROIS Henri
 M. LEMOINE Yves
 M. LESCURE François
 M. LESENNE Jacques
 M. LOCQUENEUX Robert
 Mme LOPES Maria
 M. LOSFELD Joseph
 M. LOUAGE Francis
 M. MAHIEU François
 M. MAHIEU Jean Marie
 M. MAIZIERES Christian
 M. MANSY Jean Louis
 M. MAURISSON Patrick
 M. MERIAUX Michel
 M. MERLIN Jean Claude
 M. MESMACQUE Gérard
 M. MESSELYN Jean
 M. MOCHE Raymond
 M. MONTEL Marc
 M. MORCELLET Michel
 M. MORE Marcel
 M. MORTREUX André
 Mme MOUNIER Yvonne
 M. NIAY Pierre
 M. NICOLE Jacques
 M. NOTELET Francis
 M. PALAVIT Gérard
 M. PARSY Fernand
 M. PECQUE Marcel
 M. PERROT Pierre
 M. PERTUZON Emile
 M. PETIT Daniel
 M. PLIHON Dominique
 M. PONSOLLE Louis
 M. POSTAIRE Jack
 M. RAMBOUR Serge
 M. RENARD Jean Pierre
 M. RENARD Philippe
 M. RICHARD Alain
 M. RIETSCH François
 M. ROBINET Jean Claude
 M. ROGALSKI Marc
 M. ROLLAND Paul
 M. ROLLET Philippe
 Mme ROUSSEL Isabelle
 M. ROUSSIGNOL Michel
 M. ROY Jean Claude
 M. SALERNO François
 M. SANCHOLLE Michel
 Mme SANDIG Anna Margarete
 M. SAWERYSYN Jean Pierre
 M. STAROSWIECKI Marcel
 M. STEEN Jean Pierre
 Mme STELLMACHER Irène
 M. STERBOUL François
 M. TAILLIEZ Roger
 M. TANRE Daniel
 M. THERY Pierre
 Mme TJOTTA Jacqueline
 M. TOURSEL Bernard
 M. TREANTON Jean René

Vie de la firme
 Biologie et physiologie végétales
 Algèbre
 Systèmes électroniques
 Physique théorique
 Mathématiques
 Informatique
 Electronique
 Sciences économiques
 Optique - Physique atomique
 Automatique
 Géologie
 Sciences Economiques
 EUDIL
 Chimie
 Génie mécanique
 Physique atomique et moléculaire
 Modélisation, calcul scientifique, statistiques
 Physique du solide
 Chimie organique
 Physique de l'état condensé et cristallographie
 Chimie organique
 Physiologie des structures contractiles
 Physique atomique, moléculaire et du rayonnement
 Spectrochimie
 Systèmes électroniques
 Génie chimique
 Mécanique
 Chimie organique
 Chimie appliquée
 Physiologie animale
 Biologie des populations et écosystèmes
 Sciences Economiques
 Chimie physique
 Informatique industrielle
 Biologie
 Géographie humaine
 Sciences de gestion
 Biologie animale
 Physique des polymères
 EUDIL
 Analyse
 Composants électroniques et optiques
 Sciences Economiques
 Géographie physique
 Modélisation, calcul scientifique, statistiques
 Psychophysiologie
 Sciences de gestion
 Biologie et physiologie végétales

 Chimie physique
 Informatique
 Informatique
 Astronomie - Météorologie
 Informatique
 Génie alimentaire
 Géométrie - Topologie
 Systèmes électroniques
 Mathématiques
 Informatique
 Sociologie du travail

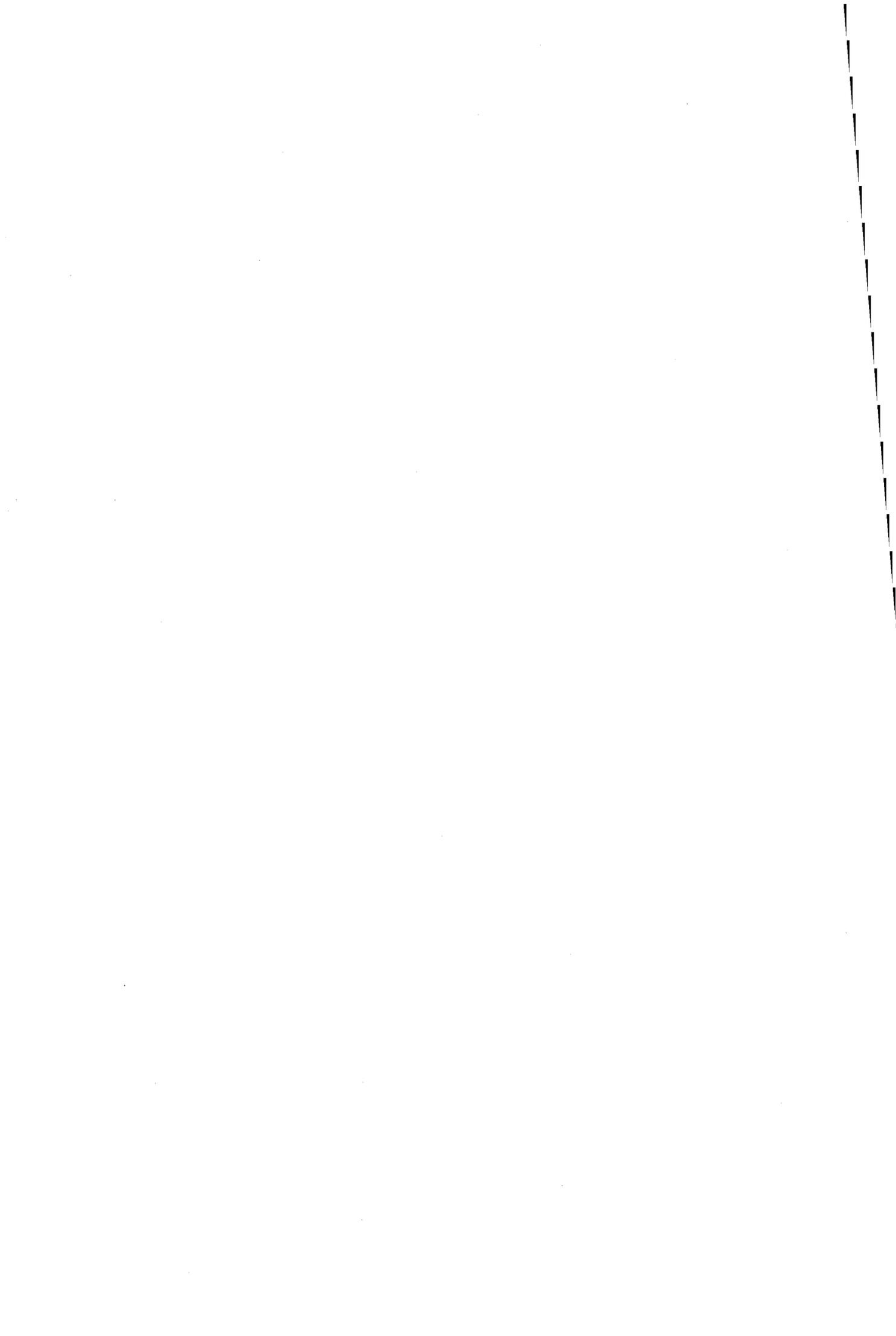
M. TURREL Georges
M. VANDIJK Hendrik
Mme VAN ISEGHEM Jeanine
M. VANDORPE Bernard
M. VASSEUR Christian
M. VASSEUR Jacques
Mme VIANO Marie Claude
M. WACRENIER Jean Marie
M. WARTEL Michel
M. WATERLOT Michel
M. WEICHERT Dieter
M. WERNER Georges
M. WIGNACOURT Jean Pierre
M. WOZNIAK Michel
Mme ZINN JUSTIN Nicole

Spectrochimie infrarouge et raman

Modélisation, calcul scientifique, statistiques
Chimie minérale
Automatique
Biologie

Electronique
Chimie inorganique
géologie générale
Génie mécanique
Informatique théorique

Spectrochimie
Algèbre



Remerciements

Je souhaiterais remercier tous les membres du jury et tout d'abord Monsieur Jean-Marc Geib pour avoir accepté de le présider.

Messieurs Michel Beaudouin-Lafon et Daniel Thalmann m'ont fait le plaisir d'accepter de rapporter mon manuscrit de thèse. C'est un honneur d'avoir pu soumettre mes travaux à leur approbation.

Merci à Monsieur Claude Kintzig du CNET d'avoir examiné ma thèse, ainsi qu'à Gérard Poulain qui a fait le déplacement pour y assister. Merci aux autres membres du CNET avec qui nous avons eu l'occasion de collaborer : Mademoiselle Bégoc, Messieurs Boucharet, Brillet et Anne.

Monsieur Alain Derycke a suivi mes travaux depuis le début de ma thèse. Sa culture, son aide, ses conseils m'ont été précieux dans ce projet pluridisciplinaire. C'est sans hésitation qu'il a accepté d'être examinateur de ma thèse, me prouvant ainsi son intérêt dans les résultats obtenus.

Monsieur Christophe Chaillou supervise l'ensemble des travaux de l'équipe graphique. Il a été mon directeur de thèse et je l'en remercie. Je dois souligner en particulier ses capacités à suivre et faire avancer un projet, sans perdre de vue le but final.

L'ENIC m'a accueilli dans ses murs, je salue donc ici toutes les personnes que j'y ai côtoyé, en particulier le personnel du département informatique et du SIR.

J'ai également séjourné longuement au laboratoire Trigone, notre partenaire dans le projet SPACE. Je salue ici tous les gens que j'y ai rencontrés.

Une thèse n'est bien évidemment pas qu'un travail personnel ! Merci à tous les membres de l'équipe GRAPHIX présents et passés pour leur dynamisme, leurs connaissances et la chaleureuse ambiance qu'ils font régner. Nos voisins de l'équipe SNCF sont au moins aussi dynamiques et sympathiques. Je me dois de mentionner les individus suivants : Chewie, la Chose, le Coyote, Froum, Gbad, Hervénounet, l'Infâââme, Karim Megastore, LUX 11-38, Meumeu, Mr. Magooille, Perux, SGBDr... *et SAM!*

Mes amis.

Ma famille.

Amaria, mon égérie.

Table des matières

INTRODUCTION.....	7
1. ÉTUDE BIBLIOGRAPHIQUE	13
1.1 LE TRAVAIL COOPÉRATIF SYNCHRONE DISTANT	13
<i>1.1.1 Les outils vidéo</i>	<i>14</i>
L'hégémonie des outils audiovisuels	14
Les outils audiovisuels informatisés	17
L'attrait pour la vidéo	18
Quelques travaux de recherche	22
i) Les Hydra units	22
ii) Clearface	23
iii) ClearBoard	24
iv) MAJIC	26
<i>1.1.2 Les « environnements virtuels collaboratifs »</i>	<i>27</i>
DIVE.....	28
MASSIVE.....	29
<i>1.1.3 Les outils « informatiques »</i>	<i>29</i>
Genesys.....	29
Teledesign	31
CSpray	34
L'environnement multimédia interactif AT&T	35

1.2 LA RÉALITÉ VIRTUELLE	38
1.2.1 Rappel sur la synthèse d'images temps réel	41
1.2.2 Les périphériques d'entrée et de sortie.....	42
1.2.3 La vision en R.V.....	45
La vue couplée aux mouvements de la tête	45
Vision périphérique en R.V. immersive	45
1.2.4 Les avatars et les clones	48
Dans DIVE	48
Dans MASSIVE.....	50
Difficulté du clonage total	51
Les travaux de l'INA.....	54
i) Analyse	54
ii) Synthèse	56
iii) Réalisme	57
iv) Le futur des clones à l'INA	60
1.3 LES INTERFACES.....	60
1.3.1 Les interfaces 2D WIMP	60
X-Window : mécanismes d'implémentation	62
1.3.2 Des proposition d'améliorations.....	63
Le concept de pièce.....	64
Le symbolisme dans les interfaces	65
La sémantique des fenêtres – rôle de la transparence	65
1.3.3 Les vues fish-eye.....	69
1.3.4 L'« Information Visualizer »	70
1.3.5 La perception et le repérage en 3D	74
2. ANALYSE ET RECOMMANDATIONS.....	75
2.1 OBSERVATIONS GÉNÉRALES	75
2.1.1 Les motivations à l'origine de ces projets	76
2.1.2 Les limites de la Réalité Virtuelle.....	76
2.1.3 L'évolution vers les interfaces 3D	78
Inconvénients des interfaces WIMP	78
Le passage à la 3D.....	79
2.2 DEUX PROJETS À LA LOUPE	80
2.2.1 MAJIC	80
2.2.2 MASSIVE.....	85
2.3 NOS RECOMMANDATIONS.....	91
2.3.1 Travail coopératif synchrone.....	91
2.3.2 Interfaces	92
2.3.3 Réalité Virtuelle.....	94

3. NOTRE PROPOSITION.....	97
3.1 LE PROJET SPACE	97
3.1.1 <i>Architecture du projet SPACE</i>	98
3.1.2 <i>ODESCA : gestion de l'activité</i>	99
Connaissance des participants, des documents et des outils	99
Connaissance du découpage de la tâche	100
Connaissance des droits de chacun	100
Maintenance de la persistance	100
3.1.3 <i>ScenIC : représentation abstraite de la scène</i>	101
Représentation abstraite de la scène et de la tâche associée.	101
Stockage des contraintes spatio-temporelles.	103
Point de passage obligatoire entre Spln et ODESCA.	103
Stockage des propriétés des différents composants de la tâche.	103
Gestion améliorée de la qualité de service.	104
3.2 NOS CHOIX ESSENTIELS	105
3.2.1 <i>Terminal de communication individuel</i>	105
3.2.2 <i>Téléprésence et perception de l'activité</i>	106
3.2.3 <i>Environnement 3D : métaphore de table de réunion</i>	107
3.2.4 <i>Une interface centrée sur l'activité</i>	109
3.2.5 <i>Tout visible</i>	109
3.2.6 <i>Adaptation de la Réalité Virtuelle : simplifications et abstractions</i>	110
3.2.7 <i>Animation et fluidité</i>	111
3.3 LA RÉALISATION.....	112
3.3.1 <i>La table de réunion</i>	112
3.3.2 <i>Navigation</i>	113
3.3.3 <i>Tout perceptible</i>	115
3.3.4 <i>Distorsion</i>	116
3.3.5 <i>Manipulation</i>	118
3.3.6 <i>Clones</i>	119
Représentation des personnes	119
Intérêt des clones	120
i) Yeux.....	120
ii) Télépointeur	121
3.3.7 <i>Vue subjective</i>	121
3.4 ÉLÉMENTS SUR L'IMPLÉMENTATION.....	123
3.4.1 <i>Plate-forme d'accueil</i>	123
3.4.2 <i>Objets graphiques</i>	123
3.4.3 <i>Apparition d'un nouvel objet dans le bandeau</i>	124
3.4.4 <i>Pointage et objets composés</i>	125
3.4.5 <i>Messages</i>	126
3.4.6 <i>L'échec de l'intégration directe d'applications Windows</i>	127

3.4.7 <i>La distorsion horizontale</i>	128
3.4.8 <i>Écrasement latéral des objets du bandeau</i>	130
3.4.9 <i>Regard</i>	132
3.4.10 <i>Télépointeurs</i>	133
3.5 LE PROTOTYPE ACTUEL DE SPIN.....	134
3.6 TRAVAUX EN COURS	140
4. CONCLUSION ET PERSPECTIVES	143
4.1 CONCLUSION.....	143
4.2 PERSPECTIVES.....	145
RÉFÉRENCES BIBLIOGRAPHIQUES	149
ANNEXE 1 — DÉFINITIONS	155
ANNEXE 2 — TABLE DES FIGURES	159

Introduction

Au fur et à mesure que les technologies de communication évoluent, les habitudes de travail sont modifiées. De plus en plus, il est facile de joindre une personne n'importe quand et n'importe où, et avec toujours plus de possibilités : communication vocale, documents papier par fax, documents informatiques par modem ou réseau informatique, vidéoconférence... Stratégiquement, c'est un atout pour les grosses entreprises implantées dans des régions ou des pays éloignés les uns des autres. La tendance est à l'abandon d'un lieu de travail commun à une équipe au profit du travail chez soi ou chez le client. Il n'est plus nécessaire de disposer d'un bureau à plein temps. On loue un bâtiment, une salle au coup par coup en fonction des besoins. Dans certaines entreprises, chacun dispose d'un casier roulant qui lui est propre et qu'il emmène dans un bureau quelconque, libre. Ces travailleurs « pendulaires » passent une partie de leur temps de travail dans leur entreprise et le reste à l'extérieur. Des bureaux de voisinage s'ouvrent, qui accueillent des employés de plusieurs firmes près de leur lieu d'habitation. Les coûts se réduisent, l'organisation du travail est simplifiée, les contraintes géographiques ne sont plus un obstacle à l'efficacité.

La demande est donc forte, surtout chez les travailleurs de bureau, à la fois pour des outils de communication évolués, « multimédia », et pour des systèmes informatiques permettant de travailler à plusieurs. Notre approche consiste à intégrer la communication à l'outil informatique et à utiliser l'ordinateur à la fois comme poste de production et comme espace de travail de groupe. Le terminal ainsi constitué aurait l'avantage d'être banalisé et de réutiliser du matériel déjà largement distribué. Les documents pourraient être créés ou modifiés sur-le-champ, sans avoir à passer d'une étape de discussion à une étape de production.

La Région Nord Pas-de-Calais soutient un pôle de recherche pluridisciplinaire sur la Communication Avancée, appelé Ganymède. Plusieurs laboratoires participent au projet, couvrant de nombreux domaines : informatique (synthèse d'image temps réel, systèmes d'exploitation à objets, cartes à puces et objets nomades, intelligence artificielle et agents...), travail coopératif et enseignement à distance, ergonomie et psychologie, médecine. Les équipes de recherches impliquées dans Ganymède sont incitées à mettre en commun leurs connaissances spécialisées, à travailler ensemble sur ce thème de la communication avancée.

De son côté, le CNET (Centre National d'Étude des Télécommunications) s'intéresse également de près aux nouvelles technologies de communication et au travail coopératif « médiatisé », c'est-à-dire s'effectuant à distance par l'intermédiaire d'outils artificiels. Les outils « synchrones », grâce auxquels les utilisateurs travaillent et communiquent simultanément, sont encore peu développés car très dépendants des technologies les plus récentes. Le CNET favorise donc la recherche dans ce domaine, notamment sous forme d'appels d'offres, les CTI.

Une coopération a été mise en place entre deux laboratoires au sein de Ganymède. Le premier, le Laboratoire d'Informatique Fondamentale de Lille (LIFL) est représenté par l'équipe Graphix, spécialiste de la synthèse d'images temps réel. Le second est Trigone, en particulier l'équipe NOCE, dont les centres d'intérêt sont le travail coopératif, les interfaces et l'enseignement à distance. Le projet commun, baptisé SPACE, s'intègre naturellement dans le pôle régional Ganymède. Un contrat de financement avec le CNET et la Région a été passé. Deux thèses, celle de Patrice Carlier au CUEEP et la mienne au LIFL, ont été financées dans le cadre de ce contrat. Le contrat, intitulé « *Interactions Homme-Machine multidimensionnelles, multimédia, multiparties pour applications de groupe* », stipule qu'un nouveau modèle d'interface adaptée au travail coopératif doit être défini. Plus précisément, trois actions ont dû être menées : veille technologique et état de l'art, expérimentations à caractère exploratoire de maquettes d'interfaces hommes-machines et enfin conception et réalisation de prototypes d'interfaces hommes-machines tridimensionnelles et multimédia pour le service de groupe. Les points forts du projet sont l'intégration complète sur station de travail informatique standard pour constituer un terminal de communication évolué à moindre coût, les techniques multimédia (son, vidéo), et les apports de la 3D temps réel. L'équipe Graphix du LIFL apporte ses compétences pour concevoir et réaliser la présentation de l'interface. Ce mémoire va donc porter principalement sur Spln¹, la partie graphique de l'interface, de sa conception à la mise en œuvre d'un prototype. Trigone, au sein du CUEEP, possède une grande expérience de l'enseignement, à distance en particulier, et comprend des informaticiens, des psychologues et des psychologues cogniticiens parmi ses membres. Ils ont pour mission d'assurer le module de gestion de l'activité et ce que nous avons pris l'habitude d'appeler le « moteur coopératif ». Cette tâche profite de l'expérience acquise notamment par Pascal Croisy [Croisy95] et Frédéric Hoogstoel [Hoogstoel95] lors de leur participation au projet Co-Learn.

Les applications visées sont la réunion à distance, l'enseignement à distance (en groupe, cours particulier avec expert...), les services d'assistance (« hot line ») etc. Notre interface devra donc être conçue pour supporter la communication interpersonnelle et le travail en groupe sur documents partagés. Cela implique des contraintes inhabituelles en raison de la dimension humaine du produit, beaucoup plus importante que dans la majorité des produits informatiques.

Le projet SPACE s'insère naturellement dans l'évolution technologique. La puissance des ordinateurs augmente régulièrement et rapidement. Les progrès se font ressentir à tous les niveaux : grandes capacités de mémoire vive et de stockage, processeur central capable d'exécuter des centaines de millions d'instructions à la seconde, accélérateurs graphiques, écrans de grande définition, son de bonne qualité. La puissance des processeurs double environ tous les deux ans. La population des utilisateurs d'ordinateurs a largement évolué. Les savants en blouse blanche des premiers temps de l'informatique ont depuis longtemps été rejoints par toutes les professions, par exemple les banquiers, qui profitent de la puissance de calcul, de stockage et de traitement des ordinateurs. Les progrès de la technique ont fait évoluer les mentalités : les ordinateurs personnels ont progressivement remplacé les outils de calcul, de stockage, d'édition. Ils ont été mis en réseau au sein de l'entreprise. Ce qui n'était au départ qu'un outil scientifique est devenu un outil de bureau performant ; c'est probablement ce qui a déterminé l'omniprésence de l'informatique. Les champs d'applications touchent maintenant tous les domaines, en particulier les loisirs (jeux vidéo, cinéma...) et tout un chacun est capable de se servir d'un ordinateur. Aujourd'hui, avec Internet, tout le monde peut accéder à un réseau mondial, ouvert, qui n'a pas fini d'être exploré. Le World Wide Web est sa vitrine la plus connue. La navigation hypertexte, l'inclusion de documents multimédia, les extensions 3D avec VRML, l'inclusion de plusieurs protocoles (FTP, WAIS, NEWS...) ont rendu les *browsers Web* incontournables.

Jusqu'à maintenant, les débits autorisés par les canaux de communication informatiques ne permettaient pas de transmettre des données de grande taille en temps réel. En particulier, des flux vidéo ou même seulement sonores, requéraient une bande passante que les infrastructures en place ne pouvaient offrir. Les réseaux informatiques n'étaient de surcroît pas conçus pour supporter la synchronisation de données ou des notions de temps réel. Le développement de réseaux rapides comme RNIS² ou maintenant ATM, ainsi que les forts taux de compressions atteints, par exemple avec la norme MPEG-2, permettent d'envisager de communiquer avec un son et une image de bonne qualité par des moyens entièrement numériques. Les modems autorisent des taux de transfert élevés sur le réseau téléphonique commuté et offrent une plus grande liberté que les réseaux dédiés. Pour le particulier, c'est encore la meilleure solution pour relier son ordinateur à d'autres machines. La faisabilité technique d'outils de communication de groupe est maintenant acquise.

¹ *SPace INterface*

² Réseau Numérique à Intégration de Services, *ISDN* en anglais.

L'évolution des postes de travail se fait aussi par la spécialisation du matériel. Les systèmes à fenêtres, notamment Windows sur PC, ont amené les constructeurs à construire des cartes accélératrices 2D prenant en charge les fonctions courantes de transferts mémoire, libérant ainsi le processeur central d'une lourde charge, optimisant les échanges mémoire. On voit maintenant apparaître des cartes de compression/décompression pour la vidéo et le son, mais aussi des accélérateurs 3D dans lesquels sont câblés de nombreux algorithmes utilisés en synthèse d'images. La nouveauté vient surtout du prix auquel les cartes 3D récentes sont proposées. On obtient quasiment la puissance d'une station graphique haut de gamme sur un PC, en partie grâce à l'adoption par Microsoft de la bibliothèque OpenGL dans ses systèmes d'exploitation 95 et NT. Le marché du jeu a également favorisé l'apparition de cartes au rapport performances/prix inimaginable sans une production de masse. Leur puissance de calcul et la qualité des images qu'elles génèrent sont étonnantes.

La technologie est capable actuellement de fournir des solutions pour le travail coopératif à distance. Mais le progrès matériel n'a d'intérêt que si les utilisateurs peuvent en profiter pleinement. Il manque encore les métaphores et les concepts permettant de développer des outils performants, utilisables par le plus grand nombre de personnes. Le confort, la facilité de prise en main peuvent et doivent être améliorés. En particulier, les interfaces doivent être repensées pour respecter les interactions homme-homme lors des sessions de travail coopératif distant. Nous pensons que les anciennes métaphores, conventions graphiques et méthodes d'interaction doivent être repensées pour tirer parti des technologies récentes, en les utilisant et en les mettant à la portée des utilisateurs.

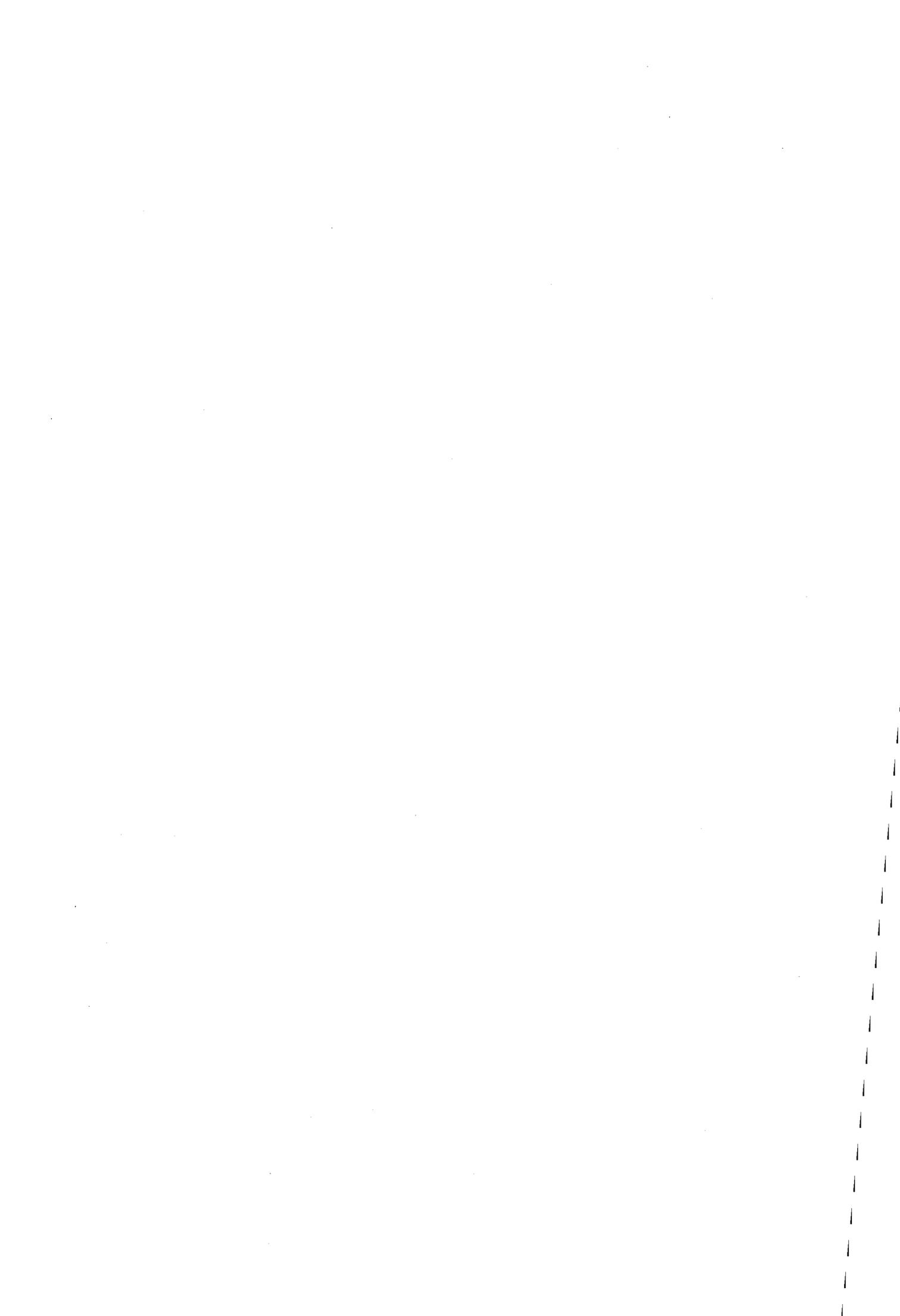
Les premiers travaux sur ce qui allait devenir SPACE ont, au LIFL, été effectués par Laurent Dumesges [Derycke93] puis Luc Wibaux [Chaillou94], sous la forme de mémoires de DEA. Les premières idées concernaient déjà la notion de pièce, la navigation et le placage de textures. Les premières maquettes étant prometteuses, le projet s'est officialisé en 1994 avec le contrat passé entre le LIFL, Trigone, la Région Nord Pas-De-Calais et le CNET.

Cette activité étant nouvelle pour le LIFL, une première étape a été d'acquérir une culture dans les domaines du travail coopératif et de l'interaction homme-machine (IHM). Une large étude bibliographique a été entreprise, pour acquérir les notions importantes dans les domaines de la perception humaine, des outils de travail coopératif informatiques ou non, des nouvelles interfaces, de la Réalité Virtuelle. Ayant grandement accru notre compréhension des attentes de la communauté IHM, nous avons donc fait le point sur les outils novateurs et leurs apports respectifs pour le travail de groupe médiatisé. Certains ont été des modèles pour nous, tant les interactions qu'ils permettent entre utilisateurs sont riches. Chacun d'eux apporte des nouveaux concepts ou une technique de mise en œuvre différente. Les similitudes entre les différents systèmes et les particularités de chaque produit ont permis de dégager une liste de directives pour la conception et la mise en œuvre d'un outil de travail coopératif médiatisé.

Parallèlement, notre démarche en tant qu'équipe chargée de la partie graphique s'est éloignée de la Réalité Virtuelle stricte. Nous cherchons à concevoir un nouvel outil bureautique qui ne soit pas une

copie fidèle de la réalité. D'une part, modéliser le réel est un problème excessivement dur déjà affronté par les nombreuses équipes de recherche en Réalité Virtuelle. D'autre part, cela ne nous a pas paru utile. Notre démarche a été de garder du monde réel ce qui aide les utilisateurs et d'ajouter des possibilités n'existant pas dans la réalité mais qu'un monde abstrait, informatique, ne répondant pas aux lois physiques peut offrir.

Le plan de ce mémoire suit globalement le déroulement de nos recherches. Nous commençons par l'étude de travaux dans les domaines du travail coopératif synchrone médiatisé, de la Réalité Virtuelle et des interfaces. Ensuite, après analyse, nous dressons une liste de recommandations formant un guide pour la réalisation d'un outil de travail coopératif (du point de vue, toujours, de son aspect visuel et non de la gestion de l'activité). Enfin nous abordons SpIn, notre proposition d'interface. Nous expliquons en premier lieu comment nous interprétons les recommandations précédentes, quelles solutions nous apportons, avant de décrire le prototype que nous avons développé. Le vocabulaire spécifique employé dans ce mémoire est défini dans « Annexe 1 — Définitions ».



1. *Étude bibliographique*

Nos travaux, et l'interaction homme/machine en général, sont intrinsèquement pluridisciplinaires. Ils concernent à la fois *l'aspect technique de l'informatique* (le graphisme, les réseaux, les périphériques), les *interfaces utilisateur* et le *travail coopératif médiatisé* (les facteurs humains, les interactions naturelles entre personnes, les éléments sociaux).

Par cette étude bibliographique, nous cherchons à faire un tour d'horizon. Les travaux que nous présentons adressent de nombreux problèmes, pas toujours centraux pour nous mais dont l'ensemble est représentatif des problèmes et des produits existants ou proposés dans ces trois domaines. Nous avons classé ces travaux en trois groupes. Le premier est axé sur le travail coopératif distant, médiatisé, et sur ses contraintes spécifiques. Le deuxième comprend les projets employant les techniques de Réalité Virtuelle. Le dernier rassemble les recherches spécifiques aux interfaces homme-machine et à leur évolution. Ces catégories ne sont pas exclusives et certains projets apparaissent dans plusieurs d'entre elles.

1.1 Le travail coopératif synchrone distant

Petit à petit les habitudes de travail se modifient. Les bureaux sont abandonnés car on préfère intervenir directement chez le client. Le travail chez soi est également en hausse. Cela diminue les

frais (locaux, transports), fait gagner du temps, augmente le rendement de l'entreprise et la qualité de vie des personnes concernées. Les moyens de communication se diversifient et prennent chaque jour plus d'importance : téléphones, répondeurs, fax, messageries vocales, portables, pagers, réseaux informatiques, e-mail... l'important étant de contacter la personne désirée le plus vite possible et de la manière la mieux adaptée à l'activité (discussion, transmission de document, simple information...).

Les infrastructures se développant, les outils s'améliorent. De la vidéo et le son qui l'accompagne peuvent être transmises sur lignes RNIS. Le travail peut s'organiser en sessions synchrones, c'est-à-dire durant lesquelles les personnes impliquées travaillent ensemble sur la même chose. La communication, la concertation, la discussion font partie du processus de création et de production. Votes, brainstormings, design participatifs distants sont techniquement accessibles.

Les outils de travail coopératif³ synchrone sont déjà nombreux. Un tour d'horizon nous a permis de les classer et d'en apprécier la diversité. Nous en avons tiré certains enseignements quant aux qualités et fonctionnalités que devrait posséder un « bon » outil, se conformant à la perception humaine, apte à supporter les conventions sociales...

1.1.1 Les outils vidéo

L'hégémonie des outils audiovisuels

Les produits disponibles dans le commerce permettant de tenir des réunions de travail distantes utilisent tous des techniques et du matériel très classiques : caméras, télévisions ou projection sur grand écran (Figure 1). Les produits se déclinent en de multiples versions, de la simple installation de bureau au dispositif reliant plusieurs grandes salles et pouvant accueillir des dizaines de personnes.

³ Certains préconisent le néologisme « collaboratif », plus large, ne présupposant pas du type d'interaction, du degré d'entraide, d'une « bonne volonté » etc.

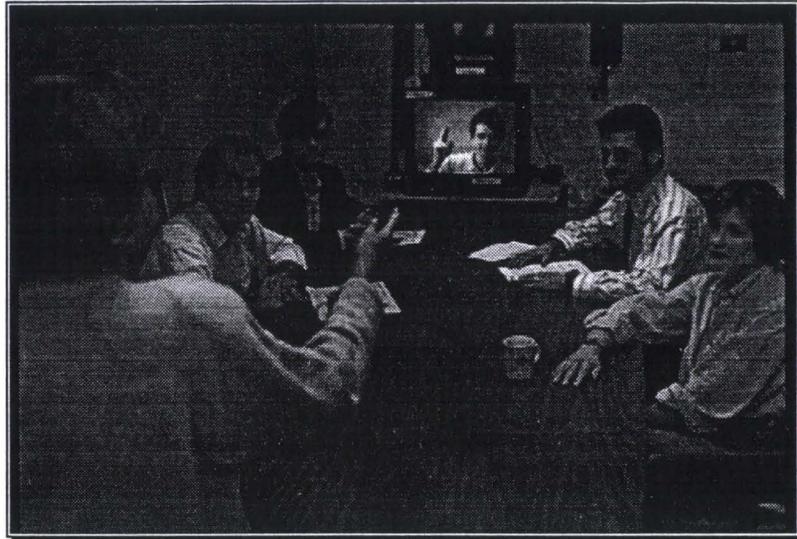


Figure 1 : une télé réunion audiovisuelle

Le système de *vidéoconférence* le plus simple, permettant un face-à-face entre deux personnes, emploie deux postes de travail, chacun étant constitué d'une caméra, d'un écran et d'un microphone. Entre les deux, une liaison spécialisée (simple câble, satellite, réseau téléphonique commuté, lignes RNIS...) assure le transport du son et de l'image entre les deux sites. Il est recommandé [Buxton92] d'utiliser une orientation « portrait » (verticale) des écrans, plutôt que « paysage » (horizontale) pour que les bras de la personne distante, et donc ses gestes, soient visibles, sans pour autant devoir réduire la taille du visage.

Remarquons tout de suite que de tels systèmes ne permettent pas aux interlocuteurs de se regarder directement, dans les yeux. L'explication est la suivante : l'écran et la caméra, ne pouvant être à la même place, sont décalés. Si l'on prend l'exemple d'une caméra posée sur l'écran, la personne filmée aura l'air de regarder sous la caméra. Des solutions, à base de miroirs semi-transparents [Buxton92] existent mais sont rarement utilisées (Figure 2).

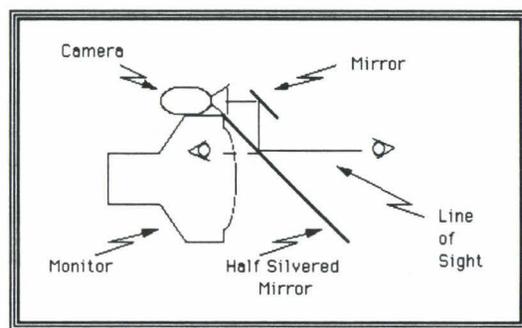


Figure 2 : miroirs semi-transparents pour préserver le contact visuel

Des dispositifs plus complexes existent. Par exemple, il est possible à un professeur de dispenser un cours à plusieurs groupes, au même moment et sans souci de proximité géographique. L'École Nouvelle d'Ingénieurs en Communication (ENIC) possède un tel système [Peters94], relié à plusieurs

sites dispersés dans toute la France. L'orateur, depuis une salle de contrôle spécialisée, est capable de parler aux étudiants, d'être vu par eux, de montrer des documents (papier à l'aide d'une caméra, diapositives, vidéos sur magnétoscope...), d'avoir des vues d'ensemble sur les différents amphis (Figure 3). L'interface de pilotage est entièrement informatisée. De leur côté, les élèves suivent le cours dans un amphithéâtre disposant d'un écran géant et d'un système de sonorisation (Figure 4). Chacun trouve devant lui un microphone et un bouton. Ce dernier permet de demander la parole pour intervenir, poser une question. Le professeur, averti par l'ordinateur, peut alors engager le dialogue. Une caméra motorisée lui permet d'identifier celui ou celle qui a attiré son attention. Plusieurs lignes RNIS et des ponts audiovisuels sont nécessaires pour le transport des divers flux audio et vidéo.



Figure 3 : la salle du professeur à l'ENIC

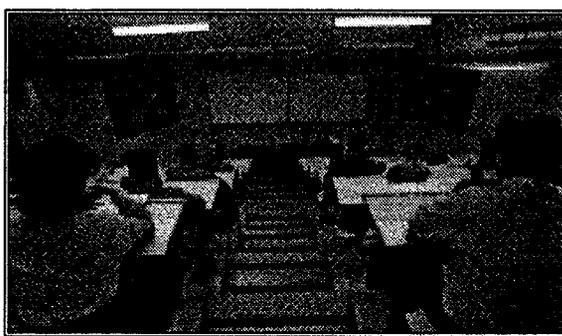


Figure 4 : le « visiocentre » de l'ENIC

Citons aussi les *mediaspaces* [Beaudouin-Lafon96]. L'idée générale est de donner la possibilité aux utilisateurs, par exemple les personnes travaillant dans un même bâtiment, de jeter un œil à tout moment dans les autres bureaux et, éventuellement, de dialoguer (avec la vidéo) avec l'un de ses collègues. Le principal apport est que cet outil est toujours disponible et est, par là même, propice aux relations spontanées et informelles. Techniquement, chacun dispose d'un écran, d'une caméra et d'un module de contrôle (souvent un ordinateur). Un dispositif se charge du routage de tous les signaux audiovisuels. Pour préserver la vie privée de chacun, il est possible de déclarer que son bureau est « fermé » et inaccessible de l'extérieur, les personnes cherchant à « venir » étant alors averties. Des mécanismes comme les *portholes* de Xerox (Figure 5, où l'on remarque une porte fermée) permettent

d'avoir une vue d'ensemble des autres lieux en montrant une mosaïque de vues prises à intervalles réguliers.

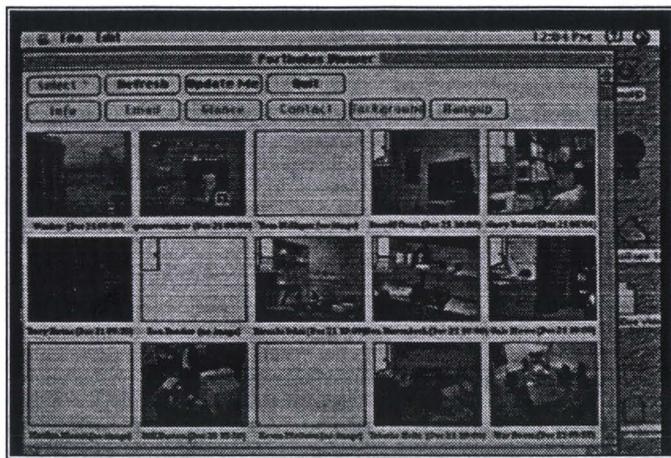


Figure 5 : les *portholes* de Xerox

Les outils audiovisuels informatisés

Les outils audiovisuels ont été transposés à l'informatique (Figure 6). L'image et le son sont alors entièrement numériques. Une caméra, couplée à une carte d'acquisition vidéo, fournit l'image. Un convertisseur analogique/digital permet d'enregistrer le son capté par un micro. Les transmissions se font à travers des réseaux informatiques classiques ou même par modem en point à point à travers le réseau téléphonique. Des CODEC formatent le flux audiovisuel et assurent son transport. Généralement, pour minimiser les flux d'informations à émettre et à traiter, on fait appel à des algorithmes de compression dédiés à l'image et au son. Le plus souvent un matériel spécialisé prend en charge cette étape, qui nécessite de gros calculs effectués en temps réel.

Il est ainsi maintenant possible, de son bureau, sur son poste de travail individuel, de tenir une réunion avec un ou plusieurs interlocuteurs. Ce mode de communication est connu sous le nom de *visioconférence* ou « *desktop conferencing* ». Le deuxième est plus large et ne présuppose que de l'utilisation d'un ordinateur, le premier évoque en plus le transfert de vidéo. L'offre est importante, de nombreux produits sont disponibles sur le marché, commercialisés par plusieurs sociétés : Factory Systèmes (On LAN et On WAN), ZDS/PictureTel, CUSeeMe, EGT/Intel (ProShare), SAT (Meet-me PC), Creative Labs (Sharevision), Matra (Plug and See) etc. La « *série 101 d'introduction aux systèmes de visioconférences PictureTel* » de PictureTel Corporation (1994) expose, dans plusieurs documents courts, les systèmes de visioconférence : choix d'un système, utilisations possibles, maintenance, formation des utilisateurs...

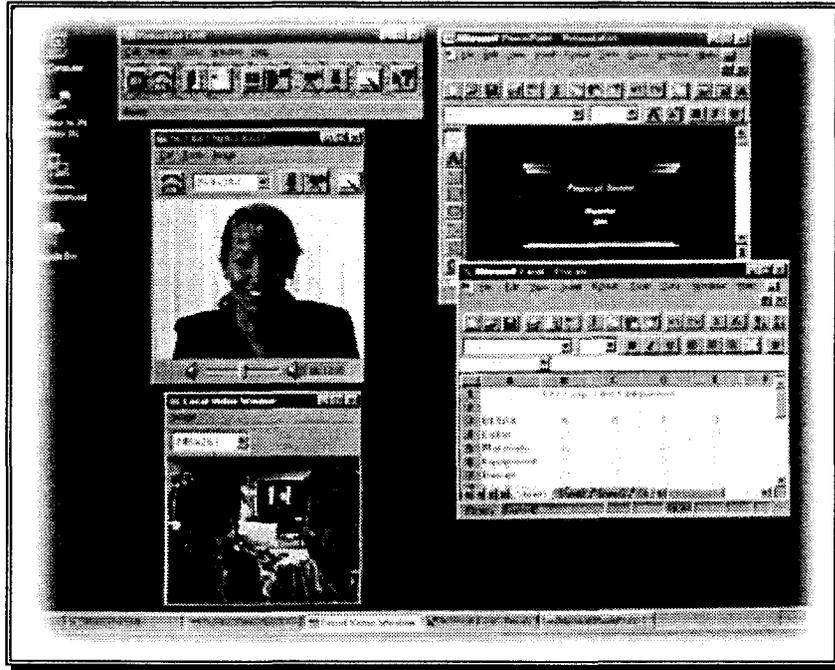


Figure 6 : un outil de visioconférence

Mais ces outils qui utilisent l'ordinateur font mieux que copier les produits audiovisuels simples. Ils leur ajoutent la possibilité d'utiliser en groupe des applications informatiques classiques (rédaction de textes, tableur, design...). Les échanges ne se limitent plus à une simple conversation. Il est maintenant possible de produire à plusieurs des documents, de les enregistrer pour les retravailler par la suite. L'activité, par la sauvegarde de ces documents, acquiert une persistance que ne permettent pas de simples outils audiovisuels. Étant donné que nous nous intéressons principalement à la présentation des outils coopératifs, nous ne détaillerons pas ici les très nombreux travaux visant à partager des applications existantes ou à en créer de nouvelles conçues spécifiquement pour le travail de groupe.

L'attrait pour la vidéo

Les nombreux prototypes et produits commerciaux, utilisant des technologies analogiques ou numériques, ont démontré la faisabilité technique du desktop conferencing. Le son, la vidéo et les réseaux d'ordinateurs peuvent donner la possibilité de s'affranchir des distances entre collaborateurs.

La grande diffusion de ces outils collaboratifs multimédia est freinée par le coût des équipements nécessaires (réseaux très performants, extensions multimédia), ainsi que l'incertitude quant aux bénéfices à tirer de tels investissements. Des études [Williams77] n'ont pas fait ressortir l'apport de la vidéo à laquelle on aurait pu s'attendre. Il semble même que l'image soit superflue du moment que les utilisateurs disposent d'un canal audio et que ni le temps de communication ni le comportement des utilisateurs ne sont modifiés de manière significative par l'existence ou non d'un canal vidéo [Ochsman74].

La collaboration médiatisée par ordinateur a été spécifiquement évaluée [Gale90]. Une comparaison de trois situations a été effectuée sur des tâches expérimentales : partage de documents uniquement, puis partage et canal sonore et enfin partage avec canaux sonore et vidéo. Les mesures se sont faites sur la qualité des résultats obtenus et sur les temps mis pour achever les tâches. Il est apparu que les trois conditions possibles fournissaient des résultats comparables. Il a néanmoins été constaté que, plus les moyens de communication étaient riches, plus les utilisateurs avaient l'impression d'une grande productivité.

Il semble quand même que la présence de la vidéo aide à faire oublier l'outil de communication, laissant les utilisateurs se concentrer sur la tâche à accomplir [Smith89]. C'est un argument que l'on retrouve dans les documentations de présentation PictureTel de la série « 101 » [PictureTel94] : « *au bout de quelques minutes, l'écran semble s'effacer et vous vous retrouvez face à face et en direct avec une personne réelle* ». Le desktop conferencing ne remplace pas vraiment le face-à-face réel, mais est plutôt utilisé comme un téléphone évolué ou comme une sorte de messagerie électronique [Fish92].

Les résultats cités jusqu'ici, assez pessimistes quant à l'intérêt de la vidéo, ont été obtenus à partir de l'observation de tâches artificielles, éloignées des activités habituelles et réelles des cobayes. A l'inverse, une petite équipe de Sun Microsystems a été suivie dans son travail réel et effectif. Nous allons résumer cette étude, dont la description est très détaillée [Tang92].

Un sondage préliminaire a d'abord été fait sur le système de vidéoconférence classique (un modèle PictureTel) qui était utilisé, dans des salles dédiées, chez Sun. Les questions portaient sur ce que les utilisateurs préféraient dans ce système, ce qu'ils lui reprochaient et les améliorations qu'ils souhaiteraient voir apporter. La plupart ont dit aimer voir leurs interlocuteurs, comme à l'habitude. Beaucoup ont déclaré apprécier les économies de voyage et de temps dues à la vidéoconférence. La nécessité de réserver la salle à l'avance a été très critiquée. La mauvaise qualité du son a également été ressentie comme une gêne véritable, alors que la vidéo a semblé moins importante. Le retard dans les transmissions (0,57s) perturbait également les interactions entre utilisateurs. Le besoin d'un document commun, genre tableau, a été largement exprimé. Des améliorations techniques ont été proposées, par exemple la liaison avec plusieurs sites différents. Quelques rares personnes ont désiré, très fortement, disposer d'applications informatiques partagées. Pour résumer, les personnes interrogées demandaient un système disponible à volonté, doté d'un espace de travail partagé et dont le son serait de grande qualité.

Pendant deux mois, les vidéoconférences, les réunions en face-à-face et une réunion téléphonique ont été enregistrées pour analyse. La vidéoconférence souffrait, par rapport au face-à-face, de trois problèmes : collisions sonores, difficultés pour diriger l'attention des interlocuteurs distants et faible interaction. Le délai de transmission amenait souvent deux personnes à commencer à parler en même temps, les obligeant à s'arrêter pour organiser un tour de parole. Des gestes et des regards peuvent ne pas être perçus. Le manque de vision périphérique et la division d'attention entre la vidéo et l'espace partagé ont été mis en cause. En conséquence, les tours de parole duraient plus longtemps,

les interventions spontanées disparaissaient, la conférence se résumant à une suite d'exposés, sans interactions complexes.

Les utilisateurs préféraient utiliser le téléphone à la place du canal audio du système de vidéoconférence pour éviter de subir le délai de transmission, bien que le son et la vidéo ne soient alors plus synchronisés. Cela confirme l'importance d'un son de bonne qualité et sans délai, quitte à dégrader la qualité de la vidéo. La possibilité de se voir entre participants leur permettait toutefois d'appuyer leur discours par des gestes, de demander la parole sans perturber le locuteur et de voir l'attitude des autres.

Un prototype de visioconférence a été fourni à chaque utilisateur, permettant de contacter une ou plusieurs personnes du groupe pour les inviter à se réunir. Chacun pouvait alors voir ses collaborateurs, son image pour contrôler ce qu'il émettait et une application partagée de dessin et d'annotation sur une image. Les délais audio ont été améliorés par rapport aux performances de la vidéoconférence. La transmission se faisant sur réseaux informatiques, à charge variable, la vidéo était dégradée si besoin était pour préserver le canal audio, indépendant. Le rafraîchissement de l'image, configurable par les utilisateurs, était pré réglé seulement à cinq images par seconde pour économiser de la bande passante.

L'activité des volontaires a été suivie pendant quatorze semaines : trois semaines sans visioconférence, avec les moyens habituels ; sept semaines avec le prototype de visioconférence utilisé sur leur station de travail ; quatre semaines pendant lesquelles le canal vidéo avait été supprimé du prototype. Les coups de téléphone, les messages par courrier électronique, les sessions de visioconférence et les réunions réelles ont été relevés. Des questionnaires ont été remplis par le groupe à différents stades de l'étude.

Le volume de communications interactives n'a pas été modifié par la présence du nouvel outil de desktop conferencing. Celui-ci s'est plutôt substitué à d'autres moyens de communication. Les résultats indiquent clairement que la présence d'un canal vidéo conditionne l'utilisation ou non du prototype de visioconférence. La vidéo favorise les échanges plus personnels, incite aux discussions informelles et aide à savoir si l'on est bien compris par son interlocuteur. En fournissant des informations sur l'environnement de la personne distante et sur ce qu'il était en train de faire, elle rendait acceptables les silences, les interruptions (répondre au téléphone ou à un e-mail...) en les expliquant, à l'inverse du téléphone qui force à demander et émettre des signaux visant à s'assurer que l'autre personne écoute et est toujours là. Les signes de compréhension mutuelle, les gestes naturels passaient bien, la faible fréquence de rafraîchissement n'ayant pas été perçue comme un problème. Le tableau partagé n'a été utilisé que dans quelques occasions mais a contenté les utilisateurs.

Le courrier électronique, pourtant asynchrone, a en partie cédé la place à la visioconférence. Les problèmes qui auraient nécessité de nombreux échanges successifs de messages ont été traités directement en synchrone. Ce résultat indique la préférence des utilisateurs pour des outils possédant

la vidéo : le téléphone, qui suffirait à la majorité des échanges, n'est pas utilisé à la place du courrier électronique. D'ailleurs, les coups de téléphones longs ont eux aussi disparu au profit du desktop conferencing.

La salle de vidéoconférence, pourtant réservée à intervalles réguliers, a été abandonnée. La disponibilité totale de la visioconférence, qui propose pourtant un mode de communication différent, a éliminé le besoin de ces sessions régulières. Les auteurs mettent en gardent que, pour des groupes moyens (cinq-dix personnes), le prototype ne suffirait plus.

La visioconférence se place à cheval entre des outils focalisés, comme le téléphone ou une réunion, et d'autres pour lesquels l'attention de l'utilisateur n'est pas requise en permanence, celui-ci pouvant avoir d'autres activités en parallèle, ses collègues pouvant se rendre compte de sa situation directement grâce à la vidéo. Les conventions sociales d'un face-à-face réel étaient respectées pendant l'utilisation du prototype, les interactions plus faciles et moins artificielles qu'en vidéoconférence (interruption de la personne qui parle, humour, une personne finit la phrase d'une autre...). La direction du regard, à défaut d'être exacte et ne permettant pas le contact visuel, était perçue et influait sur les attitudes des participants, notamment après un temps d'adaptation à l'outil.

Des phénomènes d'écho parasitaient le canal sonore, dont la qualité était médiocre. Des casques étaient disponibles à la place des haut-parleurs, qui auraient éliminé le problème d'écho, mais quasiment tout le monde trouvait cela pénible à utiliser.

L'établissement d'une communication se faisait comme avec un téléphone. Néanmoins, les tentatives d'ouverture d'une réunion se soldaient le plus souvent par un échec. L'impossibilité de savoir à tout moment si une personne est présente, le manque d'une fonction type répondeur enregistreur ont provoqué des réticences quant à l'utilisation systématique du prototype. De telles fonctions existent dans ICQ de Mirabilis⁴, qui regroupe *e-mail*, *chat*⁵ et transfert de fichiers en une seule application : la liste des personnes connectées apparaît en permanence, avec des indications supplémentaires (présent mais éloigné, ne pas déranger...). Le passage quasi immédiat d'un mode asynchrone à un mode synchrone est d'une grande souplesse.

Deux conclusions ressortent de cet article. La première est que les utilisateurs veulent la vidéo. Non seulement elle enrichit la communication mais en plus elle améliore la qualité des interactions entre les participants. C'est une nouveauté, les études précédentes n'ayant considéré que des résultats, obtenus dans des tâches spécifiques. La seconde est que le son doit être de la meilleure qualité possible, sous peine d'anéantir le réseau d'interactions qui apparaît entre les interlocuteurs. Enfin, les auteurs insistent sur le fait que chaque mode de communication (desktop conferencing,

⁴ <http://www.mirabilis.com>

⁵ bavardage

réunion en chair et en os, vidéoconférence, etc.) est spécifique et ne saurait en remplacer un autre. La visioconférence à tout de même l'avantage d'être adoptée facilement, au point de voir resurgir spontanément des comportements collaboratifs et sociaux naturels.

Quelques travaux de recherche

Nous allons maintenant présenter quatre projets de recherche ayant pour but le développement d'outils construits autour de la vidéo et capables d'assurer la communication entre personnes distantes.

i) Les Hydra units

Plusieurs outils utilisant les techniques audiovisuelles classique sont présentés dans [Buxton92], dont le Liveboard, le Videowhiteboard. Nous allons détailler le système utilisant les Hydra (Figure 7). Chaque Hydra est composée d'un petit écran vidéo, d'un haut-parleur et d'une caméra miniature. Durant la télé-réunion, chaque interlocuteur distant est représenté par un Hydra. Les appareils peuvent être disposés comme le seraient des interlocuteurs physiques lors d'une réunion, par exemple autour d'une table. Chaque boîtier, et donc chaque caméra, sera tourné vers la personne réelle. Celle-ci s'adresse alors naturellement à l'un ou à l'autre de ses interlocuteurs. Cette configuration a été conçue pour procurer une impression de réalisme. La voix d'une personne provient de l'endroit où se trouve son image. Une personne s'aperçoit qu'on la regarde grâce aux caméras multiples. C'est le contact visuel (« *eye contact* »).

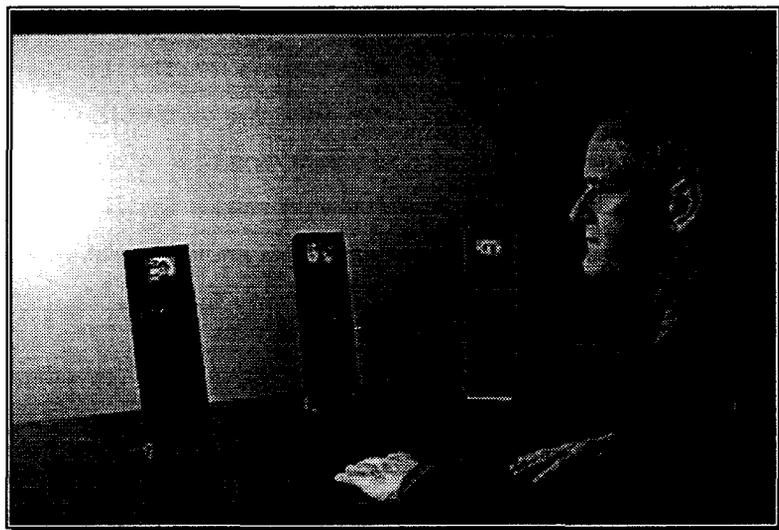


Figure 7 : Bill Buxton converse par l'intermédiaire de trois Hydra Units

Ce dispositif a un côté pratique évident dû à sa petite taille, comparativement à certains des autres systèmes présentés qui nécessitent parfois l'utilisation d'un local dédié. On peut toutefois reprocher aux Hydra de ne donner qu'une impression de présence limitée. Il n'est pas facile, en effet, d'avoir la même sensibilité pour un interlocuteur effectivement présent que pour une de ces « boîtes noires » qui souffrent de leur aspect artificiel. L'écran, s'il permet d'identifier les gens que l'on a « en face » de soi,

est trop petit pour que les expressions sur les visages soient lisibles et déchiffrables. Le fait même de parler à des personnes en miniatures fausse les relations sociales ou hiérarchiques au sein du groupe.

ii) Clearface

Ishii et Arita [Ishii91] utilisent un système, appelé TeamWorkStation (Figure 8), qui ne dispose que d'un seul écran. Ils étudient le problème du placement sur cet écran à la fois d'un tableau partagé et des images vidéos en direct de collaborateurs. Après avoir écarté la « mosaïque » (Figure 9) où chaque fenêtre occupe une zone d'écran propre, sans recouvrements, ils se tournent vers la transparence : les images vidéo seront surimposées au tableau blanc (Figure 10). Une caméra filme le visage, tandis qu'une autre filme les mains sur le plan de travail. Les mains (donc les gestes) apparaissant transparentes au-dessus du tableau. Ils s'inspirent des expériences menées par Neisser et Becklen [Neisser75] qui démontrent la capacité humaine à distinguer deux scènes surimposées et à passer de l'une à l'autre à volonté. L'aptitude naturelle que l'on a à reconnaître les gens par leur visage et par leurs gestes améliore encore les choses. Le suivi simultané des deux plans, par contre, est plus difficile mais est rarement utile dans la situation de travail présentée. Par contre, si les utilisateurs arrivent à faire passer facilement leur attention du document aux vidéos et inversement, ils se sentent gênés pour l'écriture sur le tableau. En effet, ils évitent de dessiner aux endroits où leurs collègues apparaissent. Il apparaît également que donner une taille fixe aux vidéos limite leur nombre maximal à l'écran (quoique les expériences n'aient été menées qu'avec deux ou trois utilisateurs).

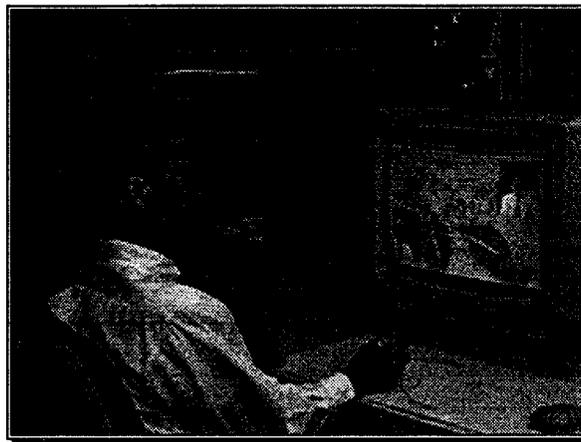


Figure 8 : la TeamWorkStation, avec ses deux caméras

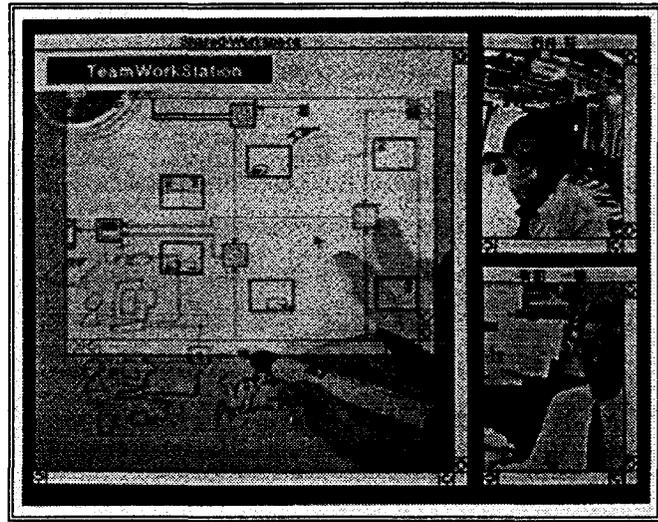


Figure 9 : présentation en « mosaïque » sur la TeamWorkStation



Figure 10 : visages surimposés et mains en transparence

La solution retenue est donc d'offrir le moyen à l'utilisateur de déplacer et de dimensionner les vidéos à volonté. Il semble également souhaitable que les personnes soient filmées devant un fond uni, sinon il devient difficile de faire la distinction entre le dessin du fond et les images des interlocuteurs. Sous ces conditions, le système se révèle utilisable sans que l'utilisation de la transparence occasionne des difficultés particulières. Notons que ce système constitue un premier pas dans les environnements *continus* (« *seamless* ») où documents et personnes sont représentés dans un seul espace homogénéisé. Les visages ne s'intègrent pas au reste et confèrent un caractère encore artificiel à la vue présentée.

iii) ClearBoard

Le ClearBoard est un système de travail pour deux personnes distantes (« pairware » plutôt que « groupware ») avec un tableau partagé. L'originalité est d'avoir adopté comme métaphore non pas une situation autour d'une table ou devant un tableau mais où les deux personnes se parlent à travers une vitre sur laquelle ils peuvent dessiner. Il est intéressant de constater que cette situation n'est jamais exploitée dans le monde réel mais qu'elle se prête bien à la médiatisation. L'intérêt est que le

partenaire est vu à travers le document, et non à côté, ce qui permet de présenter de manière *continue* le document et l'autre personne. Un premier prototype, le ClearBoard 1 [Ishii92] (Figure 11), utilisait un système vidéo (caméras, projecteurs, écrans transparents) pour mélanger le document et l'image de la personne distante. Par rapport à la situation réelle, deux améliorations rendent le ClearBoard exploitable. D'une part, le plan de travail est incliné, et non pas nécessairement vertical. D'autre part, l'image est inversée pour annuler l'effet miroir de la vitre et ainsi fournir des schémas entièrement à l'endroit pour les deux participants. Les gestes sont compréhensibles, le visage est expressif, les deux personnes peuvent se regarder, tout est fait pour rendre naturelle la situation de travail.



Figure 11 : le ClearBoard 1

Le Clearboard 2 [Ishii94] (Figure 12) utilise l'informatique et non plus la vidéo. L'avantage est de pouvoir enregistrer et retrouver plus tard le travail déjà effectué. L'absence de persistance des documents réalisés à deux dans le ClearBoard 1 avait été très critiqué par les utilisateurs et rendaient l'expérience un peu vaine.

L'intégration d'une application informatique et de la vidéo est un progrès par rapport aux outils coopératifs utilisant les systèmes 2D. L'inconvénient de celles-ci était de présenter une collection de fenêtres, chacune contenant une vidéo ou une application, sans que rien ne soit fait pour assurer la cohérence globale ou montrer les liens entre ces fenêtres. Ici l'écran affiche une image *continue* qui contient le document et l'interlocuteur dans un même espace ou, formulé autrement, dans un *même repère*, sans séparation artificielle entre eux. C'est grâce à cela que les gestes et le regard sont si naturels. En comparaison, Clearface réussit à intégrer gestes (désignation) et documents mais les visages des interlocuteurs n'ont aucun rapport, au niveau du placement, avec le reste.

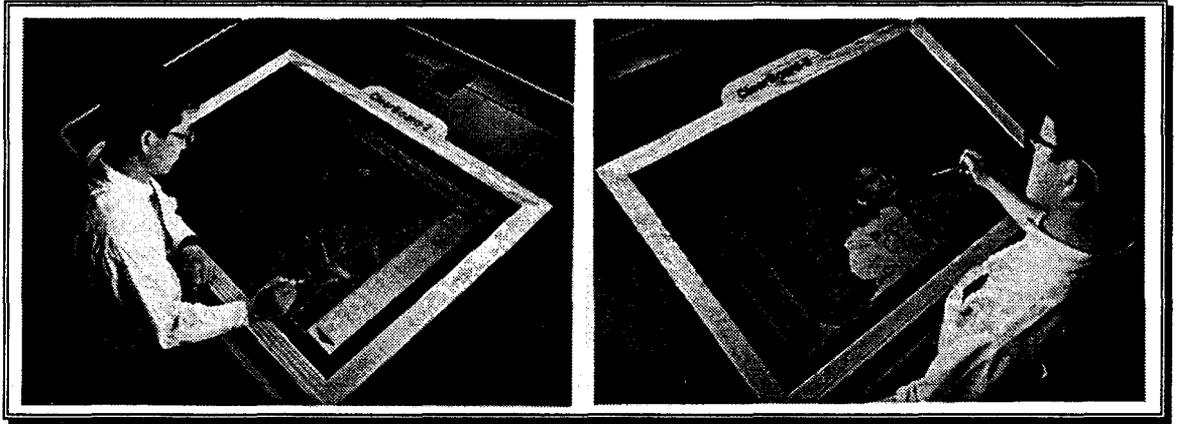


Figure 12 : le ClearBoard 2

iv) MAJIC

Le système présenté dans [Okada94], MAJIC (*Multi-Attendant Joint Interface for Collaboration*), se propose de permettre à plusieurs personnes (4 au maximum) dispersées géographiquement de tenir une réunion « de travail ». Le principal objectif est de simuler une vraie réunion, de la manière la plus réaliste possible. A cela s'ajoute le souci de respecter certaines conventions sociales particulièrement chargées de sens au Japon (la distance entre les gens, le poids du regard).

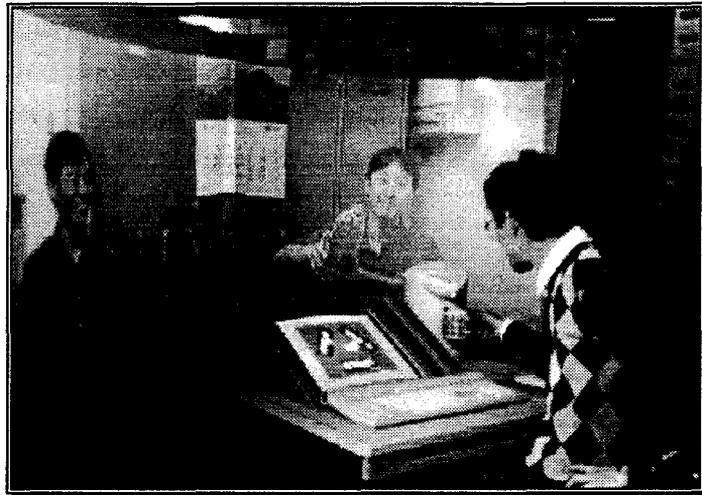


Figure 13 : MAJIC

L'utilisateur est partiellement entouré par un écran géant courbe (2m×4m) sur lequel sont projetés ses interlocuteurs en grandeur réelle (Figure 13). Derrière l'écran, semi-transparent, un ensemble caméra + haut-parleur est placé de telle manière que :

- quand je regarde un interlocuteur, il en a conscience – deux personnes peuvent se regarder dans les yeux, le *contact visuel* (« *eye contact* ») est préservé ;

- quand je regarde quelqu'un, les autres le savent en suivant la direction de mon regard : c'est la *conscience du regard* (« *gaze awareness* ») ;
- quand un intervenant parle, sa voix provient de l'endroit où son image apparaît. Autrement dit, voix et image ont la même position dans l'espace.

De plus, grâce à l'écran de grande taille :

- les expressions du visages sont lisibles ;
- les gestes, composante importante du dialogue, sont retransmis ;
- l'affichage « grandeur nature » renforce le réalisme et la sensation de présence. De surcroît, il peut être socialement important de respecter la taille des personnes que l'on a en face de soi.

Ajoutons que les personnes distantes sont « détournées » avant d'être projetées, ce qui permet de les intégrer à un décor unique pour rendre la scène plus homogène. Comme pour le ClearBoard, on retrouve le souci de *continuité* : les personnes ne sont pas représentées indépendamment mais ensemble, ce qui correspond bien au fait qu'elles ont une activité en commun. C'est d'autant plus important dans MAJIC que tout est fait pour montrer les signes de cette activité commune (regard...).

Les informations sont émises sur des canaux vidéo et audio classiques. Leur nombre dépend de celui des personnes participant à la réunion et influe sur le coût d'installation et d'exploitation de MAJIC.

1.1.2 Les « environnements virtuels collaboratifs »

Les techniques dites de « Réalité Virtuelle » visent à simuler sur ordinateur un monde artificiel mais semblable par ses propriétés au monde réel. L'une des applications les plus prometteuses est de faire partager un même monde synthétique par plusieurs intervenants humains. Dans ces CVE⁶, terrain propice au travail coopératif de groupe, les utilisateurs se retrouvent ensemble malgré leur éloignement physique et sont plongés dans un environnement sans surprise, puisqu'il reproduit et simule la réalité. Il est donc dans la nature même des CVE de supporter le travail collaboratif, pour peu que les outils proposés soient bien conçus et que la simulation soit de bonne qualité.

Nous allons parler succinctement ici de DIVE et de MASSIVE, deux mondes virtuels gérant des groupes importants de personnes. La Réalité Virtuelle, qui est avant tout un ensemble de nouvelles technologies, sera développée en détail à la section 1.2.

⁶ *Collaborative Virtual Environment*

DIVE

Ce qui fait l'originalité et l'intérêt principal de DIVE (Distributed Interactive Virtual Environment), c'est qu'il constitue une mise en œuvre expérimentale d'un modèle spatial d'interaction. L'idée derrière ce modèle est de laisser les utilisateurs agir entre eux comme ils le feraient naturellement, c'est-à-dire se regrouper pour tenir des réunions ou simplement discuter [Benford93b]. Tout d'abord, chaque utilisateur possède une *aura* (Figure 31). C'est une portion de l'espace qui entoure l'utilisateur, dans laquelle il sera considéré comme présent et, réciproquement, au-delà de laquelle il ne percevra pas le monde dans lequel il évolue. L'intérêt est de limiter le nombre d'objets et de personnes avec lesquels un utilisateur peut entrer en interaction. Le modèle confère également à chacun un champ de présence, appelé *nimbus* (traduisant à quel point les autres participants auront conscience son existence), et un champ de perception appelé *focus* (représentant l'attention qu'il porte à ce qui l'environne). En fait, il existe plusieurs focus et nimbus, autant que de médias gérés par le système. Par exemple, le champ visuel sera limité en angle alors que le champ sonore sera sans direction particulière.

L'existence de ces champs permet directement à toute personne d'aller écouter un autre participant en s'en approchant jusqu'à ce que le recoupement du champ de perception de l'auditeur et du champ de présence du locuteur soit suffisant. Réciproquement, ce dernier peut voir son attention attirée par l'auditeur si ce dernier vient se placer de manière à être perçu. Groupes et sous-groupes se forment implicitement, dynamiquement. Il n'y a pas besoin d'un mécanisme supervisant la mise en place et le déroulement d'activités, tout est géré par les participants eux-mêmes.

L'interaction ne se limite pas à des échanges entre personnes. L'univers de DIVE contient également des outils qui eux aussi possèdent une aura. Pour utiliser un outil il suffit de s'en approcher assez près pour que DIVE ouvre un canal d'interaction entre l'utilisateur et l'outil. Par exemple, pour lire et écrire sur un tableau, il suffit de se placer à proximité dans son aura. Certains outils comme la table de réunion ou le porte-voix servent de relais entre les utilisateurs ou modifient leur aura. Toutes les personnes qui viendront se placer autour de la table pourront s'entendre, la table étendant leur champ de perception. Quant au porte-voix, il va étendre très largement le champ de présence (sonore) de la personne qui le possède, laquelle pourra se faire entendre à une distance bien supérieure à celle couverte par son nimbus et son aura normaux.

Différents lieux peuvent être définis dans DIVE. Le passage d'un lieu à un autre se fait par téléportation : le lieu de départ est simplement remplacé par le lieu d'arrivée. Des portes constituent le passage d'un lieu à un autre. L'espace peut donc être organisé en créant des pièces spécialisées comme une salle de réunion, un bureau d'accueil chargé de diriger les nouveaux arrivants, etc.

L'architecture ouverte de DIVE en fait une plate-forme de choix pour des expérimentations en Réalité Virtuelle collaborative. Comme nous l'avons vu, le travail sur les avatars est très important. Les périphériques d'entrée gérés sont nombreux. Les applications ne sont pas limitées à des rencontres

entre personnes distantes, mais s'étendent jusqu'à la robotique, par exemple des interfaces de contrôle de robots [Simsarian97].

MASSIVE

MASSIVE est un projet proche de DIVE : le même modèle spatial est utilisé pour l'interaction entre personnes. L'implémentation qui en est faite est par contre différente. MASSIVE entend démontrer que ce modèle peut être adapté à des interfaces très différentes, permettant l'interaction entre personnes équipées différemment. Un autre objectif de ce projet est d'être capable de gérer des grands groupes d'utilisateurs, des centaines voire plusieurs milliers.

Les utilisateurs se scindent en deux groupes principaux. L'un est constitué de ceux disposant d'une machine capable de générer des graphiques 3D en temps réel, sur écran ou dans un dispositif immersif. L'autre accueille les utilisateurs travaillant en mode texte (terminal dédié ou fenêtre contenant ce texte). Notons que cela signifie que tout le monde peut utiliser MASSIVE et qu'il sera effectivement possible de mener des expérimentations sur des grands groupes. Un autre paramètre est celui de la possibilité pour la station de travail d'émettre du son. Des messages sonores peuvent être émis par le système. L'utilisateur peut aussi entendre réellement ses interlocuteurs. A défaut, MASSIVE contient des outils de *chat* qui sont en fait des tableaux partagés prévus pour accueillir du texte. C'est typiquement dans ce style « IRC⁷ » que les utilisateurs en mode texte s'expriment.

En accord avec le modèle spatial, les différentes interfaces proposées ont en commun de pouvoir transcrire explicitement la position et l'orientation de chacun et de permettre à tous les utilisateurs de se déplacer dans le monde virtuel géré par MASSIVE.

1.1.3 Les outils « informatiques »

Genesys

La société Genesys est spécialisée dans la conférence⁸ par téléphone. L'organisateur de la réunion, par un simple numéro de téléphone suivi d'un code secret, ouvre une salle qui va accueillir les autres participants. Ceux-ci appellent à leur tour au même numéro, à l'heure convenue de la réunion. L'organisateur a des droits particuliers : faire sortir un intrus, fermer la salle ou clore la réunion. Hormis sa simplicité d'utilisation (pas de réservation, pas d'opérateur), ce service n'est pas novateur.

⁷ Internet Relay Chat, une application en mode texte où tous les utilisateurs connectés à un canal particulier peuvent s'envoyer des messages en mode quasi synchrone et transmettre une description succincte de leurs émotions.

⁸ Il s'agit bien ici de réunion, donc uniquement de communication entre des personnes. Ce système offrant un service téléphonique ne supporte bien entendu pas de documents partagés.

Ce qui fait l'originalité des téléconférences de Genesys, c'est le « MultiConference Manager », un logiciel qui permet à l'organisateur de piloter la téléconférence (Figure 14). L'écran est partitionné en quatre salles : la salle de réunion, le hall d'entrée, le salon d'accueil et le salon d'aparté.

Les personnes qui arrivent, signalées par un signal sonore, attendent dans le hall d'entrée : la liste de leurs noms s'affiche dans cette salle. Ils y restent tant que la porte entre le hall et la salle de réunion est fermée. Le salon d'accueil permet à l'organisateur de prendre à part l'un des nouveaux arrivants avant la réunion. Dans la salle de réunion proprement dite, chaque personne est représentée par un dessin (homme ou femme) derrière un bureau équipé d'une diode (Figure 15). Le nom de chacun apparaît sur le devant du bureau. La diode clignote en vert lorsqu'un participant prend la parole. L'organisateur peut emmener une ou plusieurs personnes dans le salon d'aparté pour discuter en privé pendant la réunion. Une barre d'icônes offre encore d'autres fonctions à l'organisateur : empêcher les autres de parler, leur permettre quand même de demander de parole (par la touche « * » du téléphone : l'avatar, derrière son bureau, lève alors la main et la diode clignote en rouge), un outil de vote (chacun choisit par les touches 0–9 du téléphone) avec affichage des résultats.

Le MultiConference Manager ajoute de la dynamique par rapport à une audioconférence classique. Les tours de paroles se négocient rapidement, le contrôle est facilité, on saisit mieux la situation, des outils automatisent des tâches pénibles par téléphone (le vote). C'est d'autant plus réussi que le matériel requis est à la portée (voire entre les mains) de tous : deux lignes téléphoniques, un petit PC (4 Mo de RAM) sous Windows, un modem lent (9600 bauds). C'est un produit commercial modèle, qui mêle une technologie simple et éprouvée (le téléphone) à une aide visuelle simple mais efficace.

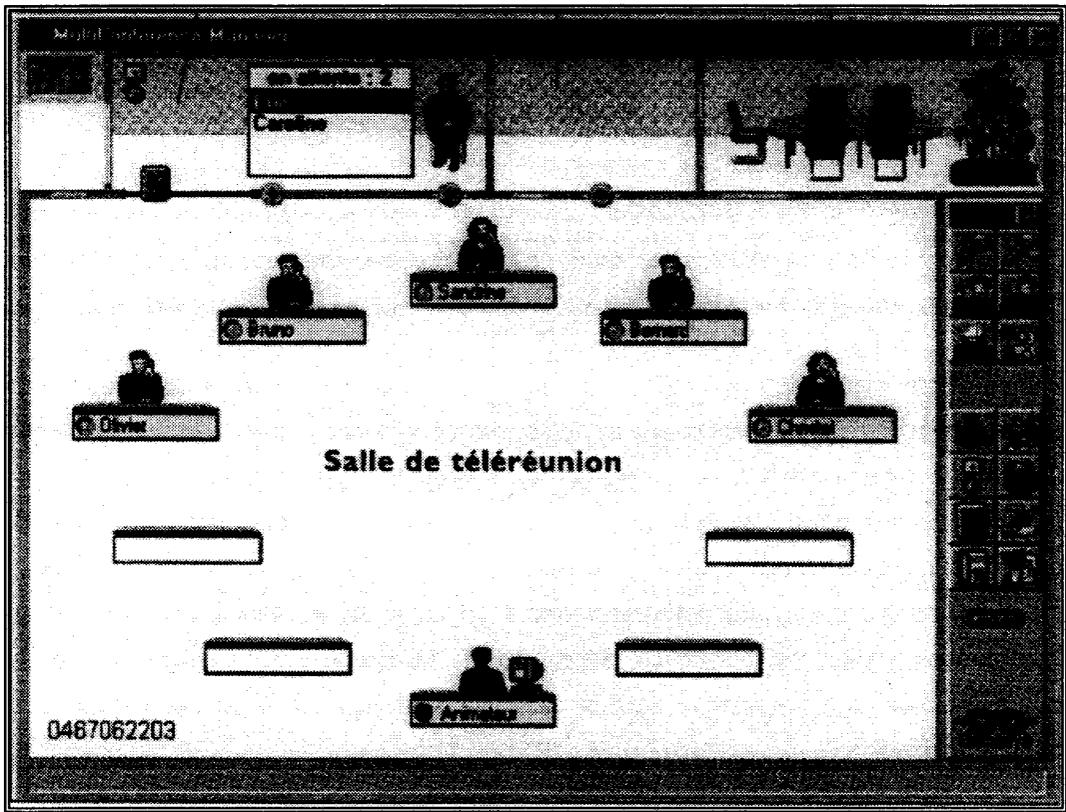


Figure 14 : le MultiConference Manager de Genesys



Figure 15 : une interlocutrice

Teledesign

Les créateurs, designers, architectes et artistes montrent une préférence pour les espaces de travail ouverts, accessibles à leurs collègues, plutôt que des bureaux privés à l'abri d'interruptions et d'intrusions. Teledesign [Shu94] a été développé spécifiquement pour étudier les méthodes de travail et les préférences des utilisateurs pour une application particulière, la CAO⁹ 3D en groupe. Une première version a servi à comparer deux méthodes d'accès à l'objet partagé, soit à tour de rôle, avec prise de contrôle implicite, soit de manière concurrente, les conflits étant gérés en annulant (en

⁹ Conception Assistée par Ordinateur

défaisant) les actions inconsistantes. Un système d'édition graphique tourne en local sur chaque machine, une couche de communication assurant les communications pour mise à jour entre les différents postes. Une seule fenêtre montrant l'objet créé s'affiche à l'écran, avec possibilité de visualiser les axes. Les manipulations s'effectuent en 2D, à la souris. L'objet peut être projeté en perspective sur l'écran, ou orthogonalement à un axe du repère pour aligner les composantes avec précision. La création et la modification de sous-objets se font par menu ou au clavier. Cet éditeur se veut le plus simple possible à manipuler et à prendre en main par les nouveaux utilisateurs.

Après une phase d'adaptation à l'outil, séparément puis à deux, les cobayes sont séparés pour se parler sans se voir. Des tâches, plus ou moins libres, leur sont alors confiées. Il est rapidement apparu que chacun doit avoir une vue propre sur l'objet, même si les deux personnes travaillent sur un même objet : point de vue, type de projection, taille de la fenêtre, variables d'état de l'éditeur (mode déplacement ou mode redimensionnement par exemple)... Ces points de vue indépendants posent des problèmes, en particulier pour faire référence à un objet ou un point précis de l'espace. Les positions décrites verbalement le sont par rapport à une vue précise, non partagée et n'ont pas de sens pour le collègue. Le vocabulaire déictique du langage courant (ici, à gauche, devant moi...) est inadapté à cette situation.

Toutes les personnes interrogés ont préféré le mode d'édition simultané au mode n'acceptant qu'une action à la fois, à tour de rôle. Ce mode séquentiel ne permet pas de scinder le travail à faire en deux, chacun travaillant à une partie de son côté et donne l'impression de perdre son temps sans justification. Dans le mode concurrent, les collisions (modifications contradictoires) observées sont très rares. Le contact verbal permet la prévention et la résolution rapide de ces collisions. Les comportements enregistrés se conforment aux usages sociaux, chacun respectant l'autre, prenant en compte la hiérarchie et essayant de se montrer coopératif. Si cette situation est satisfaisante à deux, un grand groupe serait probablement plus difficile à gérer rien que par la parole et le mode d'accès libre deviendrait probablement moins efficace, nécessitant la mise en place de mécanismes de contrôle plus directifs.

La deuxième expérience consiste à faire apprendre à un novice par un utilisateur déjà expert à se servir de Teledesign, les interactions lors de l'apprentissage étant différentes d'un travail collaboratif classique. Le besoin d'un *télépointeur*, c'est-à-dire un pointeur visible sur plusieurs affichages différents à la fois, a surgi immédiatement. Son rôle le plus important serait de montrer l'objet dont on parle, rendant ainsi possible des phrases du style « pour changer la forme de cet objet selon cette dimension, il serait plus facile de sélectionner ce coin-ci en se plaçant dans cette vue », truffées de mots déictiques de position. Les pointeurs de Teledesign étant 2D (pilotés à la souris), transmettre leurs coordonnées n'aurait servi à rien, il aurait été impossible de les projeter dans une vue 3D différente. Intuitivement, les novices pensent que le maître voit leur pointeur. Quand ils se rendent compte que ce n'est pas le cas, ils détournent le système pour désigner un objet, soit en le prenant puis en l'agitant, soit en créant un nouvel objet « bidon » près de l'objet qu'ils veulent désigner.

L'impossibilité de partager les points de vue ou tout au moins de communiquer son point de vue gêne les explications. Expliquer l'intérêt de passer de la vue perspective à une vue orthogonale est difficile sans pouvoir montrer la différence. A l'inverse, le maître, ne connaissant pas la vue du novice, comprend mal les erreurs commises et les difficultés rencontrées par celui-ci. Le WYSIWIS¹⁰ strict (la même fenêtre exactement est affichée chez les deux participants) pourrait résoudre ce problème, en particulier dans cette relation novice/maître.

Teledesign a été modifié pour résoudre le problème de la connaissance de la vue de l'autre. La solution implantée consiste non pas à partager la vue (l'image à l'écran) d'une autre personne mais à pouvoir se rendre compte de quel est son point de vue sur l'objet 3D. Une petite pyramide colorée, appelée *viewpoint*, est placée dans l'espace à l'endroit où est placée la personne distante relativement à l'objet observé, la pointe dirigée dans la direction où elle regarde. Le pointage, i.e. la désignation, est visualisé par une ligne qui va de la pointe de la pyramide au point sélectionné (Figure 16).

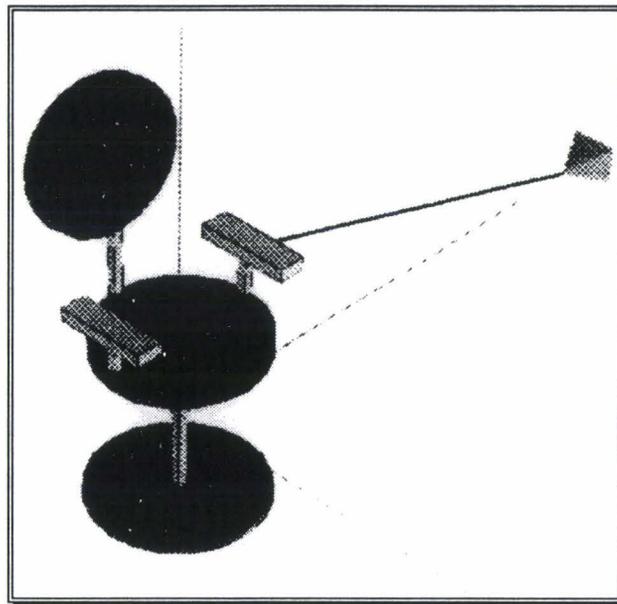


Figure 16 : le *viewpoint* avec la ligne matérialisant le point d'action

Ce *viewpoint* est un moyen simple de connaître, dans le cadre précis d'une application de CAO, le point de vue d'une personne ainsi que l'objet qu'elle pointe. Des développements à apporter pourraient s'attacher à représenter d'autres aspects comme l'activité d'un utilisateur. Les auteurs conseillent de s'attacher aux possibilités de l'outil avant tout, la représentation réaliste des utilisateurs ne leur semblant pas primordiale pourvu que les informations importantes soient transmises de manière satisfaisante.

¹⁰ What You See Is What I See, ce que tu vois est ce que je vois.

CSpray

CSpray [Pang97] s'attaque au problème de la visualisation scientifique de données par plusieurs utilisateurs simultanément. Il constitue une version « collaborative » de Spray [Pang94]. Chacun doit pouvoir sélectionner le point de vue qu'il désire tout en étant conscient de ceux des autres. La solution proposée est proche de celle développée dans Teledesign. Ici, un observateur est représenté par un œil coiffé d'un chapeau et sur lequel apparaît un nom en surimpression, écrit normalement (i.e. ne subissant ni rotations ni autres transformations purement 3D). L'ensemble est nommé *EyeCon* (Figure 17).

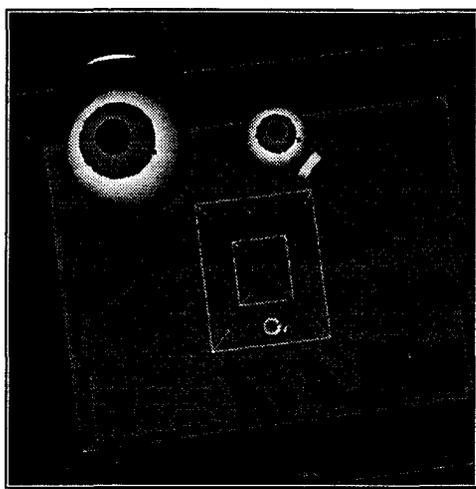


Figure 17 : EyeCons

Bien que très simple, cette représentation permet de connaître la position d'une personne dans l'espace 3D, de son orientation (grâce à la pupille et au chapeau placé sur le « dessus »). Le nom écrit de manière toujours lisible (contrairement à MASSIVE par exemple) permet une identification rapide.

La désignation se fait à part, par l'intermédiaire de bombes (*spray can*) desquelles sort une flèche, un faisceau dirigé vers les données désigne une zone intéressante. Elle peut être commentée, légendée par son propriétaire. Ladite zone peut également être colorée à la bombe pour la marquer et la repérer ultérieurement (Figure 18).

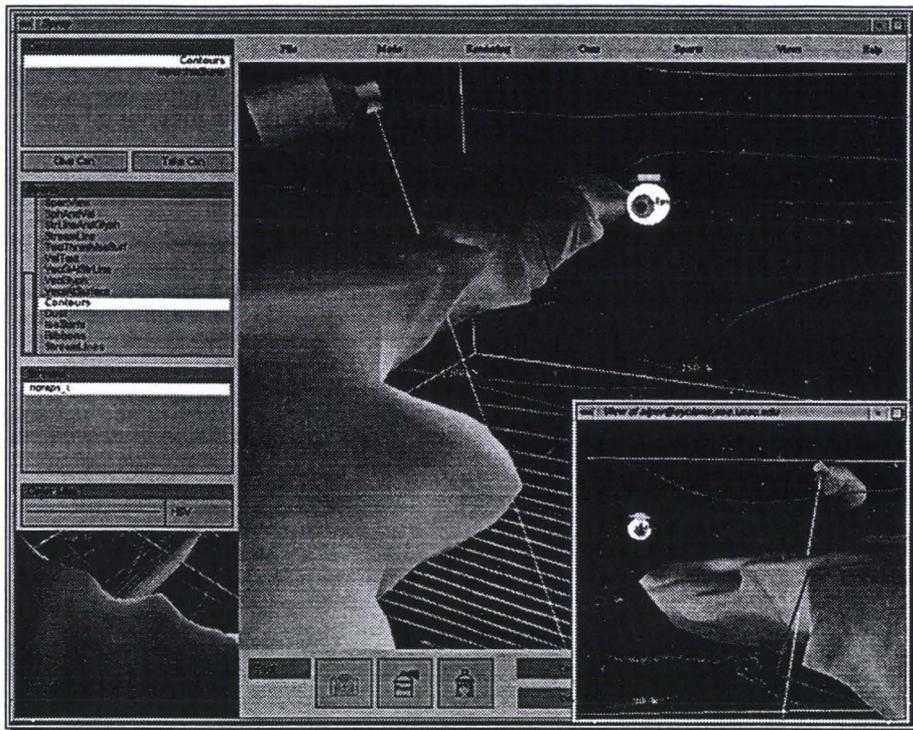


Figure 18 : EyeCons, Spraycans et vues séparées

Il est possible d'adopter le point de vue d'un autre participant soit en cliquant sur son EyeCon, soit en sélectionnant son nom dans la liste des personnes « présentes ». La fenêtre principale est à caractère privé. Quand le besoin s'en fait sentir, une fenêtre publique peut être ouverte. Elle va refléter chez tout le monde la vue qu'a une personne particulière. Elle sert quand un orateur privilégié doit s'adresser à tous, tandis que les autres vont juste regarder. Chacun pourra suivre ses déplacements (changements de point de vue) et les zones qu'il va marquer ou désigner.

On retrouve donc dans CSpray une solution similaire à celle de Teledesign pour le partage de points de vue. En plus, le « point d'action » est découplé du « point de vue » (selon la terminologie de Benford [Benford95]) et il est possible d'avoir un vrai mode WYSIWIS dans une fenêtre secondaire tout en gardant sa vue propre dans la fenêtre principale.

L'environnement multimédia interactif AT&T

Le projet décrit dans [Seligmann95] intègre quatre éléments : un service de salles de réunion virtuelles, un service de son spatialisé, un service de partage d'applications et une visualisation 3D. L'accent est mis sur la mise en place de ce que les auteurs appellent des « assurances », c'est-à-dire des indices permettant de se rendre compte du bon fonctionnement de l'outil de communication (bruit de fond) et de l'attention que nous portent les autres personnes (« Mmmhh... » au téléphone). La représentation graphique rend compte plutôt de la présence tandis que le son donne une idée de l'activité en cours.

Sur la Figure 19, on peut voir deux personnes « fantomatiques » dans leurs bureaux, le matériel dont elles disposent (ordinateur, téléphone). Leur présence n'y est que suggérée. En effet, une pièce supplémentaire, virtuelle, a été créée (elle flotte au-dessus des bureaux) pour accueillir ces deux personnes, opaques pour signifier leur présence effective.

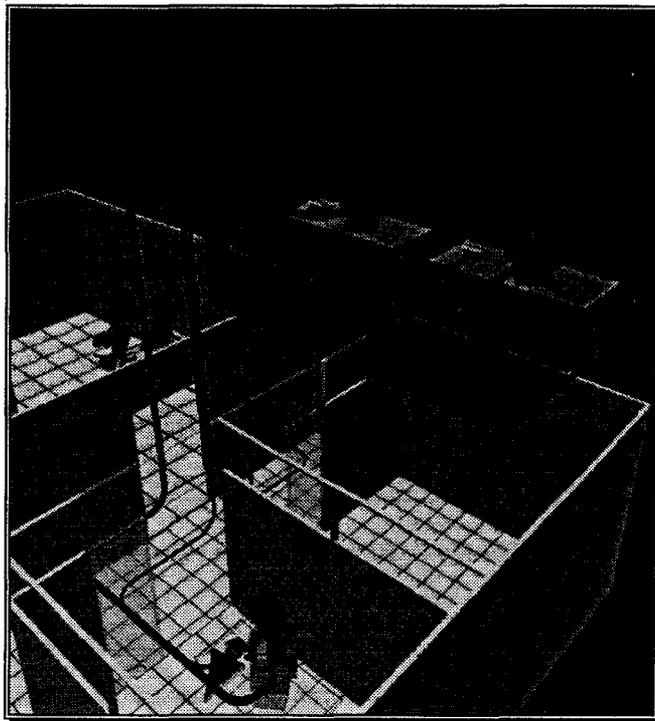


Figure 19 : deux personnes se rejoignent dans une pièce virtuelle

Une personne passant à proximité sera perçue, grâce au son, par les deux occupants de cette nouvelle pièce, qui pourront l'inviter à les rejoindre.

Sur la Figure 20, on peut voir une réunion à trois. Les deux interlocutrices sont visibles ainsi que, sur la table, une application partagée. La vue est construite pour satisfaire le point de vue de l'observateur (par exemple, la fenêtre est orientée de manière lisible). Chaque voix est placée dans l'espace à la même place que l'image de la personne qui l'émet. Des sons sont générés artificiellement, spatialisés eux aussi : touches de clavier quand quelqu'un tape, frappe de machine à écrire près du document. Quand la personne de droite désigne le document, une flèche représente cette action.

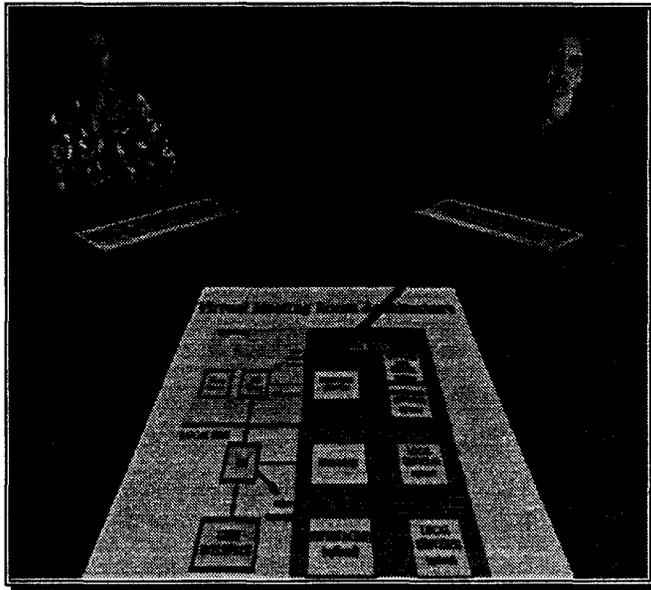


Figure 20 : l'utilisateur voit ses interlocutrices et un document

Les vues d'une salle virtuelle générées par le système pour les différents utilisateurs sont cohérentes. Les personnes à qui l'on fait face apparaissent sous forme de photographies, de vidéos ou simplement par l'affichage de leur nom. Les services fournis s'adaptent localement aux ressources dont dispose chaque utilisateur (par exemple, un écran fenêtré remplace la vue 3D, Figure 21).

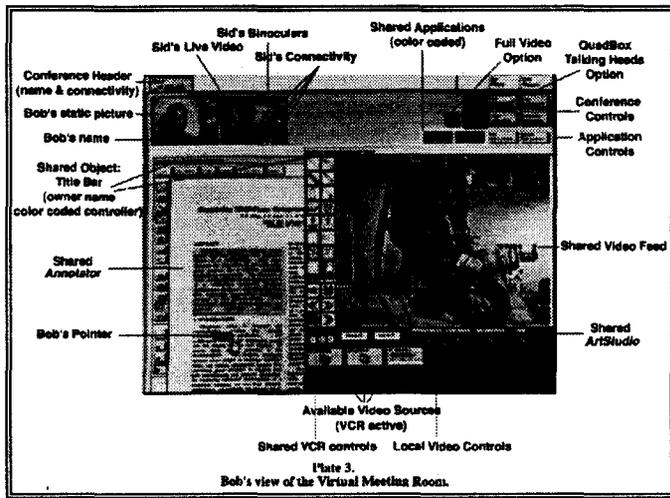


Figure 21 : l'activité vue dans des fenêtres

Les différents services proposés fournissent de nombreuses informations. L'utilisateur peut anticiper ce qu'il va rencontrer, par exemple par le son qui provient d'une salle. Il peut identifier les personnes, les lieux par le son, le décor. Son et image sont associés. Des objets similaires peuvent être distingués. La persistance des salles facilite la mémorisation des situations. Les positions relatives des objets sont globalement cohérentes.

1.2 La Réalité Virtuelle

Le terme « Réalité Virtuelle » (R.V.) englobe toutes les techniques visant à simuler un monde réaliste sur un ordinateur. C'est un terme imprécis, qui regroupe des produits faisant appel à la synthèse d'images 3D temps réel, au son spatial, à l'immersion, à la simulation, à l'interaction directe, cette liste n'étant pas exhaustive et ce cadre n'étant pas restrictif. Les travaux les plus anciens et certainement les plus nombreux et importants ont porté sur la réalisation de simulateurs, militaires en particulier. L'entraînement au pilotage de véhicules de combats (avions, blindés...) n'a d'intérêt que s'il est réaliste (images, interactions, etc.). Les qualités pédagogiques proviennent de l'habitude à manipuler l'appareil que l'on acquiert ainsi qu'à la grande variété des situations face auxquelles l'élève peut être confronté. Les algorithmes et les machines ont été améliorés pour permettre la génération d'images de qualité en temps réel. Heureusement, il est devenu possible de développer des simulateurs dans d'autres domaines avec la baisse du prix et la montée en puissance des machines. Les scientifiques, la médecine, la robotique, l'automobile et de nombreux autres secteurs industriels utilisent les simulateurs pour développer de nouveaux produits sans passer par des maquettes réelles. De plus, un simulateur peut être directement manipulé par un utilisateur, ce qui permet de tester l'ergonomie d'un outil. La Figure 22 présente un engin d'exploitation des forêts, capable de couper et débiter un arbre sur place. L'ergonomie de son poste de pilotage a été testée puis améliorée grâce à un simulateur.

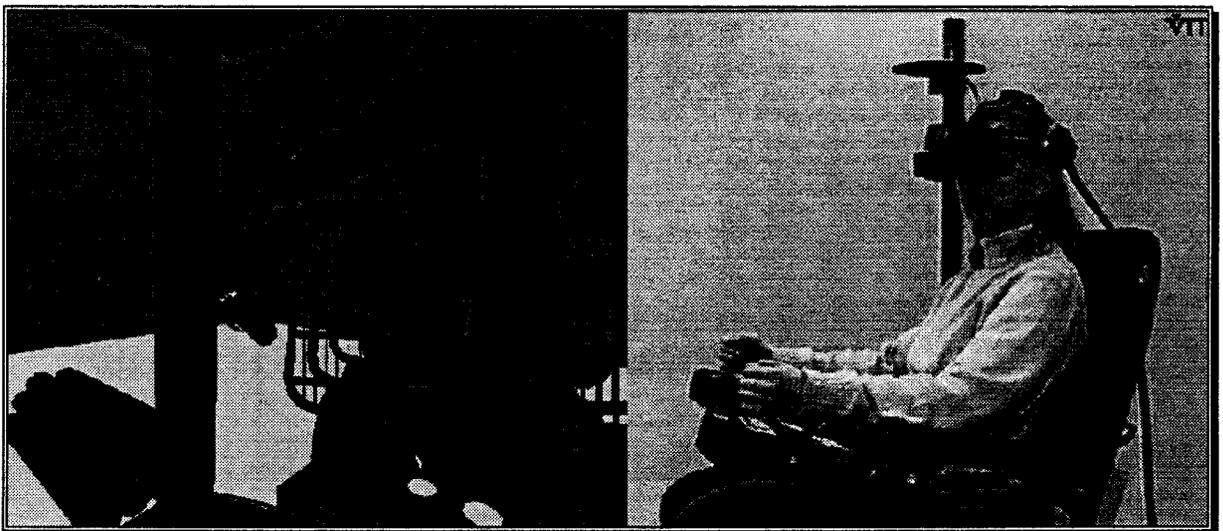


Figure 22 : un cockpit virtuel, vues intérieure et extérieure [Viitaniemi97]

Les MUD (Multi-User Dungeon), des environnements ludiques partagés, existent depuis longtemps. Ils étaient dans un premier temps en mode texte (TinyMUD, LPMUD, DikuMUD, AberMUD), puis graphique comme BSX (une extension de LPMUD, ayant donné naissance à Regenesis ou

Regenbogen¹¹, Figure 23). L'émergence de la Réalité Virtuelle a coïncidé avec la généralisation de la mise en réseau des ordinateurs et la vulgarisation d'Internet. Il était inévitable que la représentation graphique 3D attrayante de l'une se joigne aux possibilités de communication rapides, presque en temps réel. Les CVE sont la version « Réalité Virtuelle », plus luxueuse, des MUD.



Figure 23 : une capture d'écran d'un client BSX

Du matériel spécialisé comme les casques (*Head Mounted Display*, Figure 24), qui incluent deux écrans pour l'affichage stéréo et des dispositifs permettant de suivre le mouvement de la tête, ou les gants (*datagloves*) sont apparus, popularisés notamment par la firme VPL Research, Inc (Visual Programming Language) dirigée par le très médiatique Jaron Lanier. Il devint ainsi possible de se sentir naturellement présent dans un environnement entièrement synthétique en vraie 3D, partagé par d'autres personnes comme DIVE dont nous avons déjà parlé.

¹¹ <http://mud.inf.tu-dresden.de:4780/>

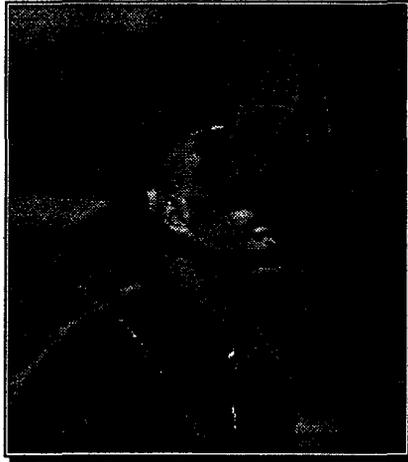


Figure 24 : « Head Mounted Display »

La Réalité Virtuelle, selon les mots de Jaron Lanier lui-même, est passée du statut d'interface utilisateur à celui d'outil de communication.

Paradoxalement, l'interaction avec un simulateur est plus facile à réaliser qu'avec un monde virtuel comme DIVE. En effet, le système de commande constitue l'interface entre l'utilisateur et l'appareil à simuler. Il est relativement facile de détecter la position d'un manche à balai, la rotation d'un volant, l'enfoncement d'une pédale ou d'un bouton pour faire réagir le simulateur proprement dit à ces informations. Dans les CVE, les personnes sont elles-mêmes représentées directement. Le nombre de degrés de liberté du corps humain est élevé et, de plus, il est difficile d'y placer des capteurs à la fois nombreux, précis et qui n'occasionnent pas de gêne. Les gants par exemple donnent la position des doigts et des phalanges ainsi que la position de la main dans l'espace mais la précision et l'encombrement sont perfectibles. Les problèmes posés par la navigation (les déplacements d'une personne dans le monde virtuel) sont énormes. S'il est possible de suivre la « vraie » position du corps pour la projeter dans l'environnement virtuel, cette technique implique que l'observateur ait la place de se mouvoir. La collision avec les objets qu'il voit doit influencer sur ses déplacements : il doit être bloqué contre un meuble, un arbre... On voit souvent en Réalité Virtuelle un objet passer à travers un autre, d'une part parce que la détection de collisions est un problème difficile, gourmand en temps de calcul, d'autre part parce qu'une fois ces collisions détectées il faut les prendre en compte et les restituer. Cela implique d'utiliser des dispositifs à retour d'effort, capables par exemple d'arrêter la main « contre » un objet virtuel. C'est possible mais jusqu'à maintenant les solutions apportées sont loin d'être parfaites (onéreuses, complexes, limitant les mouvements). Ajoutons que si l'œil se contente d'images à 25 Hz voire moins, le retour d'effort ne devient vraiment efficace qu'aux alentours de 1000 Hz [Cadoz92], i.e. quand le phénomène est rendu au moins 1000 fois par seconde.

Devant l'impossibilité de simuler intégralement la réalité, les concepteurs se sont livrés à de nombreuses simplifications et approximations. Certaines sont grossières (mais efficaces) : un bouton suffit pour avancer, se mettre à marcher ; on considère qu'une personne se saisit d'un objet dès que sa main en est proche. D'autres sont plus subtiles, comme les « auras » de DIVE. Les *widgets* sont

également utilisés en 3D. Ce sont des objets graphiques associés à des objets de la scène qui agissent comme des « périphériques virtuels ». Une action simple sur un widget déclenche des mécanismes compliqués qu'il aurait été difficile d'exécuter par de la manipulation directe pure.

1.2.1 Rappel sur la synthèse d'images temps réel

Dès lors que l'ordinateur n'a plus été limité à l'affichage de texte, on a cherché à lui faire représenter des scènes d'une manière réaliste. La Réalité Virtuelle hérite de l'histoire déjà longue de la synthèse d'images. Nous allons, succinctement, rappeler comment on en est arrivé aux techniques utilisées actuellement. On peut distinguer deux types d'approches :

- le rendu projectif [Foley90] :

Dans ce premier cas, chaque objet de la scène est décrit comme un ensemble de sommets, d'arêtes et de facettes triangulaires. Les objets sont placés dans la scène par rapport à l'observateur (par rotations, translations et mises à l'échelle), l'éclairage de Phong est calculé, la projection sur l'écran est appliquée (en perspective ou orthogonalement), puis vient l'étape du rendu qui peut être réalisé en « fil de fer », avec une couleur par facette ou lissé par interpolation (Gouraud, Phong). Chaque étape étant indépendante, cette méthode se prête bien à la parallélisation des différentes étapes (pipe-line de rendu, voir Figure 25). Les ombres portées doivent être calculées explicitement si besoin est. En qualité, si l'on s'en tient aux algorithmes purement synthétiques, les images obtenues sont très artificielles. Le placage de textures donne la possibilité de rendre les objets modélisés plus réalistes. La technique consiste à projeter une image 2D sur l'objet à représenter. Cette image peut être une photographie d'un objet réel. Ainsi les objets apparaissent comme s'ils étaient en bois, en marbre, etc.

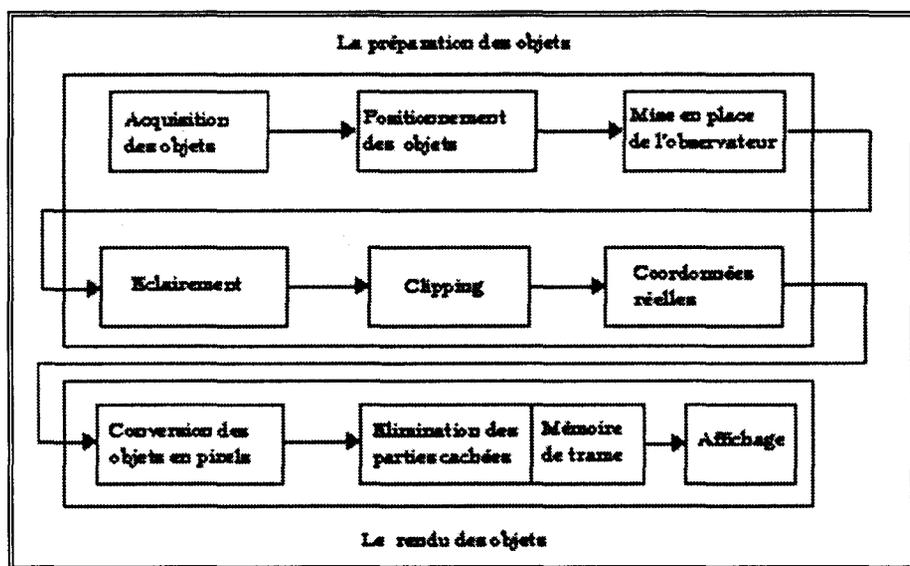


Figure 25 : le pipe-line de rendu classique

- les méthodes globales : les deux principales sont le lancer de rayon et la radiosit .
 - ◊ le lancer de rayon [Whitted80] traite l'image point par point. On « lance » un rayon lumineux de l'œil de l'observateur vers la sc ne   travers chaque pixel de l' cran.   chaque collision du rayon avec un objet, celui-ci sera susceptible de se s parer en un rayon r fl chi (effet miroir) et un rayon r fract  (transparence) et l'algorithme se poursuit r cursivement. Les objets peuvent  tre d crits par des primitives math matiques simples (plan, facette, sph re) ou plus complexes (surfaces implicites). L'algorithme calcule implicitement les ombres port es, de mani re exacte.
 - ◊ la radiosit  [Goral84] effectue le calcul des  changes lumineux dans la sc ne avant de passer par un rendu projectif classique pour l'affichage des objets. Pour avoir une bonne pr cision, chaque objet est d coup  en petites facettes. Chacune des N facettes de la sc ne enti re est susceptible de recevoir de l' nergie lumineuse des N-1 autres facettes et d'en r mettre une partie qu'il faudra prendre en compte. La majeure partie des calculs consiste   d terminer quelles facettes sont visibles depuis une facette particuli re et dans quelles proportions (*facteurs de formes*). Le temps de calcul de ces millions d'interactions augmente rapidement avec la complexit  de la sc ne et se chiffre couramment en heures.

Si les m thodes globales ont pour avantage la qualit  de l'image obtenue, les temps de calcul qu'elles n cessitent excluent aujourd'hui leur utilisation pour des applications temps r el,   l'animation et donc le calcul des images successives doit avoir lieu environ 20 fois par secondes.

Les machines sp cialis es dans la synth se d'images temps r el fonctionnent actuellement toutes en rendu projectif. Les objets sont repr sent s sous forme d'ensembles de facettes triangulaires. Le placage de textures apporte   moindre c ut (calcul et m moire) un aspect r aliste aux objets. Les algorithmes employ s (produits de matrices, interpolations...) sont simples (quelques lignes de programme suffisent pour tracer un segment de droite ou dessiner une facette triangulaire) et parfaitement ma tris s. Des acc l rateurs mat riels sp cialis s ont pu  tre r alis s (il en existe  galement pour les m thodes globales mais ils ne permettent pas encore le temps r el). Cela a permis de d tailler finement les objets, de les mod liser par un grand nombre de facettes. Aujourd'hui, le rendu projectif est la seule m thode exploitable pour obtenir des images anim es en temps r el.

1.2.2 Les p riph riques d'entr e et de sortie

Les interfaces WIMP¹² (cf. 1.3.1), les plus r pandues, allient plusieurs concepts visuels   un p riph rique d di  simple, la souris. Si l'utilisation qui en est faite (utilisation des boutons par exemple) peut varier l g rement d'un gestionnaire   l'autre, elle permet d'effectuer simplement et de mani re standard de nombreuses interactions dans ces environnements 2D. La pr sentation graphique en fen tres et la souris sont fortement li es.

¹² Windows Icons Mouse Pointer.

La 3D n'a pas encore trouvé son périphérique d'entrée (i.e. de l'utilisateur vers l'ordinateur) universel. Au contraire, les dispositifs variés foisonnent, chacun étant spécialisé, que ce soit à un type d'utilisation (simulateur, « cybersex »), ou à une partie du corps (gants, suivi de l'œil—Figure 26).

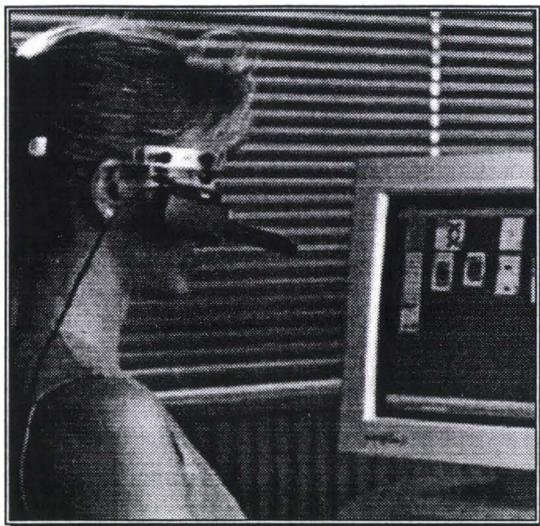


Figure 26 : un dispositif de suivi du regard, remplaçant ici la souris

Les expériences de Réalité Virtuelle font appel à une multitude de capteurs pour enregistrer les variations du plus grand nombre possible de degrés de liberté du corps humain. Certains (gant) mesurent les angles des articulations, d'autres fournissent la position, voire l'orientation, dans l'espace d'un point de référence. On distingue trois types de périphériques « simples » (Figure 27). Les capteurs isométriques (\approx même déplacement) comme les Polhemus ou l'OWL de Pegasus captent une position dans l'espace et éventuellement une orientation. Les périphériques isotoniques (\approx même force) comme la Spaceball mesurent des forces ou des couples. Ils sont souvent utilisés pour le contrôle de robots. Peu exploités, les périphériques élastiques utilisent force et déplacement et disposent d'une position centrale de repos.

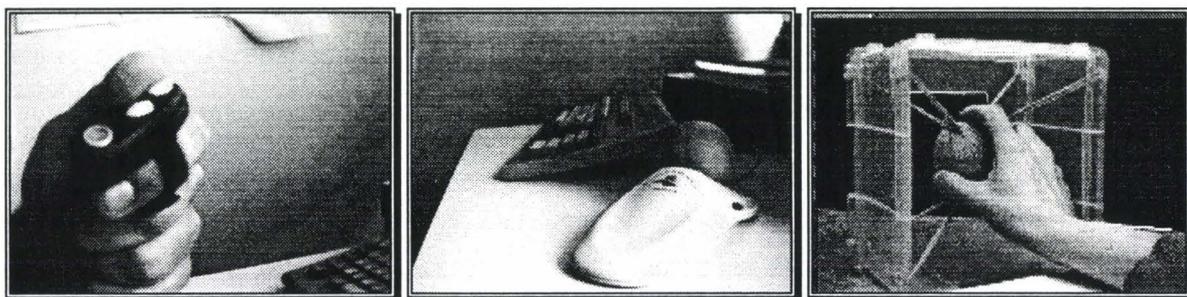


Figure 27 : périphériques isométrique, isotonique et élastique

Les dispositifs de sortie ne sont pas en reste. Les casques immersifs fournissent image stéréoscopique et son stéréophonique. La qualité des écrans (cristaux liquides le plus souvent) est encore malheureusement médiocre. Le son spatial, étudié depuis longtemps [Barme96], est bien

maîtrisé. Il nécessite néanmoins une puissance de calcul importante (utilisation de DSP¹³ pour le traitement du signal, circuits intégrés spécialisés) et fait intervenir des facteurs subjectifs (écho dépendant de la forme des épaules par exemple). Le retour d'effort, permettant d'appréhender au mieux les objets que l'on touche, nécessite des installations (mécanique, moteurs) à la fois précises et puissantes, réservées à des applications particulières.

Claude Cadoz [Cadoz94] [Cadoz96] fait le point sur les différentes natures des gestes et sur les dispositifs d'entrée les utilisant. Le geste est particulier du fait qu'il sert à la fois à émettre et à recevoir de l'information. On distingue trois fonctions complémentaires : ergotique (action matérielle sur l'environnement), épistémique (connaissance de l'environnement) et sémiotique (émission d'information vers l'environnement).

La fonction épistémique ne s'arrête pas à la perception tactile cutanée (état de surface, température) concentrée sur la face interne des doigts. La main est l'organe du sens tactilo-proprio-kinesthésique. La perception tactilo-kinesthésique, ou haptique, renseigne sur le monde environnant par des mouvements d'exploration et l'évolution des informations reçues. Tout comme la vision, ce sens ne se réduit pas à prendre des « photographies tactiles » de l'environnement mais construit une image mentale à partir des informations perçues par exploration, en l'occurrence la forme, l'orientation, la distance et la taille des objets palpés. La proprioception provient des récepteurs des articulations et de l'oreille interne. Elle nous renseigne sur le poids, les trajectoires.

Dans la fonction ergotique, la main est en contact avec la matière, qu'elle va modifier. Les objets reçoivent de l'énergie, et non de l'information, et peuvent en renvoyer une partie en retour.

La main peut aussi être un organe d'expression : c'est la fonction sémiotique du geste. Elle est naturelle et spontanée. Non seulement les gestes accompagnent la parole mais ils partagent avec elle certaines étapes de traitement. C'est un mode d'expression en soi à part entière et non pas seulement un ajout à la parole.

Ces fonctions coexistent dans ce que Cadoz nomme le *geste instrumental*, celui du musicien. En frappant sur une touche de piano, on communique de l'énergie à l'instrument pour le faire vibrer mais le but est de s'exprimer, de faire passer une émotion.

Dans la relation avec la machine, un *capteur* doit permettre de convertir les gestes en information pour la machine, tandis qu'un *effecteur* transforme une information en énergie physique. Le *transducteur*, qui est soit un capteur, soit un effecteur, soit les deux, assure les fonctions nécessaires à l'interaction. Les transducteurs sont de deux types. Sans contact, ils ne font qu'envoyer de l'*information* à la machine en captant à distance les phénomènes gestuels (caméra, Polhemus électromagnétique, Pegasus OWL à ultrasons...). Les transducteurs avec contact disposant d'un

¹³ Digital Signal Processor.

effecteur peuvent fournir deux sortes de retour. Celui-ci peut être juste *tactile*, s'adressant à la perception tactile cutanée (certains gants disposent de ballonnets, de minuscules pistons ou de vibreurs). Il transmet alors juste de l'information mais n'empêche pas la main de traverser un objet (virtuel). Si au contraire il s'adresse à la perception tactilo-proprio-kinesthésique, c'est un retour d'effort complet. Dans ce cas, les contraintes mécaniques (puissance élevée, bande passante de l'ordre du kiloHertz, nombreux degrés de liberté—23 pour la main) interdisent la construction de dispositifs simples, petits et pratiques.

1.2.3 La vision en R.V.

Les dispositifs d'affichage utilisés en R.V. immersive sont encore pauvres et limités (taille, résolution). Des travaux, comme ceux de Ware et Usoh indiquent deux directions pour les améliorations futures. La première est de mettre à profit les facultés humaines d'apprentissage et de reconnaissance de son environnement par l'exploration. La seconde est d'exploiter la vision périphérique, d'agrandir l'angle de vision.

La vue couplée aux mouvements de la tête

Le système de réalité virtuelle « Fish Tank » [Ware93] a pour particularité d'être couplé aux mouvements de la tête de l'utilisateur. A la différence d'autres systèmes immersifs, le mouvement de l'utilisateur dans l'espace réel est impossible. L'affichage se fait en stéréoscopie sur un moniteur 120 Hz (soit 60 Hz/œil), ce qui donne de meilleurs résultats, en qualité, que les dispositifs généralement employés en réalité virtuelle. Les expériences les plus importantes menées sur le Fish Tank ont eu pour but d'évaluer les apports de la visualisation stéréoscopique par rapport à la perspective simple et du couplage de la direction de vision avec les mouvements de la tête. Leurs conclusions montrent que, si les deux se complètent bien, le couplage tête/mouvement apporte plus d'informations (sensation de profondeur, placement dans l'espace), par le mouvement relatif des objets observés, que la simple stéréoscopie. Ces observations corroborent les expériences de Wanger et al. [Wanger92] (voir 1.3.5, « La perception et le repérage en 3D »).

Vision périphérique en R.V. immersive

Dans les environnements virtuels immersifs (IVE¹⁴), l'utilisateur doit être littéralement plongé dans le monde synthétique, aussi faut-il éviter qu'il ait encore conscience du monde réel. Les écrans utilisés jusqu'alors dans les casques HMD sont petits, plats, et ne couvrent donc pas le champ de vision humain (représenté sur la Figure 28, son étendue dépend de la direction) qui s'étend sur les côtés de

¹⁴ Immersive Virtual Environment

la tête¹⁵. Seule la vision centrale est mise à contribution, tandis que la vision périphérique est inutilisée, l'observateur étant privé des informations qu'elle lui fournit normalement. Rappelons que la vision centrale est plus précise et distingue mieux les couleurs. La plus grande acuité est atteinte dans la « tache jaune » de la rétine, elle est maximale dans la fovéa. Le diamètre de celle-ci correspond à un angle de vision de seulement un degré. La vision périphérique est essentiellement monochrome et son acuité est réduite mais elle réagit au mouvement. Les mouvements de l'œil ne sont pas continus mais alternent saccades et pauses (que l'on classe selon leur durée). Les sauts les plus grands (en angle) surviennent quand un mouvement excite la rétine. On est presque obligé, par réflexe, de tourner les yeux, voire la tête et le corps, très vite et inconsciemment vers ce que l'on a aperçu du coin de l'œil. Dans un environnement réel, elle aide à se repérer et à se rendre compte de l'espace environnant. Le problème de repérage est commun dans les IVE, par manque de points de repères mais aussi à cause des difficultés de navigation et d'exploration.

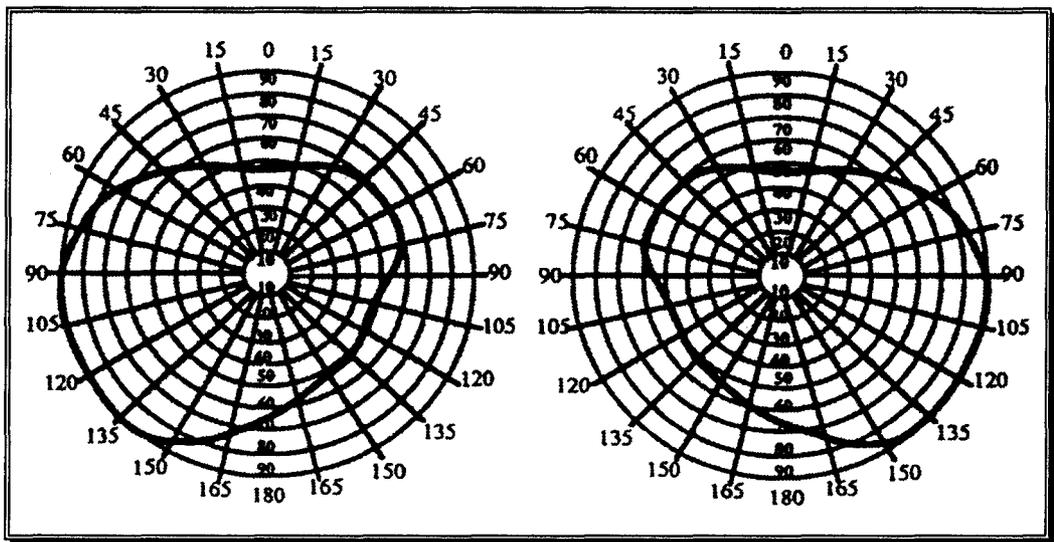


Figure 28 : champs visuels de l'œil gauche et de l'œil droit [Emslie-Smith89, p. 451]

Après une étude préliminaire prouvant que, sous un casque, les utilisateurs réagissent normalement à des mouvements sur les bords des écrans, Slater et Usoh présentent un nouvel algorithme simple de rendu [Slater93] tel que la majeure partie de l'écran est dessinée de manière classique, alors que les bords contiennent la périphérie de la scène. Bien entendu, les déformations écrasent les objets situés sur les côtés mais tous les objets présents dans la limite de 90° par rapport à l'axe de vision sont présents à l'écran : la personne peut donc voir distinctement ce qui est devant elle et son attention peut être attirée par d'éventuels mouvements alentours. L'idée des auteurs a été de modifier le pipe-line de rendu classique pour visualiser deux pyramides de vision, l'angle de la

¹⁵ les casques très haut de gamme comme le Datavisor 80 de n-vision atteignent 120°, la moyenne se situe seulement vers 70°. A titre comparatif, l'œil humain voit sur 190° et le champ de vision binoculaire est de 220°.

deuxième étant très grand. L'image issue de la pyramide « normale » est affichée, centrée, sur la plus grande partie de l'écran, les bords accueillant les objets périphériques (Figure 29). Il en résulte des discontinuités à la jonction des images issues des deux pyramides, un écrasement de la zone périphérique (large zone projetée sur les bords de l'écran) et donc une représentation géométriquement fautive. Néanmoins, la méthode présentée permet de simuler dynamiquement un champ de vision proche de 180°. Elle permet également de séparer l'espace en deux zones, l'une d'entre elles pouvant être affichée avec une moindre qualité et donc plus rapidement si besoin est. Le pipe-line de rendu avec déformation est très proche du pipe-line classique, les accélérateurs graphiques existants peuvent donc être utilisés presque sans modification. Il reste à prouver que, comme on l'espère, les utilisateurs réagiront bien comme s'ils avaient un champ visuel de 180°.

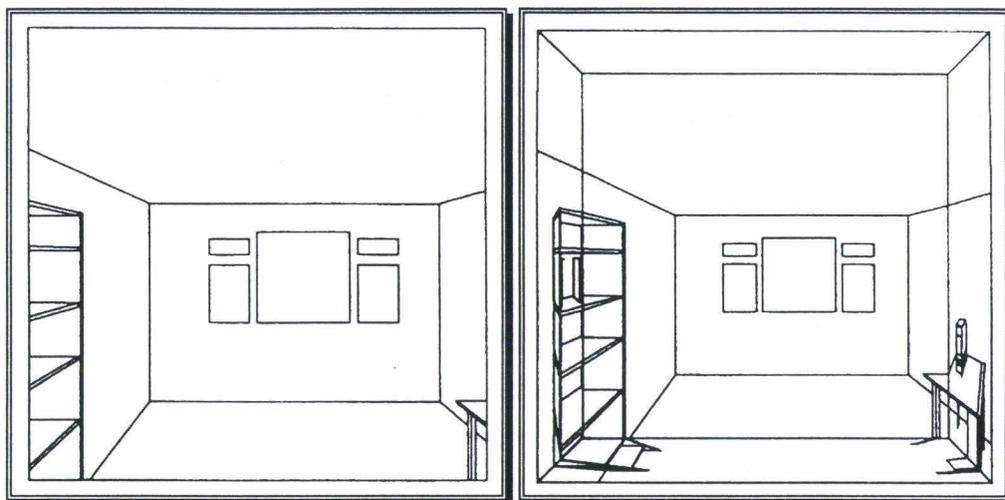


Figure 29 : rendu normal (à gauche) et avec zone périphérique (à droite)

Voici deux schémas (Figure 30), l'un représentant le rendu classique (équivalent à la « boîte noire » des appareils photographiques antiques, dont l'objectif est un trou très petit) et l'autre la version améliorée avec plusieurs plans de projection :

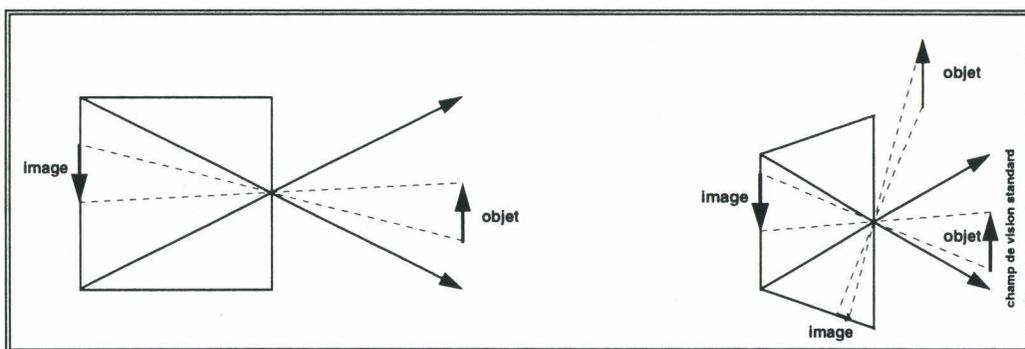


Figure 30 : projection classique et projection modifiée « grand angle »

1.2.4 Les avatars et les clones

Les plates-formes de Réalité Virtuelle étant devenues multi-utilisateurs, il a fallu chercher comment représenter les personnes partageant le même monde. Le terme utilisé pour désigner une représentation quelconque d'une personne est « *avatar* ». Cet avatar peut être un smiley, un personnage de dessin animé, un animal, un robot... Le clone est un avatar particulier qui copie l'apparence physique d'une personne. Le degré de clonage peut varier. Une photographie, par exemple, recopie exactement les traits de la personne (en cela c'est un « clone 2D ») mais reste figée et est incapable de traduire des émotions. Pour simplifier, nous appellerons *maître* ou *propriétaire* la personne qui se trouve représentée sous la forme d'un avatar.

Nous présenterons tout d'abord les avatars et les clones à travers deux projets de Réalité Virtuelle, DIVE et MASSIVE. Ensuite nous détaillerons la difficulté d'un clonage complet de qualité à travers le projet VLNet. Nous décrivons enfin les travaux sur les clones réalisés par l'équipe Télévirtualité de l'Institut National de l'Audiovisuel.

Dans DIVE

DIVE permet de construire des environnements distribués 3D, multi-applications, multi-utilisateurs, multi-plates-formes. Le but du projet CyCo, dans lequel s'inscrit DIVE, est de supporter des communautés virtuelles à large échelle. A l'origine, l'implémentation de DIVE avait pour but de valider, par l'expérimentation et l'évaluation, un modèle spatial d'interaction dans de grands environnements virtuels [Fahlén93] [Benford93c] (voir 1.1.2 pour une description complète). Ce modèle est applicable à tout système où une métrique spatiale peut être définie ou, en d'autres termes, où il est possible de mesurer une distance et une orientation.

DIVE est un « véritable environnement de Réalité Virtuelle » : le monde dans lequel l'utilisateur évolue est modélisé géométriquement pour ressembler à la réalité. On retrouve des éléments habituels : reliefs du terrain, bâtiments, meubles, panneaux d'affichage... L'entrée dans un nouveau lieu se fait par téléportation (la description de l'endroit où l'on désire se rendre est chargée et celui que l'on quitte disparaît pour être remplacé purement et simplement).

Les premiers avatars de DIVE étaient du type « blockies », c'est-à-dire qu'ils étaient composés de quelques objets à la géométrie simpliste (Figure 31). Quelques traits statiques symbolisaient le visage. Ces corps pouvaient être définis par l'utilisateur ou simplement choisis dans une petite bibliothèque de blockies prédéfinis. Malgré leur dépouillement, ils étaient tout de même porteurs d'informations importantes, la position d'une personne et la direction vers laquelle elle est tournée. Les utilisateurs sont sensibles à ces données et cherchent naturellement à ne pas se tourner le dos, à essayer de voir la personne qui leur parle (bien que cela n'influe pas sur le canal sonore). Le modèle spatial, par les notions d'*aura*, de *focus* et de *nimbus* tire parti de ces conventions sociales. Ceci dit, la manière dont il a été implémenté dans DIVE peut être vu à l'inverse, comme une forte incitation à les respecter.

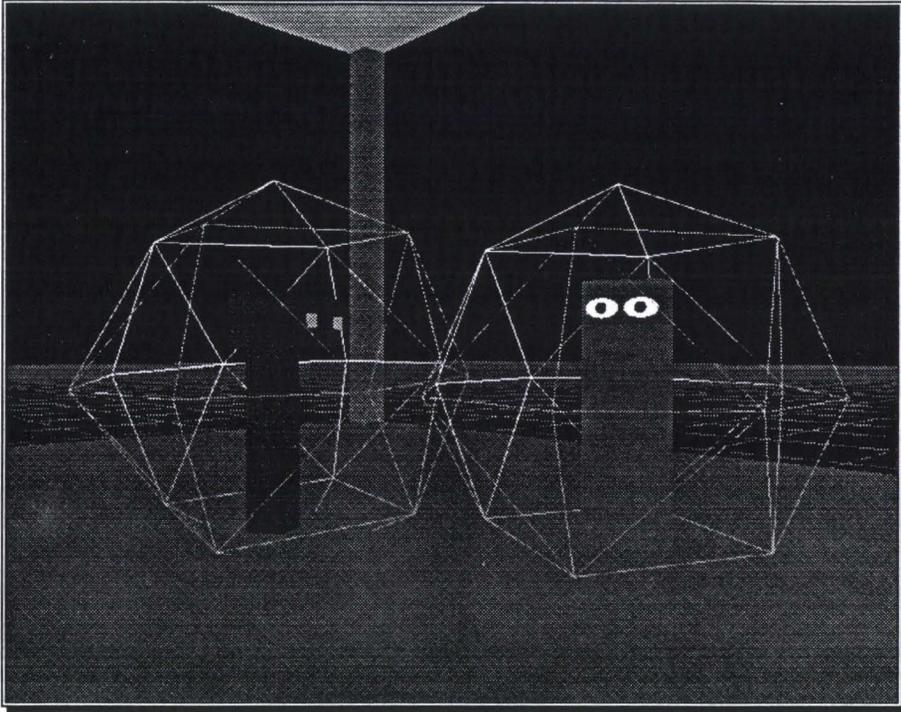


Figure 31 : « blockies » de DIVE avec leurs auras

Quantité de nouveaux types d'avatars ont depuis été développés [Benford95]. Il devient plus facile de trouver une représentation pertinente, appropriée, de soi pour les autres parmi toutes celles qui sont proposées. Les nouveaux corps fournissent des informations immédiates et continues sur la présence, l'activité, l'attention, la disponibilité, l'humeur, le statut, la position, l'identité, les possibilités, etc. Ils permettent une communication explicite, par les gestes notamment, ou accompagnent d'autres modes d'interaction. Sur la Figure 32, on peut voir de gauche à droite : un utilisateur « immergé » avec un corps humanoïde, le visage texturé et dont la tête et un bras sont repérés par des capteurs ; un « blockie » portant un chapeau à hélice ; un utilisateur qui participe par l'intermédiaire d'un système de vidéoconférence ; un autre utilisateur immergé. Ces différents corps sont avant tout le reflet du matériel dont chacun dispose, déterminant directement les possibilités et la capacité d'expression de l'avatar. Les textures sur le visage sont de simples photographies fixes. Un capteur 3D suffit à donner la position du bras. La tête est repérée avec six degrés de liberté (position et orientation), ces données étant fournies par le casque HMD. Elle donne des informations sur le point d'intérêt de la personne, son activité et son degré de présence. Une couleur peut être affectée à chaque participant. Son avatar revêtira cette couleur, qui peut aussi être transmise aux objets qu'il manipule.

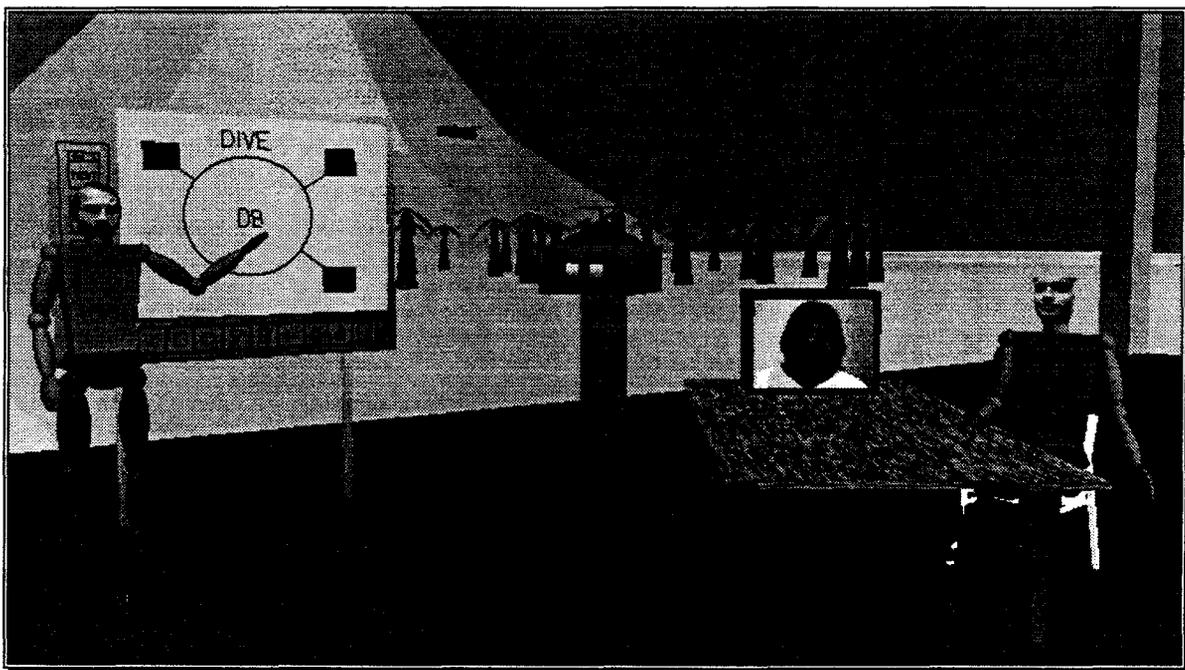


Figure 32 : différents avatars et clones dans DIVE

Dans MASSIVE

MASSIVE (Model, Architecture and System for Spatial Interaction in Virtual Environments) est un système de conférence en Réalité Virtuelle. Ce projet a deux objectifs : supporter des mondes à grande échelle (en particulier un grand nombre d'utilisateurs) et permettre l'interaction entre personnes disposant d'équipements et d'interfaces très différents.

Les interfaces utilisables dans MASSIVE sont : graphique, audio, texte. L'interface graphique permet d'afficher un environnement 3D dans lequel l'utilisateur peut se déplacer. L'audio fournit d'une part des sons préenregistrés et d'autre part des conversations en temps réel. L'écran texte rappelle les MUD dans lesquels l'espace est représenté en 2D vu de dessus. Chaque utilisateur apparaît sous la forme d'un caractère, généralement l'initiale de son nom, épaulé par un trait indiquant vers où il est tourné (huit directions possibles). Des messages et émotions peuvent être envoyés aux autres participants. Ces trois interfaces peuvent être combinées selon ce que le matériel est capable de faire. On peut parler même si on ne voit que du texte. Si l'on dispose d'un affichage 3D, une fenêtre texte peut servir de carte pour se repérer. Par contre, un utilisateur peut exporter une vue de lui qui ne correspond pas au type d'interface qu'il utilise. Grâce à cela, il est possible de faire interagir des personnes disposant d'équipements hétérogènes.

Les avatars 3D ont deux rôles à remplir. Le premier est de faire connaître l'interface qu'utilise son maître et donc quel moyen il faut utiliser pour communiquer avec elle. Le second est d'informer sur la position et l'orientation des autres utilisateurs. La forme générale de ces avatars est un blockie en forme de « T », bien que des formes humanoïdes soient également disponibles. Deux yeux indiquent un utilisateur immergé. L'utilisateur d'un système 3D non stéréoscopique a droit à un cyclope. Celui

utilisant une interface texte se verra marqué d'un « T » sur son front (la face avant). Des oreilles traduisent la présence d'une interface sonore (Figure 33). Parmi les améliorations apportées, le propriétaire d'un avatar peut maintenant explicitement lui faire bouger les oreilles indépendamment. Nous verrons que cette possibilité, pas vraiment réaliste, peut quand même servir à s'exprimer et suggère des gestes naturels. Pour faciliter l'identification des personnes en train de parler, leurs avatars sont momentanément pourvus d'une bouche symbolique.



Figure 33 : quelques avatars de MASSIVE

Difficulté du clonage total

Le clonage réaliste, qui constitue une part importante des problèmes des environnements de Réalité Virtuelle, peut être séparé en deux tâches distinctes et complémentaires. La première est de fournir une représentation crédible de l'utilisateur. La seconde est que les mouvements doivent être réalistes. Nous allons étudier ces deux aspects à travers le système VLNet [Capin97] (Figure 34).

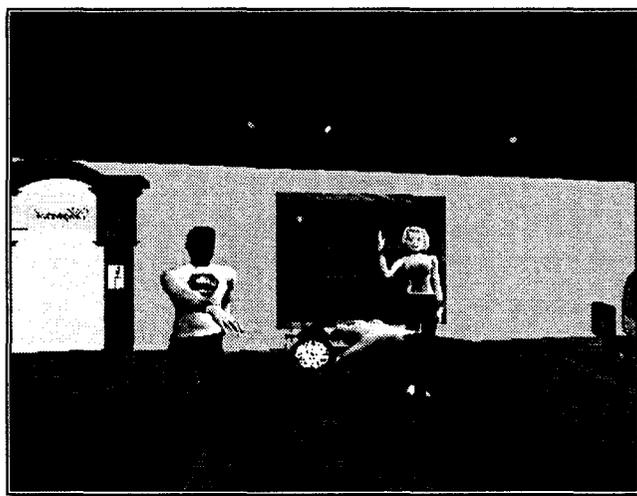


Figure 34 : des clones dans VLNet

Commençons par rappeler que le corps humain est pour le moins un système complexe [Cadoz94] : le système de la motricité humaine gère quelques 792 muscles, 110 articulations et 100

degrés de liberté. La main seule possède 23 degrés de liberté. Ces chiffres suffisent pour donner une idée de l'ampleur du problème de la reproduction du corps humain !

L'homme est modélisé dans VLNet par un squelette possédant 75 degrés de liberté, plus 30 par main (Figure 35). Des « blobs », ou « metaballs » [Blinn82] sont fixés dessus pour représenter peau et muscles (Figure 36). Le corps est divisé en 17 parties : tête, cou... Le contour de la peau est précalculé et déformé dynamiquement en suivant le squelette. Les paramètres dynamiques de description concernent le positionnement global (positions des épaules, de la tête, des coudes...), les articulations entre les différentes parties du corps, les mains et les doigts, et enfin le visage (déformé par FFD¹⁶ [Sederberg 86]). Au plus bas niveau, 65 « actions perceptibles minimales » peuvent être contrôlées (elles correspondent aux muscles). Au plus haut niveau, la parole ou des émotions peuvent être décrites par scripts. Il est possible de plaquer une texture sur le visage.

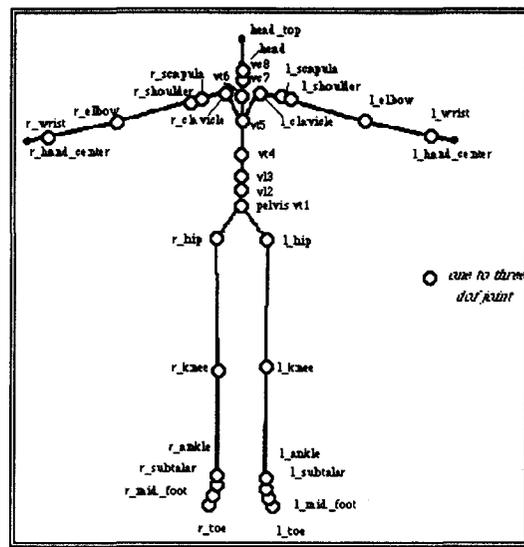


Figure 35 : le squelette et ses articulations

¹⁶ Free Form Deformation

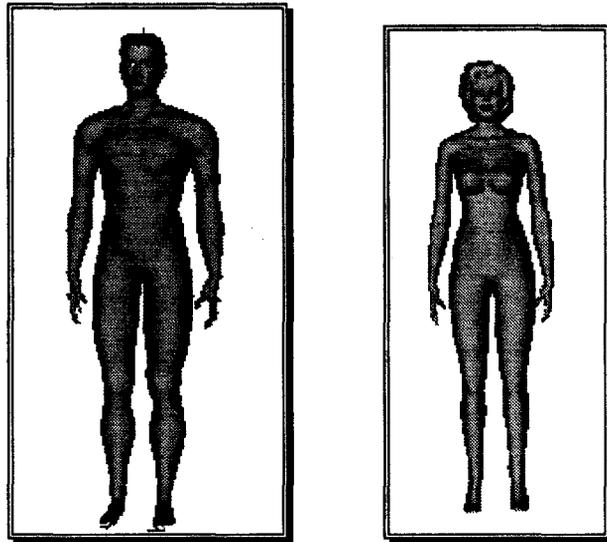


Figure 36 : les squelettes recouverts de metaballs

Le contrôle de l'humain virtuel se fait par des mécanismes de haut niveau. Ils sont rendus nécessaire par la complexité du corps décrit précédemment, qui rend impraticable un contrôle fin, articulation par articulation. On distingue trois méthodes d'animation de l'humain virtuel : le contrôle direct, le guidage par l'utilisateur et les être virtuels autonomes.

Le contrôle direct consiste à reproduire les articulations et le visage de la personne réelle. Idéalement, il devrait y avoir autant de capteurs que d'articulations réelles. Des méthodes existent qui se contentent de 14 capteurs, voire même 10 ou même seulement 6 dans VLNet. Naturellement, il faut procéder à diverses approximations : extrapolations, cinématique inverse, choix de postures réalistes, tables acquises expérimentalement, etc.

Le guidage s'apparente à la manipulation de marionnettes. Certaines actions de l'utilisateur ; comme le mouvement des doigts si on utilise un gant, fournissent des données qui sont transposées à d'autres effets (comme la position des yeux). Ces données peuvent alimenter des processus de simulation, de la marche par exemple. Tout type de périphérique peut être utilisé.

Les participants virtuels autonomes sont artificiels. A partir d'une tâche à accomplir et de divers comportements possibles, ils vont agir comme s'il s'agissait de personnes réelles. Ils interviennent auprès des utilisateurs pour améliorer le système, par exemple en leur fournissant de l'aide, en les orientant, etc.

Le visage peut être rendu par quatre techniques : texture vidéo, modélisation des expressions faciales, synthèse du mouvement des lèvres à partir de la parole et expressions ou animations prédéfinies. L'utilisateur se plaçant face à une caméra, on peut extraire de l'image vidéo une texture dynamique de son visage. Une étape d'analyse est nécessaire pour déterminer où son visage est situé dans l'image et éliminer l'arrière-plan. La texture est comprimée, transmise, décomprimée et finalement appliquée sur un modèle 3D simplifié de tête (Figure 37).

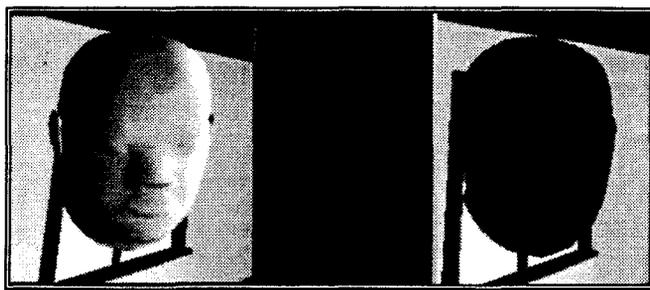


Figure 37 : placage de texture sur un modèle 3D

La deuxième méthode est d'analyser le visage dès acquisition dans le but d'extraire des paramètres décrivant l'expression du visage. Ceux-ci fournissent la position ou l'orientation de la tête, des sourcils, des yeux, l'ouverture de la bouche et des yeux, etc. Une faible quantité de données est transmise pour reconstruire à distance un visage dont l'expression correspond à celle de la personne réelle. Cette méthode existe dans VLNet. Nous verrons un peu plus loin que c'est également celle employée à l'INA.

La synthèse du mouvement des lèvres peut se faire à partir de l'analyse de la voix. C'est une solution à envisager dans les cas où on ne dispose pas de caméra (quand on utilise un HMD immersif par exemple). Cette synthèse peut être très sommaire en ouvrant et fermant la bouche dès que la personne parle (comme dans MASSIVE) ou photoréaliste [Bregler97]. La clé de la réussite est la synchronisation entre parole et mouvement des lèvres.

Une expression peut être choisie explicitement par l'utilisateur parmi un ensemble défini à l'avance. C'est en quelque sorte une version plus luxueuse des « smileys » ou « émoticônes ».

Les travaux de l'INA

Le projet « Télévirtualité » de l'INA a pour but la création de logiciels d'analyse et de synthèse d'images capables de transformer, en *temps réel*, le visage de personnes réelles en clones virtuels, en ne mobilisant qu'une faible bande passante, permettant une rencontre virtuelle de plusieurs clones. Séduits par l'élégance de leur procédé, la qualité du clonage et les possibilités offertes dans des environnements virtuels de travail coopératif, nous avons choisi de détailler leur démarche. Leurs recherches peuvent être découpées en trois parties indépendantes : l'analyse, la synthèse [Viaud95] et le réalisme. Nous présentons ici les résultats obtenus sur le clonage du visage, pour conclure par les perspectives qu'offre leur système.

i) Analyse

La simplicité du matériel et le confort de l'utilisateur ont été d'office privilégiés. Alors que d'autres projets déjà existants imposaient des contraintes encore trop fortes pour beaucoup d'applications (dispositifs de contact, capteurs, immobilisation de la tête face à la caméra), le système de l'INA se compose d'une chaîne analyse/synthèse (Figure 38) faisant intervenir de faibles contraintes. Aucun

capteur ni marqueur spécifique n'est utilisé : le visage de synthèse est piloté par un visage réel filmé par une simple caméra.

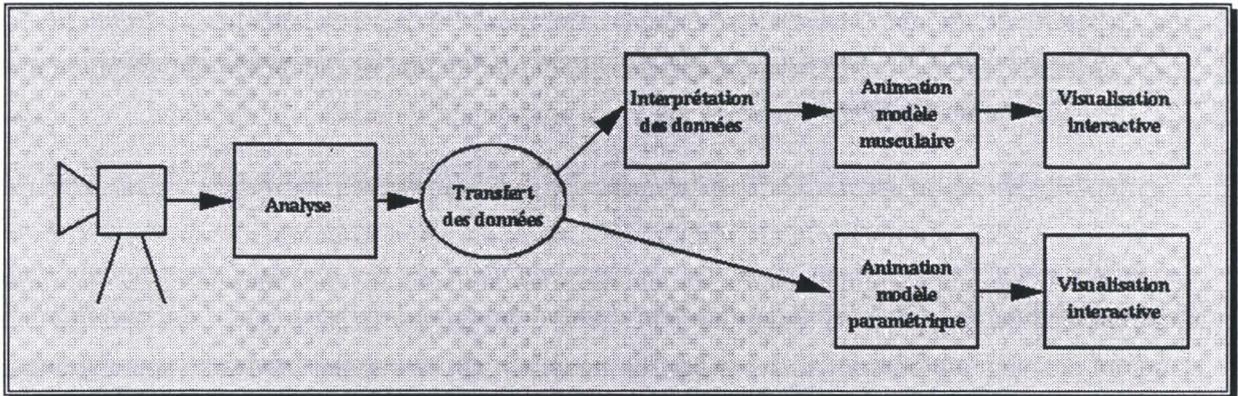


Figure 38 : chaîne de traitement temps réel

La détection des mouvements du visage est effectuée par un système passif. Placé dans une situation de travail, face à son écran, le participant est assis et peut bouger la tête. Une caméra est positionnée face à lui, un éclairage type « lumière du jour » est maintenu constant et le visage doit parfois être rehaussé d'un léger maquillage de studio (nécessité d'un rouge à lèvres). L'image vidéo est numérisée et traitée sur une SGI-Indigo.

C'est alors que commence véritablement la phase d'analyse. Des zones sont successivement reconnues, permettant au final de localiser des points caractéristiques du visage ainsi que d'autres informations comme la position de la tête. L'analyse permet d'extraire des paramètres caractéristiques du visage (mouvements de la tête, déformations de la bouche, direction du regard, position des paupières...). Voici en détail la succession de traitements appliqués à chaque image :

- localisation du visage sur l'image. C'est une recherche qui a pour but de déterminer un masque homogène correspondant à l'emplacement du visage dans l'image. Une fois le masque délimitant le contour du visage établi, une fenêtre de traitement est positionnée autour de chaque élément caractéristique du visage ;
- recherche des points caractéristiques autour de la bouche. Le contour extérieur des lèvres est calculé à l'aide d'un automate de suivi. Puis sur ce contour, 6 points caractéristiques sont positionnés (2 à la commissure des lèvres, 1 au milieu de la lèvre supérieure, 1 au milieu de la lèvre inférieure, et 2 points intermédiaires sur la lèvre supérieure). L'aire intérieure de la bouche est aussi calculée ;
- détection des yeux et des sourcils. La détection des yeux consiste tout d'abord à rechercher 2 formes, plus ou moins circulaires, correspondant aux iris. Puis le reste du contour des yeux est complété par propagation. De même est effectuée la recherche des sourcils au dessus des yeux. Après sont positionnés 2 points caractéristiques au centre des yeux ainsi que 2 points au

centre des sourcils. La hauteur de chaque œil est calculée ainsi que la position relative de chaque iris par rapport au reste de l'œil ;

- détection du nez. Un point est positionné au niveau du nez. Il correspond au centre de gravité des deux narines, ou bien il est calculé à partir de l'ombre du nez quand ces dernières ne sont pas visibles (tête baissée).

Après cette étape un certain nombre de points caractéristiques ont été reconnus ainsi que le contour du visage. Pendant l'apprentissage, à partir de ces données, plusieurs distances, correspondant à des mesures effectuées sur un visage au repos doivent être calculées. Ces mesures seront utilisées pour le calcul du repère de la tête ainsi que pour le calcul de toutes les déformations qui seront données en fonction de la position au repos. On détermine ainsi pendant l'apprentissage, par exemple, la largeur de la bouche, la hauteur des deux lèvres, la distance entre les deux yeux, la distance yeux-nez....

- calcul du repère de la tête (transformations globales). Un modèle topologique simple de la tête est utilisé pour établir des relations géométriques entre les différents points caractéristiques. Ces relations, ainsi que les distances au repos calculées pendant la période d'apprentissage, permettent de retrouver les rotations 3D à partir des données 2D détectées sur l'image. On calcule ainsi les trois angles de rotation ainsi que le changement d'échelle. Remarquons que la détection du changement de zoom (i.e. l'approche et le recul du visage par rapport à la caméra) permet de donner beaucoup de « vie » à l'expression du clone ;
- calcul des déformations du visage dans le repère de la tête (transformations locales). Connaissant les différents angles de rotation de la tête et les projections des points caractéristiques sur l'image, il est alors possible de calculer la position de ces points et leurs distances relatives dans le repère de la tête. Un filtre de Kalman est utilisé pour lisser les différents angles de rotations et à celui-ci ont été ajoutés de nouveaux filtres.

La robustesse des algorithmes d'analyse est appréciable. Ils fonctionnent correctement même s'il y a des mouvements dans le fond, derrière la personne à analyser. En cas d'anomalie, il ne diverge pas. Si le participant sort momentanément du champ de la caméra, le programme s'arrête plutôt que de fournir des informations erronées. L'analyse redémarre d'elle-même dès le retour de la personne.

Les treize paramètres extraits (soit treize octets seulement pour coder une expression) [Viaud96] sont communiqués à distance pour reconstituer un clone adoptant l'expression analysée.

ii) Synthèse

Le transfert des données s'effectue par réseau ou ligne téléphonique vers un autre site. Elles servent alors d'entrées pour l'animation de modèles de synthèse 3D. Les paramètres sont, soit transmis directement à un modèle paramétrique stylisé (BD ou caricatural, Figure 39), soit interprétés en terme d'actions musculaires (modèle facial « réaliste »).

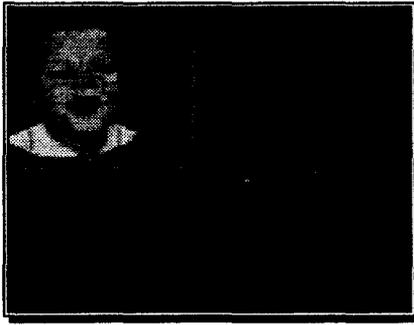


Figure 39 : un clone « bande dessinée »

La présence de détails, tels les rides d'expression, les marques de l'âge ou les signaux conversationnels auxiliaires présents lors du discours est très importante pour la perception du visage humain. L'INA propose un modèle d'animation faciale prenant en compte de tels détails.

Dans un premier temps, un masque de référence défini par une surface spline et contenant la carte de rides générique est déformé, et adapté au nouveau visage. L'animation faciale est ensuite générée par un système dynamique : un maillage de ressorts est associé à cette surface spline et des forces, dont l'action sur le réseau de ressorts correspond aux déformations dues aux muscles peauciers, sont définies. Les rides sont alors modélisées géométriquement et reliées automatiquement aux déformations musculaires : le mouvement et la forme du bourrelet dépend du système dynamique et d'une fonction associée à l'âge du visage donné.

Une simulation des trois phénomènes qui caractérisent le vieillissement d'un visage est proposée : l'apparition et la persistance des rides d'expression, le changement du modelé dû au relâchement du tissu peaucier combiné à l'action de la pesanteur, et le changement de texture de la peau.

La notion de généricité du visage est prise en compte afin de permettre la création de nouveaux visages par déformations et adaptations des différentes structures décrites plus haut. Le passage d'un visage à un autre peut aussi servir à créer de nouveaux visages.

iii) Réalisme

Une préparation importante est réalisée pour la construction des modèles de clones. Ici encore, le matériel requis est simpliste (une caméra et un éclairage de bonne qualité). A partir de photographies de la personne à cloner et d'un maillage 3D général qui sera adapté à la morphologie du modèle, le clone final est obtenu. Il ne reste plus qu'à lui appliquer les déformations correspondant aux données envoyées par le module d'analyse pour retrouver le visage de la personne réelle.

- prise d'images avec une seule caméra :

Le clone 3D réaliste est construit à partir d'une image de face et de profil (Figure 40). L'acteur se positionne face à la caméra, son visage remplissant l'image. L'éclairage doit être diffus, aucune ombre franche ne doit apparaître sur la surface du visage. La difficulté de la reconstruction réside dans la

prise de vue avec caméra unique. En effet, les images ne sont pas homogènes entre elles, tant au niveau éclairage qu'au niveau géométrie.

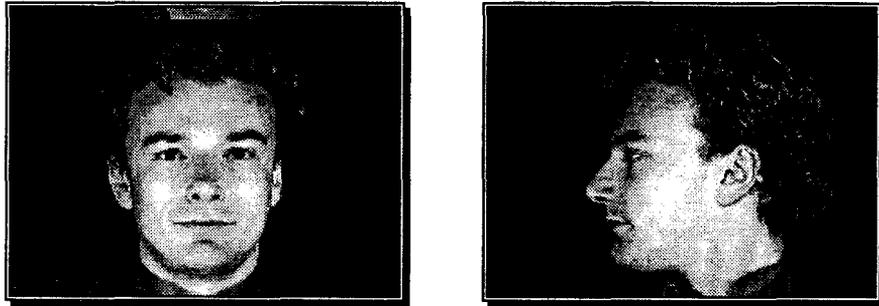


Figure 40 : vue de face et de profil

Les images sont d'abord recalées entre elles deux à deux.

- prétraitement optionnel des images :

L'acteur peut présenter des couleurs très différentes sur les images de face et de profil pour des zones identiques du visage, effet qui peut perturber la reconstruction de la texture générant des effets de type « maladie de peau » pas nécessairement agréables pour l'acteur !

Ces effets peuvent être compensés en sélectionnant interactivement la zone du visage puis en appliquant sur cette zone des filtres pour « lisser » la texture du visage sur les images de face et de profil (Figure 41).

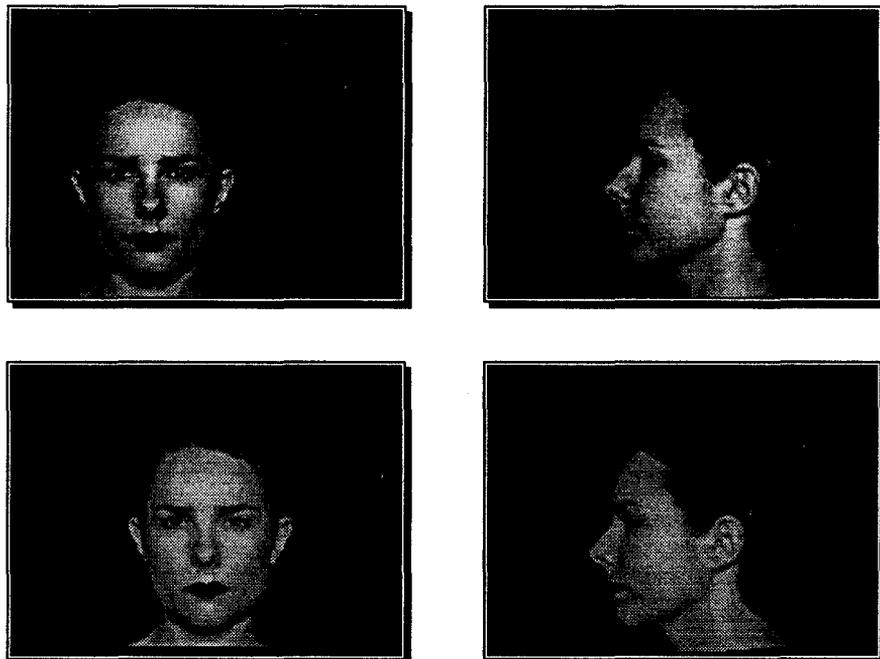


Figure 41 : visages bruts (en haut) et lissés par filtrage (en bas)

- reconstruction géométrie et texture :

Le volume général d'un visage est déterminée par environ une vingtaine de points caractéristiques, en partie symétriques. Une structure géométrique plus complexe du visage est associée à l'ensemble des points caractéristiques et déformée lorsqu'on positionne chacun des points sur le visage traité.

La texture est calculée en fonction de la géométrie précédemment obtenue et des images de face et de profil (Figure 42).

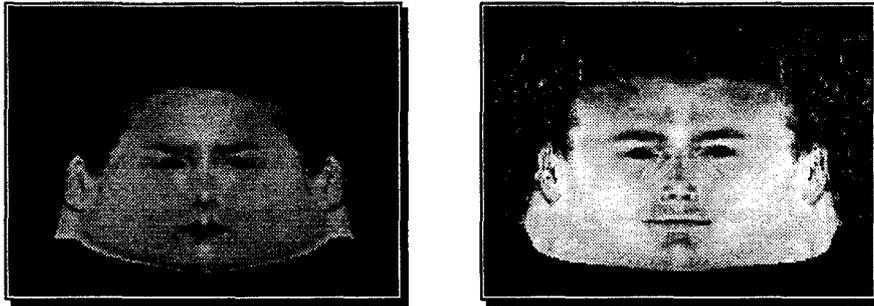


Figure 42 : textures à appliquer aux clones

- le modèle tridimensionnel :

Le modèle 3D est obtenu en plaquant la texture sur la géométrie (Figure 43).

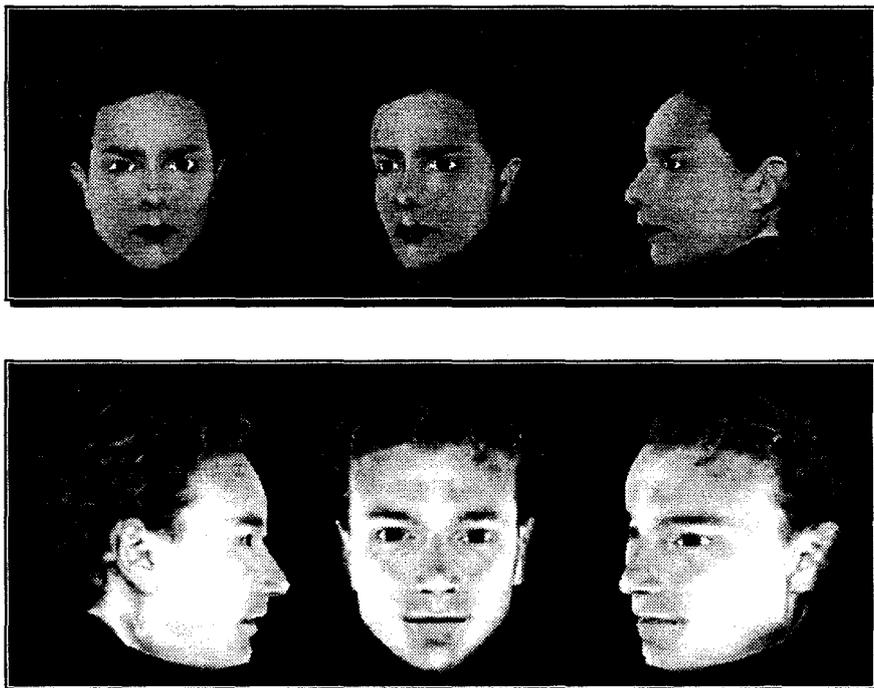


Figure 43 : modèles 3D de clones

iv) Le futur des clones à l'INA

Les améliorations du système de clonage se poursuivent. L'analyse et la synthèse (animation du modèle) sont stables mais la génération initiale du clone (création de la texture et du modèle 3D) nécessite encore une intervention humaine. Il faut faire à la main la correspondance entre certains points caractéristiques sur la texture et sur le treillis 3D initial « moyen » pour adapter celui-ci à la forme de la tête à reproduire. Ce processus est en phase d'automatisation, c'est-à-dire que les points clés seront détectés et ajustés par l'ordinateur. Une fois la partie de création rendue entièrement autonome, n'importe qui pourra se forger son clone pour peu qu'il dispose du matériel nécessaire (juste une caméra et un éclairage adapté). Il existe des algorithmes [Proesmans97] permettant, à partir d'une image, d'extraire des données sur la forme (en 3D) et la texture d'un objet quelconque et, dans le cas qui nous intéresse, d'un visage.

Le reste du corps doit aussi être pris en compte pour réaliser un clone « complet ». Des travaux sont actuellement en cours à l'INA pour reproduire les gestes des bras et les mouvements du tronc, plus importants dans un environnement virtuel que le bas du corps (souvenons-nous des problèmes pour se déplacer). Un seul capteur est placé sur chaque main, la position des bras étant extrapolée par un système de cinématique inverse.

La possibilité de cloner au moins partiellement un être humain pour l'intégrer dans un environnement virtuel 3D existe. Jusqu'à maintenant on trouvait d'une part les avatars du type « blockies », simples à réaliser et à animer et d'autre part des travaux de recherche plus luxueux mais complexes. Cette complexité est normale quand on considère la subtilité du phénomène à reproduire. L'amélioration des ordinateurs mettra très prochainement les clones les plus évolués techniquement à la portée de tous. En parallèle, on leur trouve toujours de nouvelles applications qui les rendent d'autant plus intéressants.

1.3 Les interfaces

Aujourd'hui, les interfaces fenêtrées, ou WIMP, sont universellement présentes. Tous les ordinateurs actuels en sont dotés. Plusieurs systèmes existent mais ils partagent tous des concepts identiques. Il semble très difficile de passer à une « nouvelle génération » qui puisse les supplanter. Pourtant, de nouvelles propositions d'amélioration ou d'évolution éclosent, tandis que le matériel récent permet d'envisager un passage à la 3D qui pourrait révolutionner les outils de travail informatiques.

1.3.1 Les interfaces 2D WIMP

On désigne par WIMP (Window Icon Mouse Pointer) les interfaces informatiques issues des travaux de Xerox, reposant sur la métaphore de bureau. Elles présentent un espace apparenté à un

plan de travail sur lequel sont posés les outils, les documents... Ce plan est morcelé en fenêtres, pouvant se recouvrir et « s'empiler » (Figure 45). Cela permet d'avoir de nombreuses fenêtres de grande taille, contrairement à l'ancienne disposition en « mosaïque » (par exemple dans l'environnement Smalltalk) dans laquelle les fenêtres se juxtaposaient (Figure 44). L'utilisateur agit par l'intermédiaire du couple pointeur/souris. Les icônes sont des pictogrammes symboliques représentant des fenêtres « réduites », des outils, des applications... Certains widgets se sont imposés comme standards, essentiellement les menus déroulants, les boîtes de dialogue ou les ascenseurs.

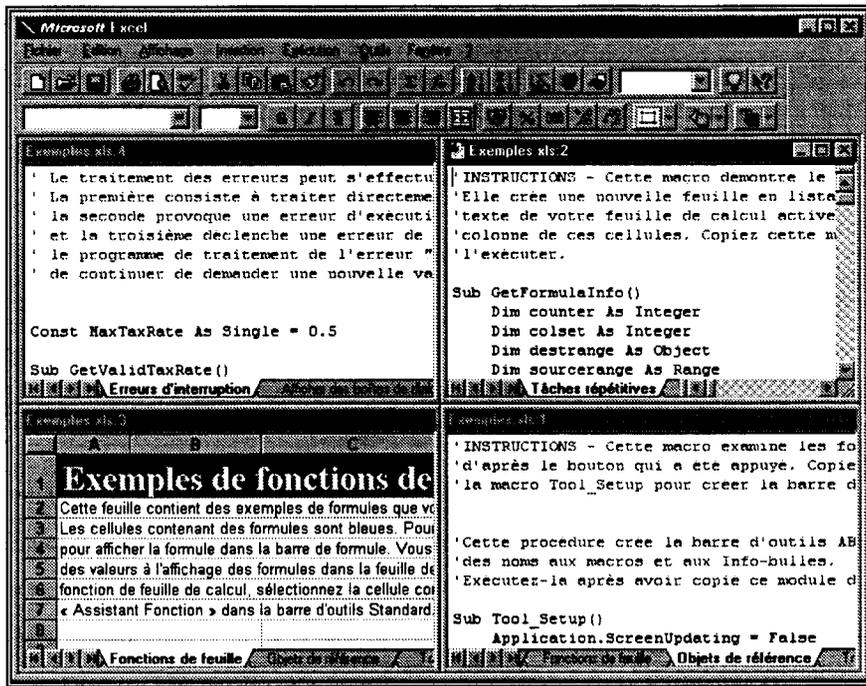


Figure 44 mosaïque de fenêtres

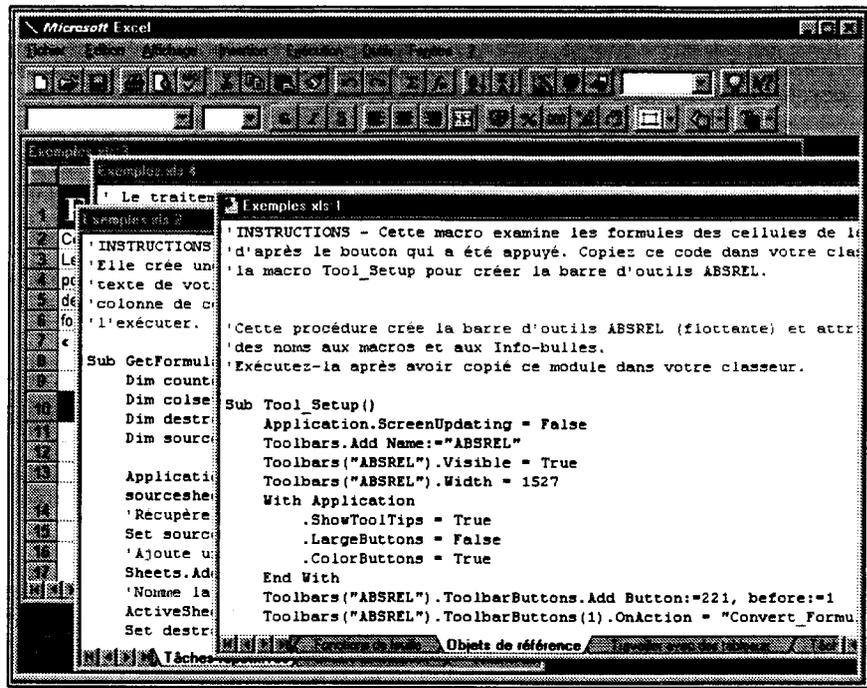


Figure 45 : recouvrements entre fenêtres

Le style de programmation a évolué avec l'arrivée des systèmes de fenêtrage. La programmation procédurale, dans laquelle le programme s'exécutait séquentiellement avec des instructions de contrôle et de branchement explicites, obligeait le programmeur à prévoir toutes les situations auxquelles il allait devoir réagir. Dans la programmation par événement, chaque application dispose d'une file d'attente de messages, auxquels elle va réagir en exécutant les fonctions idoines. Le développement est simplifié, le programme n'a plus à scruter ses entrées sans arrêt pour se dérouler convenablement. Cette méthode est bien adaptée à la conception objet.

Ces interfaces ont accompagné l'explosion de la micro-informatique. Les applications qui les utilisent sont innombrables : traitements de texte, tableurs, environnements de développement, retouche d'images, séquenceurs musicaux, courrier électronique, et même les récents *browsers* du World Wide Web qui permettent la navigation hypertexte dans des documents 2D multimédia. La métaphore de fenêtre a même été étendue au domaine du son [Cohen93]. Une part de plus en plus grande de la production de logiciels concerne l'interface du produit.

X-Window : mécanismes d'implémentation

Si les systèmes de gestion de fenêtres sont nombreux, on peut parfois leur reprocher d'avoir une architecture trop fermée ou limitée. C'est pourquoi nous allons décrire ici X-Window. Ce système, apparu au M.I.T. en septembre 1987, s'est imposé auprès des constructeurs comme un standard dans le monde des stations de travail en uniformisant les interfaces et en permettant le travail sur une machine distante à travers un réseau. Il est maintenant également disponibles sur tous les types d'ordinateurs dotés du système d'exploitation UNIX.

Ce système [Scheifler92] [Nye92] régulièrement mis à jour et en constante amélioration, permet d'écrire une application indépendamment de la plate-forme d'accueil. De surcroît, X-Window est conçu pour fonctionner à travers un réseau, l'exécution de l'application et l'affichage s'effectuant sur des machines différentes. Le parc de machines peut donc être hétérogène, le système d'affichage fournissant un lien permettant d'utiliser des applications ne pouvant être exécutées que par une machine spécifique.

Sur la station utilisée comme terminal, un serveur est lancé : il s'occupe de l'affichage et de la gestion des périphériques. Les applications s'adressent à ce serveur, éventuellement à distance, par l'intermédiaire d'une bibliothèque de fonctions, appelée Xlib [Gettys96], qui envoie les requêtes d'affichage nécessaires selon un protocole prédéfini (Figure 46).

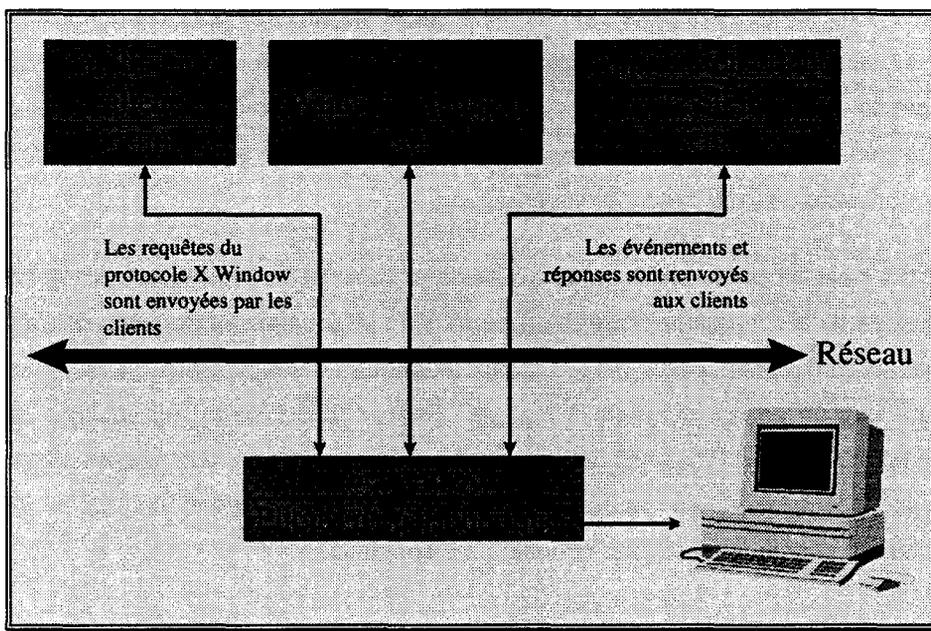


Figure 46 : le mécanisme client/serveur de X-window [Nye92]

Cette architecture offre une grande souplesse. Elle permet par exemple d'installer des gestionnaires de fenêtres divers et variés, chacun offrant une apparence ou des fonctionnalités qui lui sont propres mais dont la compatibilité est assurée par l'utilisation de X-Window. Ainsi, une application fonctionne quelque soit le window manager utilisé (OSF Motif, OpenWin, FVWM, Common Desktop Environment...).

1.3.2 Des proposition d'améliorations

Diverses recherches ont eu pour but de faire évoluer les interfaces 2D, dans l'organisation de l'espace de travail, dans la représentation graphique des fenêtres ou dans la sémantique et allant jusqu'à remettre en cause l'existence de ces fenêtres.

Le concept de pièce

La préoccupation principale qui a conduit à la mise au point par des chercheurs de Xerox [Henderson86] du concept de pièce (*room*) était de proposer une solution au problème d'encombrement de l'écran dans les systèmes à fenêtres classiques. Ils justifient le choix d'offrir à l'utilisateur plusieurs espaces virtuels séparés (plutôt qu'un seul très grand) en analysant l'accès que l'on fait aux données du « desktop » (le bureau, le plan de travail des interfaces WIMP). Ils se sont en effet aperçu que la totalité des documents de travail peut être découpée en ensembles locaux (« *locality set* ») et que l'on passe à moyen terme d'un ensemble à l'autre. Un ensemble de documents est utilisé pour une tâche particulière pendant un moment puis, quand on passe à autre chose, on utilise un autre ensemble de documents. Cette observation s'étend à l'utilisation qui est faite des fenêtres : on utilise à un moment un ensemble de fenêtres (organisées pour éviter les recouvrements entre elles) puis, quand on passe à autre chose, on ferme ou recouvre ces fenêtres pour s'intéresser à un autre ensemble de fenêtres. Pour donner un exemple, l'ensemble consacré au courrier électronique contiendra une fenêtre avec la liste des messages reçus, une avec le message sélectionné et une dernière pour composer un message à envoyer.

L'idée est de créer plusieurs espaces de travail, chacun étant assimilé à une pièce dédiée à une activité particulière. Les contraintes pour qu'un bureau virtuel à plusieurs espaces soit utilisable sont décrites. Il faut d'abord qu'une même fenêtre puisse apparaître dans plusieurs pièces, éventuellement avec à chaque fois un aspect différent. Il faut que certains outils soient accessibles partout : leur fenêtre suivra l'utilisateur dans tous ses déplacements. Par exemple, dans toutes les pièces il faut que l'on sache si l'on a reçu du nouveau courrier. Les pièces sont nommées (« courrier », « développement ») afin de leur attribuer une étiquette sémantique.

Le déplacement entre les pièces est étudié. En fait, quand l'utilisateur crée une pièce réservée à un usage particulier, il indique à quels autres endroits elle est connectée (ceci par des portes dessinées à l'écran). Le passage à travers une porte revient à se téléporter, c'est-à-dire à changer brutalement de contexte, l'ancienne pièce étant remplacée par celle de destination. Par commodité, une porte est automatiquement ajoutée pour le retour dans la pièce précédente. Pour des questions de rapidité, il y a aussi un plan accessible à tout moment (une liste des pièces en fait) qui permet de sélectionner directement une destination, ainsi que de visualiser les passages entre pièces.

Ces mécanismes existent maintenant dans la plupart des gestionnaires de fenêtres, sous le nom de « bureaux virtuels ».

Rooms est devenu plus tard *Rooms3D*, concept utilisé par exemple dans l'Information Visualizer. Là aussi, une vue d'ensemble des différentes « pièces 3D » est disponibles pour passer facilement de l'une à l'autre.

Le symbolisme dans les interfaces

Loretta Staples [Staples93] propose diverses techniques pour donner de la profondeur, réelle (perspective) ou sémantique, aux interfaces 2D classiques, en s'appuyant sur une étude de l'histoire de l'art, par laquelle elle dégage des symboles et des techniques transposables dans une interface graphique.

L'article commence par une mise en garde contre une approche complètement réaliste qui, si elle semble naturelle dans certains cas, n'est pas applicable pour toutes les applications. De plus, on retomberait sur les mêmes gênes, les mêmes problèmes que dans certaines situations réelles (classement, représentation et recherche d'information par exemple). Inversement, trop d'abstraction ferait perdre tout point de repère habituel à l'utilisateur. Le réalisme et l'abstraction doivent être utilisés conjointement. Les abstractions possibles doivent être systématisées pour que leur signification soit toujours la même.

La projection perspective permet de représenter la profondeur et peut servir à indiquer un éloignement spatial ou temporel. L'effet peut être amélioré en utilisant un éclairage ou en mettant des ombres aux objets. La transparence permet de distinguer les objets plus lointains qui autrement auraient été cachés. Les fenêtres peuvent alors avoir une épaisseur, indiquant la nature du document ou sa taille.

Un objet peut être mis en valeur en l'éclairant, en lui ajoutant un halo, voire en supprimant son ombre pour lui donner une position « irréelle ». Ces procédés sont utilisés dans la peinture religieuse, dénotant du caractère différent, sacré de la personne ou de l'objet auréolé. De la même manière, un objet rendu opaque parmi des objets transparents prend de l'importance.

La sémantique des fenêtres – rôle de la transparence

Axel Kramer [Kramer94] critique l'utilisation de fenêtres pour les tâches créatives. Elles imposent une structure rigide au travail à effectuer, qui ne conviennent pas quand on a besoin avant tout d'émettre des idées, d'en discuter, d'y revenir. Ce processus social a avant tout besoin de moyens de négociation, donc de communication, pour aboutir. Ces critiques sont plus dirigées vers l'usage qui est fait des fenêtres, ce à quoi elles sont associées. Une nouvelle métaphore, proche du papier calque, est présentée : le *patch translucide*¹⁷ (Figure 47).

¹⁷ le terme « *semi-transparent* » serait plus juste

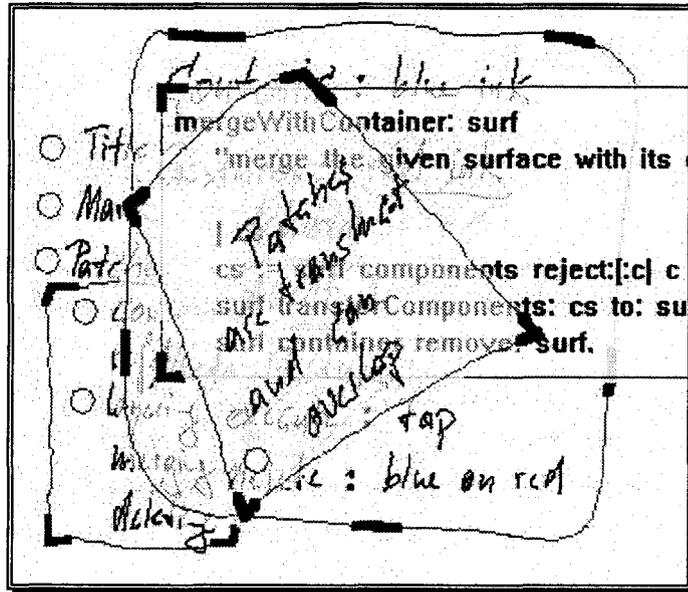


Figure 47 : patches translucides

Les fenêtres ont avant tout répondu au besoin d'afficher plusieurs contextes d'information en même temps. Après les premières expériences utilisant plusieurs écrans physiques, les fenêtres sont apparues d'abord côte à côte, réalisant un découpage, ou plutôt un pavage (*tiling*) de l'écran pour finir par avoir des fenêtres pouvant se recouvrir. La métaphore de bureau provient de Xerox Star et a peu évolué depuis, en particulier dans les produits effectivement commercialisés et utilisés.

Au début, chaque fenêtre principale correspondait à une *application*. Des systèmes pour la gestion de fenêtres dépendantes ont été mis au point, permettant de gérer menus, barres d'icônes, fenêtres « filles » (sous-fenêtres incluses hiérarchiquement dans la « mère »). Le copier-coller entre applications différentes était possible, le document récepteur se contentant d'inclure une image du document. Pour modifier l'élément étranger inclus, il fallait revenir à l'application l'ayant généré, effectuer la modification et remettre à jour la copie présente dans le document.

Pour faciliter l'inclusion de données de différentes natures, on a cherché à associer une fenêtre à un *document*. Celui-ci passait de produit d'une application à assemblage d'informations issues de diverses applications. La modification d'une composante se fait sans changer de fenêtre, directement sur le document. Par contre, la fenêtre va fournir les outils propres à l'application concernée (les menus par exemple). On trouve ce genre de mécanismes sous différents noms : Microsoft OLE, OpenDoc...

Il reste que le résultat est prédéterminé par les méthodes de travail qu'imposent les applications générant le document final. Il est difficile de griffonner, de faire des schémas intermédiaires, de mettre ses idées en images et en formes. Idéalement, l'ordinateur ne devrait s'imposer que pour réaliser des calculs ou d'autres opérations que lui seul sait faire mais sans imposer une représentation rigide pendant la phase de conception.

Kramer propose de dissoudre l'association statique entre la représentation, sa structure et son interprétation. A l'inverse, il recommande une association dynamique dirigée par l'utilisateur. Dans l'ordre, l'utilisateur va écrire (interface à stylo) sur un afficheur électronique, puis il groupera et structurera les différentes informations pour finir par leur appliquer une interprétation. L'exemple donné est celui de l'architecte qui esquisse un plan. Au fur et à mesure, il note dans un coin le coût des éléments qu'il ajoute. Quand tout est mis en place, il groupe ces chiffres et demande à faire calculer la somme. Ce genre d'activités est fortement susceptible de réunir plusieurs personnes. Les travaux présentés ici s'inscrivent d'une certaine manière dans notre problématique, même si leur contribution la plus importante ne concerne pas directement le travail coopératif distribué.

Les patches présentés sont transparents (partiellement) et de formes irrégulières (i.e. non rectangulaires). La transparence a deux propriétés importantes. La première est de préserver le contexte. Contrairement aux fenêtres opaques, un patch translucide laisse encore voir ce qui est placé en dessous de lui. La deuxième est de protéger le « contenu ». En plaçant un calque sur un plan, on peut y faire des modifications sans l'abîmer et s'y reprendre à plusieurs fois en repartant de l'original. Les rectangles apparaissent rarement dans un processus créatif. Les formes irrégulières sont par contre monnaie courante, elles permettent de structurer le travail, de grouper des idées jetées ici et là sur le papier.

Plusieurs opérations de base peuvent être appliquées à un patch : création, destruction complète, fusion (le patch disparaît en tant que tel mais les données restent, transférées dans le patch du dessous), effacement des données contenues dans le patch, élargissement, rétrécissement, monter, descendre et ajout d'une « perle ». Ces perles sont des poignées permettant de manipuler un patch (déplacement en particulier). Un petit langage gestuel spécialisé a été créé pour déclencher toutes ces opérations. Les gestes ont été étudiés pour évoquer facilement l'action qu'ils représentent. Certaines actions sont immédiates, d'autres doivent être validées. Par exemple, la destruction d'un patch (un zigzag recouvrant tout le patch, Figure 48) doit être confirmé en « tapant » avec le stylo. Les actions en attente peuvent être manipulées tout comme un patch et être éventuellement détruites. Il est également possible de changer l'interprétation d'un patch (zone de texte, calculatrice...) mais l'article ne donne pas d'indication sur la manière de le faire. Donner une interprétation particulière à un patch revient à lui ajouter des fonctions : le patch doit reconnaître de nouveaux gestes et effectuer l'action correspondante dans le contexte de cette interprétation. Le fond est un patch « basique » qui reconnaît un ensemble minimum de gestes. Quelques pistes sont données pour l'implémentation de ces patches.

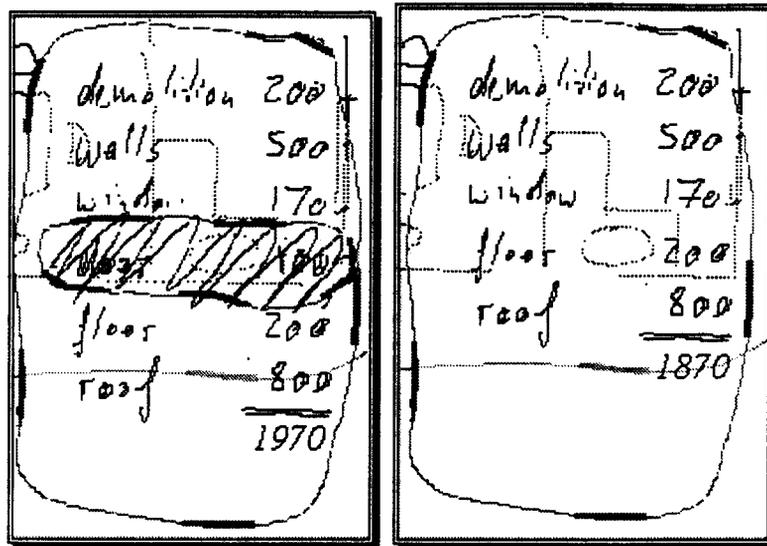


Figure 48 : destruction d'un patch

Les patches translucides unifient en un concept unique fenêtres, sous-vues et sélections. Le contenu n'est plus lié à une fenêtre. A la place, les données sont associées et interprétées dynamiquement sur demande de l'utilisateur. La structuration est rendue plus ou moins rigide à volonté. Le « vocabulaire » gestuel est long à mettre au point, il doit être à la fois cohérent et permettre une interaction efficace avec le système. Le travail encore à réaliser qui semble le plus important concerne les interprétations. La facilité à développer et ajouter des interprétations est vitale pour ce système, une méthode générale doit être mise au point. Il ne faut pas oublier que chaque interprétation suppose l'utilisation de nouveaux gestes et de nouveaux comportements qui lui sont spécifiques, et donc à chaque fois la mise au point d'un nouveau vocabulaire ! Une question qui n'est pas directement soulevée est celle de l'apprentissage. Si le vocabulaire de base semble simple à appréhender, la multiplication des interprétations possibles va entraîner un effort de mémorisation important. Il va également falloir trouver des méthodes pour expliquer ces gestes aux utilisateurs. Une « aide gestuelle en ligne » pourrait peut-être résoudre le problème. Des évaluations sont nécessaires quant à la mise en œuvre mais le concept d'association dynamique d'une interprétation et d'une structure à des données est clairement une évolution intéressante par rapport aux classiques fenêtres opaques.

D'autres travaux utilisent la transparence, comme le « *silk cursor* » [Zhai94], un pointeur 3D (un volume plutôt qu'un point comme à l'habitude) utilisant la transparence pour repérer sa position par rapport aux autres objets. Dans [Harrison95], la superposition de plans transparents différents est étudiée. L'homme est capable de passer mentalement d'un plan à un autre et de rester focalisé sur une image précise même si des événements inattendus surviennent. Quelques conseils sont donnés pour garder les plans superposés lisibles. Le « *Toolglass* » [Bier93] combine une palette d'outils semi-transparente, qui pourront être appliqués directement sur un document 2D. A noter que deux souris sont utilisées, une pour positionner la palette, l'autre pour appliquer des actions au document. D'autres exemples d'utilisation de la transparence sont présentés dans [Zhai96], notamment pour améliorer la perception de la profondeur.

1.3.3 Les vues fish-eye

Les vues fish-eye ont pour objectif d'intégrer une vue globale et une vue détaillée dans une vue unique. Elles ont été nommées par analogies avec les objectifs d'appareils photographiques à grand angle, qui fournissent les lieux proches en détail mais aussi les régions plus éloignées avec de moins en moins de détail.

Cette vue correspond à des schémas mentaux naturels. [Furnas86] a interrogé des volontaires en leur demandant de citer quelques présidents des Etats-Unis. Les listes comportaient d'une part des présidents récents (donc « proches ») et d'autres anciens mais célèbres (donc « importants »). Cette organisation se retrouve dans les quotidiens régionaux. Les événements locaux sont détaillés, ainsi que ceux éloignés mais (objectivement) de grand poids.

Furnas a formalisé le « degré d'intérêt » (DOI) qui évalue l'intérêt d'une information particulière pour l'observateur par la formule suivante :

$$DOI_{\text{fisheye}}(x|y) = API(x) - D(x,y)$$

où y est le point que regarde l'observateur, x est celui dont l'importance est à évaluer, $API(x)$ l'importance a priori de x et $D(x,y)$ la distance entre x et le centre d'intérêt y . Le degré d'intérêt augmente avec l'importance a priori et diminue avec la distance. En prenant l'exemple d'un arbre hiérarchique, la distance peut être le nombre de branches entre les nœuds x et y , l'importance a priori étant inversement proportionnelle à la profondeur d'un nœud dans la hiérarchie. Les fonctions API et D peuvent donc être très abstraites du moment qu'elles ont une signification.

Cette valeur du degré d'intérêt est ensuite interprétée pour construire une vue « filtrée » de l'information globale d'origine. Dans l'exemple de l'arbre, les branches de faible importance seront élaguées.

Les méthodes de construction d'une vue à partir du degré d'intérêt sont multiples. Des méthodes spécifiques sont développées selon la nature et la structure des informations à représenter, comme par exemple des graphes [Sarkar92]. Le problème de la navigation dans un grand nombre d'informations issues de moteurs de recherche sur le World Wide Web est explorée dans [Vernier97]. Une fonction de déformation de l'espace est définie. Elle a comme paramètres la distance (en pixels, le système exposé agissant sur une image 2D), un rayon de visualisation, et deux autres valeurs déterminant la puissance de la « loupe » ainsi créée. Il est intéressant de constater que Furnas construisait une vue à partir d'information filtrée a priori, tandis qu'ici c'est la distorsion graphique qui fait office de filtre. Déformation polaires et cartésiennes sont utilisées conjointement pour fournir différentes vues. [Sheelagh95] construit une déformation disposant de plusieurs centres d'intérêt et non plus d'un seul. Les informations peuvent être détaillées en plusieurs endroits. C'est utile par exemple pour établir un itinéraire d'une petite ville à une autre : on détaille le chemin jusqu'à un grand axe routier et aux alentours du point d'arrivée. Ces techniques de visualisation sont extensibles à la

visualisation 3D [Sheelagh97] (Figure 49), avec le problème d'occlusion à résoudre : le point d'intérêt ne doit pas être masqué par d'autres objets placés devant lui.

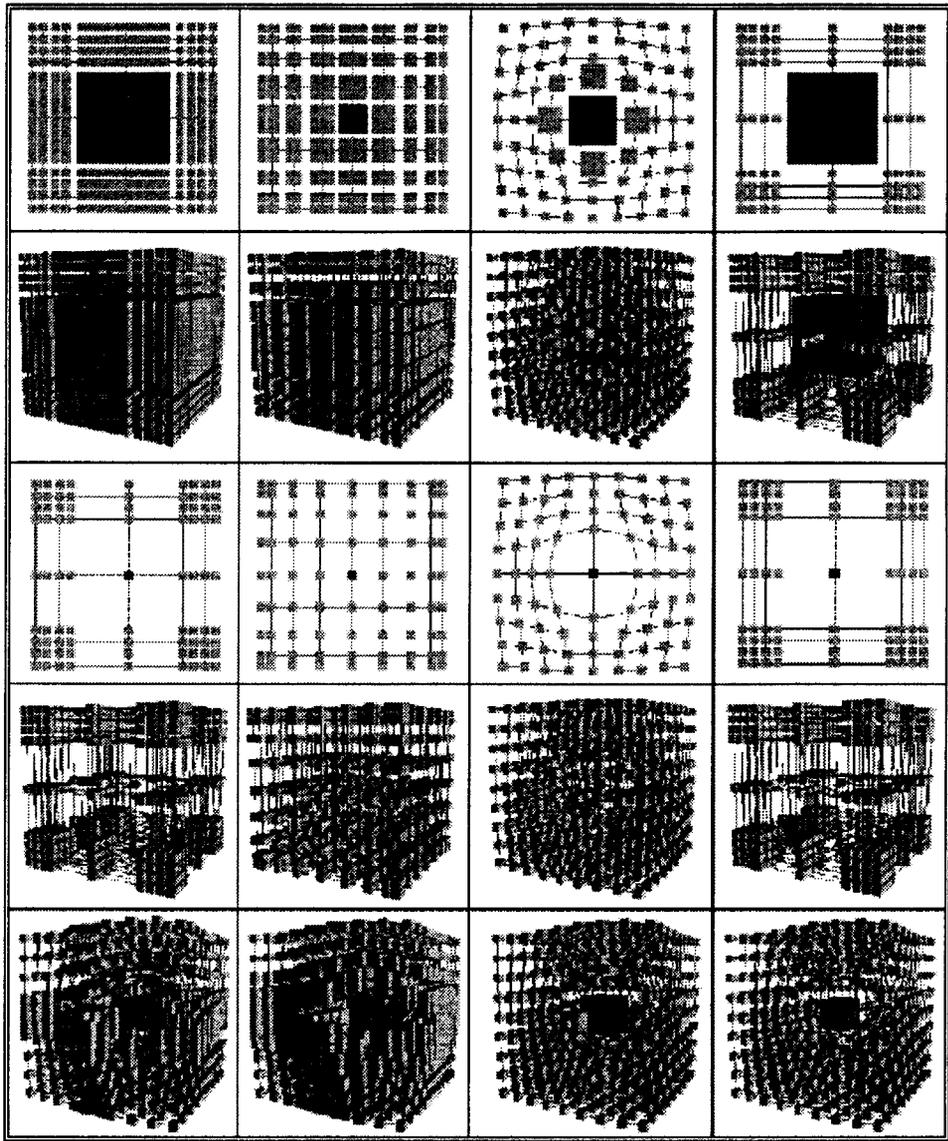


Figure 49 : différents schémas de distorsion 3D

1.3.4 L'« *Information Visualizer* »

Xerox a mené des recherches sur différentes façons d'afficher des données en utilisant une représentation 3D, aboutissant au paradigme d'*information workspace* dont l'*Information Visualizer* [Robertson93] est une mise en oeuvre. Les deux exemples les plus connus sont le « *Perspective Wall* » et le « *Cone Tree* ».

Le *Perspective Wall* [Mackinlay91] (voir Figure 50) montre une série de données dans son intégralité, malgré la place limitée sur l'écran, tout en gardant ces données exploitables. Ces données sont affichées sur trois « murs ». Au centre, un mur fait face à l'observateur. Les données qu'il contient

sont affichées normalement, lisiblement ; elles constituent typiquement le centre d'intérêt de l'observateur. De chaque côté, un mur contient les données voisines. Ces deux murs plongent vers le fond de l'écran. Les données les plus proches de l'observateur, et donc de son centre d'intérêt, sont encore bien visibles. Les plus éloignées sont illisibles mais l'observateur a quand même une idée de leur organisation. Le Perspective Wall dégrade naturellement la qualité de représentation au fur et à mesure que l'on s'éloigne du centre d'intérêt par une mise en perspective des données.

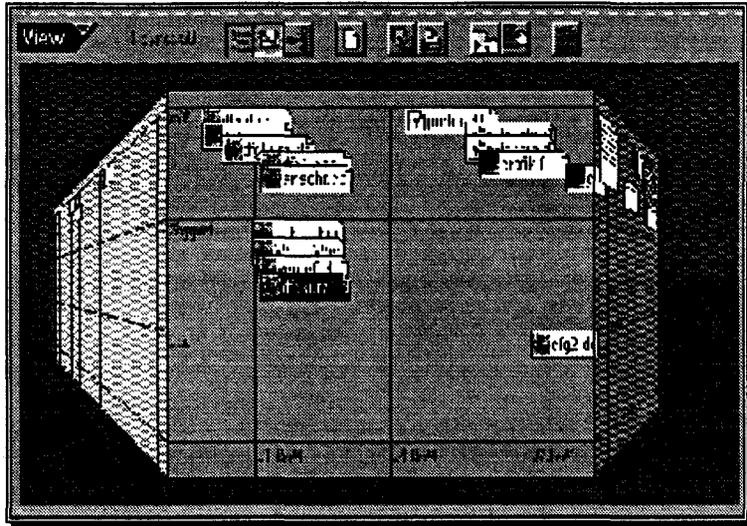


Figure 50 : Perspective Wall

Une application possible est la gestion de rendez-vous sur un agenda. Au centre, la journée courante est détaillée : heures, lieux, sujets sont lisibles. Les murs en perspective, s'ils ne permettent pas de connaître le détail des jours passés et suivants, donne quand même une idée de leur « taux de remplissage » et permet de trouver facilement un créneau libre pour fixer un nouveau rendez-vous. Ce type de représentation est particulièrement adapté aux données successives, un planning par exemple, même s'il peut être utilisé pour montrer des objets graphiques « larges » comme l'arborescence de la Figure 51 ou l'ensemble de listes de la Figure 52.

Pour rendre supportable la dégradation qu'apporte la perspective, il faut donner la possibilité à l'observateur de pouvoir consulter dans de bonnes conditions une donnée particulière, c'est-à-dire changer son centre d'intérêt. Le Perspective Wall est intrinsèquement dynamique : il est possible à tout moment de faire glisser les données le long des trois murs pour se déplacer dans l'ensemble des données. L'animation, dirigée par l'utilisateur, pallie instantanément la perte de qualité inévitable quand on densifie l'information à l'écran.

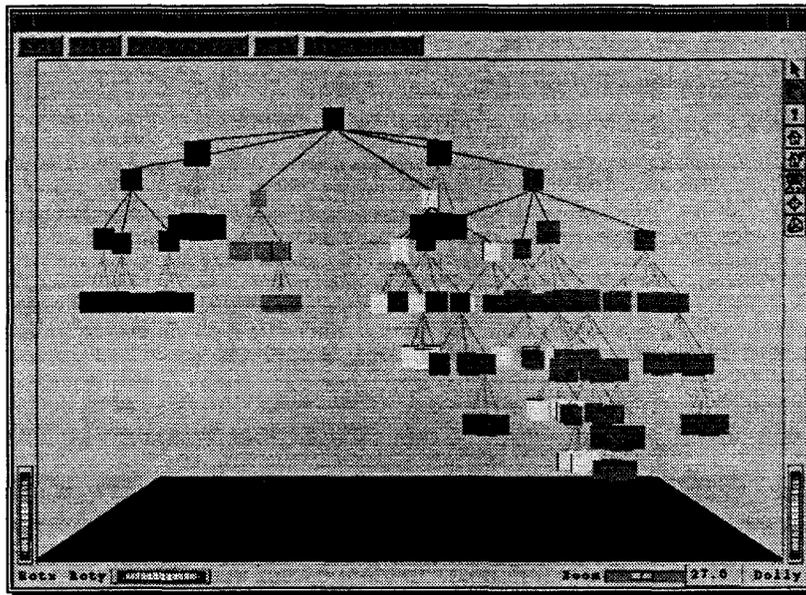


Figure 53 : une arborescence « Cone Tree »

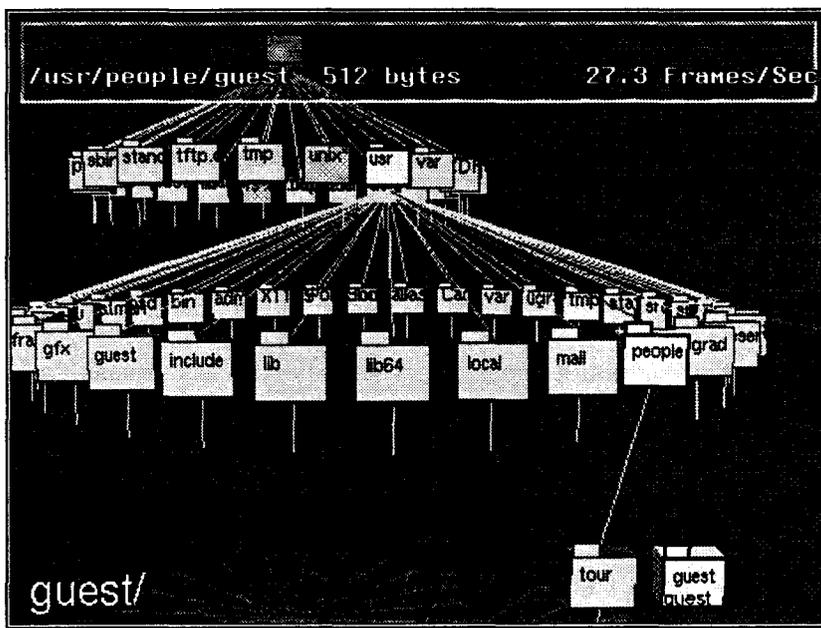


Figure 54 : une arborescence de fichiers et de répertoires sous forme de Cone Tree

L'interaction est étudiée pour rendre les manipulations plus faciles, agréables, rapides. Par exemple, la navigation se fait selon des trajectoires logarithmiques : loin de l'objet, le déplacement est rapide et il s'affine et devient plus précis au fur et à mesure que l'on s'approche.

D'autres modes de représentation ont été développés, comme les *datamaps* ou le *calendar visualizer* [Mackinlay94]. Chacun d'eux est spécialisé dans un type d'informations (champ de données 2D, emploi du temps et données à caractère temporel) mais tous conservent en commun l'affichage 3D et l'animation interactive.

La représentation 3D, alliée à l'animation interactive, facilite l'accès, la gestion, le traitement simples, rapides et assistés de l'information. Elle permet de densifier l'information présente à l'écran d'une manière facile à comprendre, n'occasionnant pas de surcharge cognitive pour l'utilisateur. Au contraire, ce qui aurait nécessité de la compréhension (s'y retrouver dans un grand nombre de données alors qu'on n'en voit que peu à la fois dans une fenêtre) fait directement appel aux capacités de perception.

1.3.5 La perception et le repérage en 3D

Un environnement 3D ne peut être utilisable que s'il est correctement perçu. En interface, et tout particulièrement pour la manipulation directe, les positions et relations des objets visibles doivent pouvoir être aisément déterminées par l'observateur. La scène doit contenir des indices visuels permettant de se repérer. Wanger a mené des expériences sur trois tâches différentes à réaliser : le positionnement, l'orientation et le dimensionnement d'objets [Wanger92]. Chaque épreuve a été subie avec ou sans certains indices supposés utiles.

Il est apparu que la projection en perspective et la projection des ombres des objets sur le sol aidaient à positionner correctement les objets (entre eux). Pour l'orientation, la perspective est un handicap et rend la tâche plus difficile. Il est par contre utile de pouvoir bouger dans la scène. Quant au dimensionnement des objets, il est plus précis dès lors que l'on est aidé par les ombres et que l'on peut bouger pour changer de point de vue. Dans tous les cas, le fait de plaquer une texture sur les objets n'a pas amené de changement représentatif.

On retrouve un phénomène connu en psychologie de la perception, notamment étudié par Gibson [Gibson66] [Gibson79], à savoir que le mouvement est bien l'une des étapes de la vision et de la compréhension de ce qui est perçu. L'obtention d'information par la perception est un processus actif et fait appel à l'exploration. C'est également valable pour le toucher [Cadoz94], l'ouïe (tourner la tête pour repérer la source d'un bruit).

2. Analyse et recommandations

2.1 Observations générales

Les travaux étudiés précédemment nous fournissent plusieurs sortes de renseignements. En retrouvant quelles ont pu être les motivations de leurs auteurs, nous comprendrons mieux comment s'est faite la conception de ces différents outils de travail coopératif et ce qui a prévalu dans les choix de design. Quant à la Réalité Virtuelle, même si elle ne tient pas toutes ses promesses, elle ouvre des pistes pour la création d'interfaces nouvelles.

Nous étudierons ensuite en détail deux projets de grande envergure que nous avons déjà évoqués. MAJIC est en quelque sorte un aboutissement des outils utilisant la vidéo classique. MASSIVE est un exemple poussé d'exploitation de la Réalité Virtuelle pour une utilisation par un grand nombre de personnes simultanément, en particulier dans un but collaboratif. L'un comme l'autre sont des références, des modèles très avancés, sans compromis dans la recherche de la qualité des services rendus. Ils ont été des sources d'inspiration tout au long de notre travail. C'est pourquoi nous jugeons pertinent de les présenter plus longuement, dans leur globalité et non sur quelques points particuliers.

Nous énoncerons finalement trois listes de recommandations et de conseils pour la réalisation d'un outil de travail coopératif, lesquels ont été mis en application dans le projet SPACE.

2.1.1 Les motivations à l'origine de ces projets

Les produits commerciaux doivent être à la fois rapides à concevoir, peu chers et faciles à utiliser. D'une manière générale, ils sont développés à partir de technologies préexistantes, de préférence largement diffusées, bien maîtrisées et évitent de trop grands investissements de recherche. A partir d'une architecture de base, ils peuvent être déclinés en plusieurs versions répondant à des besoins spécialisés. Les utilisateurs retrouveront du matériel, des méthodes de travail qu'ils connaissent déjà et dont la prise en main ne nécessite pas un trop lourd apprentissage. Les outils de visioconférence entrent dans cette catégorie. Leurs équivalents informatiques utilisent les interfaces à fenêtre classiques en bureautique. Les recherches sur les Hydra préfigurent un futur outil coopératif qui, à l'instar du téléphone, garderait le côté « petit appareil de bureau » et serait capable de devenir rapidement familier.

Certains partent du principe que l'outil doit se faire oublier et que le meilleur moyen d'y parvenir est de simuler la réalité le plus fidèlement possible. Même si MAJIC utilise du matériel audiovisuel commun (hormis l'écran novateur), l'utilisation qui en a été faite est novatrice. La conception définitive du prototype ne s'est établie qu'après de nombreuses études, plus sociologiques que techniques. L'intérêt de ce projet vient autant des préceptes énoncés que de la mise en œuvre qui en a été faite.

L'informatique a pour principaux intérêts d'être souple et adaptable, et de permettre la production et la sauvegarde de documents partagés. On l'emploie soit comme succédané d'outils vidéo (Picture Tel...), le partage d'applications en plus, soit pour la Réalité Virtuelle. Remarquons que, si la CAO profite depuis longtemps de la 3D, il n'existe toujours pas d'interface 3D généraliste complète. Les mondes virtuels sont ludiques et attractifs, mais pas encore répandus dans le monde du travail. Les approches du type « Réalité Virtuelle » veulent mettre à profit sa souplesse. Si les images gardent un caractère artificiel, la possibilité d'interagir avec l'environnement, de pouvoir se déplacer à volonté, explorer, rendent ces mondes synthétiques immédiatement attirants. A partir d'un même « moteur » de base, il est possible de produire des environnements dont les usages seront différents. C'est le cas des projets DIVE et MASSIVE, le premier permettant de développer des applications spécifiques (ajout d'images vidéo réelles à une scène synthétique, simulation, robotique...), le second se spécialisant dans le travail coopératif.

D'autres veulent proposer de nouvelles méthodes de travail. Dans cette catégorie, le ClearBoard est une réussite. Il part d'une métaphore réaliste, la vitre, mais parfaitement inutilisée. Il l'améliore en évitant l'effet « miroir » et en n'obligeant plus le plan de travail à être vertical. Pour le travail à deux, il est plus facile de simuler cette métaphore que de trouver un moyen de représenter un tableau devant lequel les deux personnes doivent se tenir.

2.1.2 Les limites de la Réalité Virtuelle

Il est clair que les espoirs fondés sur la Réalité Virtuelle doivent être tempérés. Nourrir les perceptions humaines de sensations artificielles mais réalistes risque de rester encore longtemps une

utopie. Les périphériques ne sont pas à la hauteur (en termes de précision, de vitesse de réaction, d'encombrement etc.). Le retour d'effort tactilo-proprio-kinesthésique par exemple est totalement hors de portée car il doit prendre en compte la sensibilité des doigts, la force des bras, réagir dans un laps de temps très court et à une fréquence élevée. Les contraintes mécaniques, la puissance de calcul sont insuffisantes dans un cadre général. De plus immerger un utilisateur dans un monde artificiel nécessite d'en posséder un modèle fidèle, qui doit comprendre les interactions entre l'homme et cet environnement. Des approximations des lois physiques sont insuffisantes. Le bilan que fait Claude Cadoz est éloquent [Cadoz94] : « (...) Énumérer toutes les applications ici reviendrait à écrire un chapitre sur les Réalités Virtuelles, ce qui est exclu vu la quantité d'idées fausses et d'illusions qu'il faudrait commencer par détromper. Il convient en effet, à propos de ces dispositifs, gants ou combinaisons, de mettre les choses à une plus juste place que celle que leur ont donnée les différentes publicités sous l'étiquette des Réalités Virtuelles. On a parlé de manipulation d'objets virtuels, de déplacements, de voyages dans des mondes qui ne le sont pas moins, dans l'intérieur de l'ordinateur, etc. Si le principe de la projection d'une image de l'opérateur dans l'espace informationnel de l'ordinateur est intéressant au fond des choses, ses mises en œuvre effectives sont loin de répondre aux fantasmes que les publicités n'ont cessé d'exciter depuis un certain nombre d'années. Les performances, aussi bien des capteurs que des machines et des systèmes de simulation dans cette vision totalitaire des systèmes à immersion, sont très en dessous de ce que la perception et le système cognitif humain exigent pour croire, une fois le premier effet de surprise passé, qu'il s'agit de réalité. » Au final, on adapte l'environnement simulé à ce que l'on sait faire. Certains insèrent des fenêtres 2D dans un environnement 3D, pour interagir à l'aide des widgets 2D habituels [Coninx97] (Figure 55). D'autres représentent le monde en petit dans lui-même pour faciliter la navigation, toujours difficile en Réalité Virtuelle : ce sont les « *Worlds In Miniature* » [Stoakley95] (Figure 56).

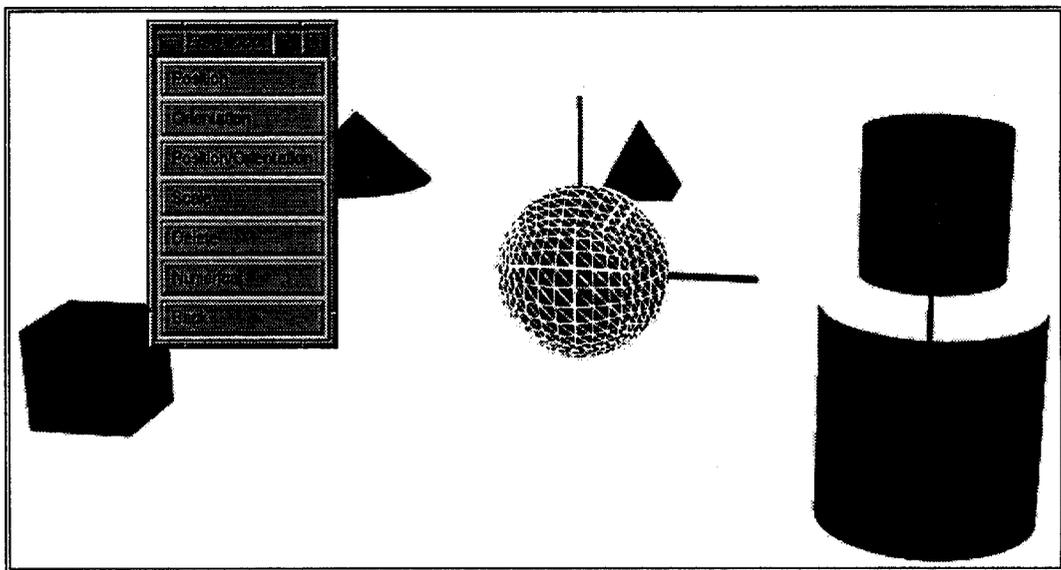


Figure 55 : réutilisation des widgets 2D en Réalité Virtuelle

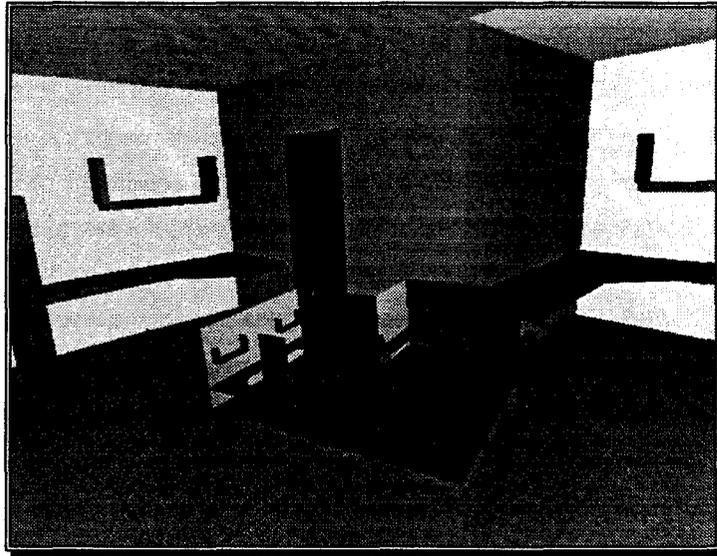


Figure 56 : les Worlds In Miniature

Les technologies développées spécialement pour la Réalité Virtuelle n'en restent pas moins dignes du plus grand intérêt et n'auraient peut-être jamais vu le jour sans l'engouement généralisé suscité par les nouveaux mondes synthétiques. La représentation de personnes sous forme d'objets synthétiques, les périphériques de sortie (images et sons spatiaux) et d'entrée (multitude de capteurs), l'animation, les agents autonomes, la synthèse d'images (le rendu mais aussi les algorithmes de reproduction de comportements ou de simulation de corps mous par exemple) ont tous progressé. En les utilisant à bon escient, sans se focaliser sur la reproduction exacte de la réalité, on peut maintenant fabriquer de nouveaux outils informatisés qu'il faut considérer comme tels et qui ouvrent de nouvelles perspectives, en particulier dans le domaine du travail coopératif.

2.1.3 L'évolution vers les interfaces 3D

Inconvénients des interfaces WIMP

Les systèmes à fenêtres, bien que très performants quand il s'agit de travailler seul sur des applications de type « bureautique » (typiquement traitements de texte, tableurs, mais aussi retouche d'images, programmation etc.) ne supportent que difficilement le TCAO¹⁸ synchrone.

Dans ce cadre, la métaphore de bureau est inadéquate. Il ne s'agit plus uniquement de représenter des documents (au sens large). Il faut aussi montrer et faire intervenir les personnes distantes avec qui s'effectue la collaboration. Il manque une « dimension » capable d'accueillir de manière satisfaisante ces personnes, distincte du plan de travail. C'est ce problème que la

¹⁸ Travail Coopératif Assisté par Ordinateur

TeamWorkStation de ClearFace essaie de résoudre [Ishii91] par la superposition et surtout l'utilisation de la transparence.

Les fenêtres WIMP sont indépendantes. Lors d'une session de TCAO synchrone, des liens sont créés entre les personnes (contact visuel par exemple) ou entre un document et une personne (tel individu manipule ce document). Rien n'existe pour représenter ces liens. Desktop-MAJIC, que nous allons aborder dans la description complète de MAJIC (voir 2.2.1) essaie de résoudre ce problème. Les solutions qu'il propose sont efficaces mais il s'agit d'astuces très simples qu'il est difficile d'étendre et d'améliorer. Le ClearBoard réussit à la fois à intégrer l'interlocuteur et le document sur le même écran, à montrer les liens les reliant mais est hélas limité à deux utilisateurs.

Pour finir, signalons que par leur nature 2D, les fenêtres et la souris sont loin d'être idéales pour les applications 3D et qu'il faut utiliser des moyens détournés pour la manipulation. Les outils de CAO, les modélisateurs de synthèse d'images travaillent avec des vues 2D sans perspective (vues de côté, de dessus...). L'information Visualizer de Xerox a été étudiée spécialement pour se manipuler à la souris et au clavier. Des actions complexes, des animations sont déclenchées d'un simple clic. Mais la mise au point de ce comportement est long et difficile et ne constitue en aucun cas un modèle général d'interaction 3D. Les manipulations sont artificielles, suffisamment éloignées des gestes réels pour que leur représentation aux autres participants pose problème.

Le passage à la 3D

Comme nous l'avons vu, les interfaces 2D utilisées actuellement ne constituent pas un aboutissement. Elles pourraient être encore améliorées. Mais tout se passe comme si la technologie avançait plus vite que la réflexion sur la création de nouvelles interfaces. Xerox Star et ses séquelles, en l'occurrence les interfaces commerciales adoptant la métaphore de bureau (Apple Macintosh, X-Window et Microsoft Windows entre autres) ont précédé les machines ayant une puissance suffisante pour les accueillir. Maintenant, la 3D est partout mais reste cantonnée aux domaines « évidents » de la simulation, du jeu, de l'animation et des trucages, ou de la Réalité Virtuelle. La représentation de données s'est tout de même bien développée, notamment pour la visualisation scientifique. En effet, une présentation bien choisie d'un ensemble de données peut permettre de percevoir naturellement une structure, des points d'accumulation, etc. qui n'apparaissent pas immédiatement dans les données brutes. Il semble acquis que l'affichage en trois dimensions, en densifiant et organisant l'information, permet de mieux la comprendre grâce à une perception plus intuitive, directe, que lorsque l'on se limite à deux dimensions.

Néanmoins il n'y a toujours pas d'interface 3D « véritable ». La Réalité Virtuelle n'est pas une interface dans le sens où elle cherche justement à reproduire les interactions réelles, directes, sans intermédiaire. Notre projet va intégrer des éléments issus de ces deux évolutions. Nous voulons mettre à profit les techniques permettant d'offrir une scène réaliste, tout en gardant l'efficacité, les méthodes d'interaction que l'on trouve dans les interfaces.

2.2 Deux projets à la loupe

Nous allons maintenant revenir plus en détail sur deux plates-formes permettant de tenir des réunions à distance. La première, jouant la carte du réalisme et de la sensation de présence, est MAJIC. Les techniques audiovisuelles sont ici exacerbées pour arriver à les faire oublier. L'autre est un représentant de la catégorie « Réalité Virtuelle », MASSIVE, dont l'histoire serait incomplète sans aborder celle de DIVE avec lequel il a de nombreux points communs. Nous avons choisi ces deux projets pour leur qualité, leur niveau d'aboutissement. Ils ont été des modèles pour nous et nous ont inspiré pour le développement de SPACE. Ce n'est qu'à la lumière de leur étude complète que nous pourrons énoncer des recommandations pour la création d'outils de travail coopératif synchrone.

2.2.1 MAJIC

Rappelons que MAJIC [Okada94] se propose de permettre à quelques personnes de tenir une réunion à distance. La simulation se veut le reflet le plus fidèle possible de la réalité, donnant au final un outil luxueux. La sensation de présence des personnes distantes, due à la convergence de multiples détails, est restituée avec la plus grande qualité possible.

La meilleure façon de tenir une réunion semble être d'asseoir les participants autour d'une table pour qu'ils puissent se voir et discuter d'un sujet particulier, sans que les distances entre eux soient trop grandes. Cette dernière condition est souhaitable car une bonne partie de la communication n'est pas verbale mais passe par une multitude d'indices corporels. Fixer quelqu'un, parler emphatiquement avec de grands gestes, détourner ou baisser les yeux sont des manifestations de l'état d'esprit, des intentions d'une personne. L'endroit où se pose le regard importe également, par exemple pour inviter quelqu'un à prendre la parole. Les postures du corps sont aussi le signe de l'intérêt, du désaccord, etc. ; elles sont souvent adoptées inconsciemment.

Quand une réunion directe en face-à-face, en chair et en os, n'est pas possible il faut utiliser d'autres moyens et médiatiser la communication. Dans les systèmes utilisant la vidéo, on peut faire la distinction entre ceux de desktop conferencing, et d'autres visant à simuler le plus exactement possible les conditions réelles. Les premiers ont pour avantage la liberté qu'ils procurent, chacun ayant la possibilité de les utiliser à tout moment à partir d'une station quelconque équipée en multimédia. Les autres, plus encombrants, semblent susceptibles de permettre des tables rondes, des réunions où les interactions entre personnes sont libres et naturelles, d'intégrer espace de communication et espace de travail et d'autoriser le contact visuel. Celui-ci est difficile à assurer quand la réunion implique plus de deux participants. Le fait de représenter les personnes distantes en taille réelle augmente immédiatement la sensation de présence au point, par exemple, d'être tenté de montrer un objet réel à l'image en question, la personne distante n'ayant pourtant pas connaissance de notre environnement [Buxton92], mais le contact visuel est encore plus difficile à garder.

Ces constatations ont conduit au cahier des charges suivant :

- MAJIC doit permettre les réunions de type table ronde en assurant le contact visuel entre les participants ;
- chacun doit pouvoir se rendre compte de ce que les autres regardent, à qui ils s'adressent (conscience du regard) ;
- les personnes distantes doivent être visibles en taille réelle, dans un décor commun et sans limites entre eux ;
- au centre, entre les participants, doit se trouver un espace de travail partagé, commun.

La taille réelle permet de lire les expressions du visage et les gestes, qu'un petit écran n'est pas capable de restituer correctement (champ de vision et résolution limités). La plupart des systèmes fournit un ensemble de vues, une par interlocuteur distant. Il semble préférable d'extraire seulement les personnes de ces vues pour les rassembler dans un décor commun, renforçant la sensation que tout le monde est dans le même lieu. Notons que les utilisateurs du prototype utilisé chez Sun [Tang92] appréciaient pourtant grandement cet arrière-plan qui leur permettait d'avoir une idée de ce qui pouvait se passer « à l'autre bout », en particulier de déceler ce qui pouvait distraire leur collègue. MAJIC visant plutôt à simuler une salle fermée, protégée des interruptions extérieures, l'environnement réel des participants importe peu. La conscience du regard, et le contact visuel qui en est un cas particulier, évitent de désigner quelqu'un explicitement quand on s'adresse à lui ou qu'on fait référence à lui. Elle permet d'attirer l'attention d'une personne, simplifiant et rendant plus naturels les interactions entre les interlocuteurs. Les auteurs de MAJIC considèrent que l'espace de travail a un rôle triple : accueillir les documents privés de chacun, recevoir les documents mis en commun et, point important au vu des conventions sociales japonaises, éviter aux gens de se regarder sans cesse en fournissant simplement quelque chose où fixer son regard, juste pour détourner les yeux. La distance entre les participants est aussi chargée de sens : trop faible, elle suggère des liens personnels, voire intimes ; trop grande elle empêche de s'impliquer dans une activité d'échanges. Entre les deux se trouve la « distance sociale » qui est adoptée au travail, lors de réunions en particulier. La « distance virtuelle » est celle que ressent une personne face à l'image de son collègue, mesurant également si cette personne a l'impression d'être face à quelqu'un ou à une vidéo artificielle. Elle est fonction de détails comme la taille, la qualité de l'image, les parasites, le timbre de la voix retransmise. La « distance sociale virtuelle » est celle qui est la plus adaptée à une réunion et c'est donc celle que chaque utilisateur doit ressentir.

L'installation de MAJIC est composée, au centre, d'une station de travail, derrière laquelle est placé un écran en arc de cercle, de quatre pieds de haut sur huit de large. Cet écran, créé spécialement pour la circonstance, a la particularité d'être réfléchissant pour l'utilisateur placé devant lui, alors qu'il est partiellement transparent (40%) vu de l'autre côté. Une multitude d'hexagones, blancs devant et noirs derrière, sont collés sur une surface transparente. Le côté blanc permet de s'en servir comme écran de projection. De derrière, la lumière passe entre les petits hexagones, on peut donc filmer à travers l'écran. L'idée est de projeter l'image d'une personne distante ; derrière cette image, une

caméra filme l'utilisateur à travers l'écran (Figure 57). C'est de cette manière que le contact visuel s'effectue, de manière semblable aux Hydra.

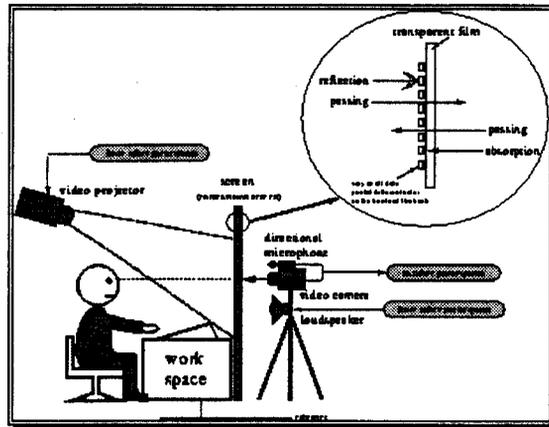


Figure 57 : le matériel de MAJIC

Le contact visuel multiple et la direction du regard sont obtenus en multipliant ce dispositif. L'important est de garder, à chaque poste MAJIC, le même ordre d'apparition des participants. Dans la Figure 58, l'utilisateur A est toujours à la gauche de B et à la droite de C. Cette cohérence spatiale assure que la direction du regard peut être interprétée correctement.

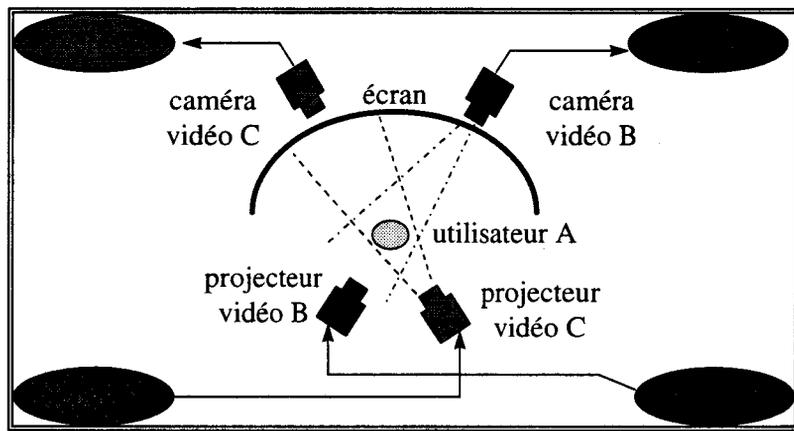


Figure 58 : dispositif MAJIC complet avec plusieurs interlocuteurs

Un haut-parleur est placé près de chaque caméra, restituant la voix de la personne projetée à l'endroit où celle-ci apparaît. Ladite voix est captée par des micros directionnels, un sur chaque caméra. Les concepteurs ont préféré utiliser plusieurs micros plutôt qu'un seul pour que la personne à laquelle l'utilisateur s'adresse reçoive sa voix clairement, tandis qu'elle parvient aux autres moins forte en raison du champ étroit du micro. Cette méthode simule donc de manière assez réaliste la direction de parole.

La station de travail encastrée dans le bureau central constitue l'espace de travail. L'article n'évoque pas clairement la possibilité de partager des documents, cet espace devant donc être

considéré comme strictement personnel. La séparation nette entre espace de travail et espace de communication est l'occasion pour l'utilisateur de quitter ses collègues des yeux momentanément, les conventions voulant que l'on évite de fixer longuement un interlocuteur. Cette séparation est conforme à la réalité, où il faut choisir entre regarder ses interlocuteurs ou travailler sur le document. L'important est de pouvoir passer d'un espace à l'autre d'une manière naturelle et immédiate [Buxton92]

Afin de valider MAJIC (Figure 13), des essais ont été faits lors d'un salon sur la collaboration. Une personne du public était invitée à converser avec deux démonstrateurs. Le contact visuel n'était que rarement établi, montrant qu'il était vivement perçu, ce qu'on confirmé les interviews a posteriori des visiteurs. Les images à taille humaine ont rempli leur rôle : elles donnent l'impression de présence proche. La conversation s'installait immédiatement avec les démonstrateurs. Un zoom sur ceux-ci mettait les nouveaux venus mal à l'aise, à l'exception d'amis des démonstrateurs qui devaient retrouver là la « distance personnelle » à laquelle ils étaient habitués. La taille du prototype a été critiquée.

Les faiblesses du prototype ont été identifiées. MAJIC est limité à un seul participant par site connecté. L'utilisateur doit rester à sa place. Ces deux problèmes pourraient être résolus par la mise en place de caméras motorisées capable de suivre une personne. Seulement quatre participants au maximum peuvent se réunir. Au-delà, dans la réalité, la distance entre les personnes augmente fatalement. Dans MAJIC, la représentation de nombreuses personnes, en petit, est envisageable, la difficulté principale est de garder la conscience du regard dans ces conditions. L'écran offre une image de qualité imparfaite, intrinsèque à sa conception qui lui confère la possibilité de réfléchir d'un côté et d'être transparent de l'autre. La structure peut certainement être améliorée mais les visiteurs du salon n'ont pas vraiment critiqué ce défaut. Une version pliable de l'écran limiterait les problèmes de place prise par le dispositif (il reste à expliquer où vont passer caméras et projecteurs !). Enfin, le décor commun pourrait être amélioré sans problème technique particulier.

L'équipe qui a monté le projet MAJIC s'est plus tard attachée à développer un outil de desktop conferencing, appelé fort logiquement Desktop-MAJIC [Tanaka96]. Les contraintes matérielles limitant sérieusement les possibilités, c'est bien un nouveau produit qui a été conçu, et non pas une copie de MAJIC, bien que certaines idées se retrouvent évidemment dans Desktop-MAJIC. Le cahier des charges indiquait cinq conditions à remplir :

- rendre la direction du regard ;
- rendre les actions et les gestes ;
- rendre les mouvements de la bouche et transmettre le son ;
- fournir un espace de travail partagé ;
- réduire le trafic réseau.

La vidéo utilisée dans d'autres systèmes de visioconférence ne permet pas de fournir d'informations sur la personne ou le document qu'un utilisateur regarde. Neuf images par personne

vont être utilisées pour remplacer la vidéo (Figure 59). Ces images montrent la personne en train de regarder devant elle, en haut, en bas, à gauche, à droite et dans les diagonales. Étant fixes, elles peuvent être transmises une fois pour toutes avant le début de la réunion. Un postulat important a été fait : le pointeur de la souris représente le centre d'intérêt d'une personne. Dans Desktop-MAJIC, l'objet regardé est assimilé à la fenêtre sur laquelle se trouve le pointeur. En fonction des positions de la fenêtre contenant le document regardé et celle où s'affiche l'image de la personne, un choix est fait entre les neuf images disponibles. Tout mouvement de la souris est interprété comme un changement de regard. Remarquons que rien n'oblige tous les utilisateurs à adopter la même disposition de fenêtres, l'image affichée étant choisie localement selon la configuration adoptée par chaque utilisateur sur sa machine. Cette indépendance est nécessaire à partir du moment où on ne se voit pas soi-même dans son interface, laquelle est donc forcément différente des autres ! L'article ne précise pas explicitement si c'est le cas dans Desktop-MAJIC.

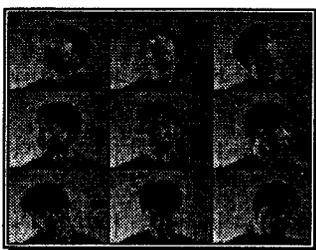


Figure 59 : les neuf positions dans Desktop-MAJIC

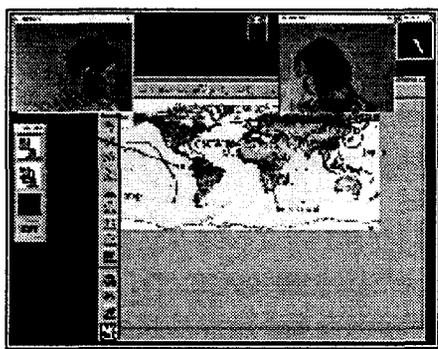


Figure 60 : Desktop-MAJIC avec photos et télépointeurs

Les actions des uns et des autres sont montrées par des pointeurs (un par participant) qui se déplacent sur le document partagé. Le lien entre personne et pointeur est perçu visuellement grâce à une ligne les reliant (Figure 60). Le principe est identique aux télépointeurs de Teledesign [Shu94]. Ces pointeurs en forme de main peuvent prendre trois postures différentes pour pointer, écrire et prendre. C'est l'utilisateur qui change explicitement la forme de son pointeur selon l'information qu'il veut faire passer quant à son activité.

Les canaux sonores restent ouverts en permanence. Pour montrer qui parle, un deuxième jeu de neuf photographies est nécessaire. Pour chaque position de la tête, une image a la bouche fermée

tandis que sur l'autre la bouche de la personne est ouverte (Figure 61). Quand quelqu'un se met à parler, les deux images sont affichées en alternance. Cette astuce simple montre clairement qui est en train de parler et facilite l'identification.



Figure 61 : bouche fermée et ouverte

Le partage d'applications est indépendant de Desktop-MAJIC et n'importe quel outil le réalisant peut être utilisé, voire plusieurs en même temps. Desktop-MAJIC reconnaît ces outils sur lesquels il vient se greffer en identifiant les fenêtres par l'intermédiaire du serveur graphique.

Une évaluation a été faite par rapport à un outil de visioconférence conventionnel, Apple Quicktime Conferencing. Les questionnaires ont montré l'importance du regard et de l'identification des actions par les formes du curseur. Le trafic réseau a été réduit soixante fois avec Desktop-MAJIC par rapport à Quicktime Conferencing.

La synthèse d'informations pertinentes sur les participants est de toute évidence une voie prometteuse qui n'a pas fini d'être explorée. Par contre, on peut reprocher à Desktop-MAJIC d'éliminer tous les indices sociaux, inconscients, involontaires que MAJIC s'efforçait de préserver, et de ne montrer que des actions explicites. L'idéal serait bien entendu un hybride entre MAJIC, pour son côté humain, et Desktop-MAJIC pour son efficacité et l'aide active qu'il apporte pour le travail collaboratif.

2.2.2 MASSIVE

Nous avons déjà évoqué cet environnement quand nous avons parlé des avatars et des clones, puis comme outil collaboratif utilisant la Réalité Virtuelle. La littérature à son sujet, ainsi que sur DIVE, étant abondante, nous avons décidé de consacrer une partie de ce manuscrit pour faire une synthèse des différents articles parus.

MASSIVE est en fait l'application de ce qui a été appelé un « modèle spatial » d'interaction [Benford93a]. Ce modèle s'appuie sur le concept de pièce. Une pièce est définie comme un espace informatique virtuel. C'est une notion très abstraite. La pièce va contenir des *objets* (personnes, données, etc.). Des portes de téléportations permettent de passer d'une pièce (qui peut être un monde entier) à une autre. Une pièce est un domaine d'interaction, entre des personnes ou entre ces personnes et les outils qu'elle contient. Ces interactions se font via un *medium* qui peut être un moyen de communiquer (sonore, visuel, texte) ou une interface très spécifique à un objet particulier. Elle peut être dédiée à un usage particulier (créée pour un usage particulier, temporaire), permanente (un amphithéâtre, une salle de conférence ou de projection), ou être juste un espace public libre (tendance pause café ou lieu de passage vers d'autres pièces). Cette définition est compatible avec celle des *rooms* [Henderson86] de Xerox. Ce concept n'est pas lié à une représentation particulière et peut avoir

un nombre quelconque de dimensions [Benford93c], même si on pense immédiatement à un espace 3D.

Cette métaphore visuelle permet de structurer assez naturellement l'espace virtuel. La pièce constitue un milieu dans lequel vont évoluer ses occupants et par lequel ils vont pouvoir interagir. Les allées et venues de chacun entre les pièces renforcent la sensation de présence. Une personne est repérée par sa position et son orientation. Là encore, ces informations n'impliquent pas nécessairement une réalité géométrique mais peuvent représenter des notions plus abstraites. Une condition nécessaire pour appliquer le modèle est de pouvoir évaluer des distances dans l'espace virtuel.

Dans le cas d'environnements très peuplés, les interactions se multiplient, les informations reçues sont excessivement nombreuses. Des mécanismes dits de médiation doivent être mis en place pour éviter une surcharge d'information. La plupart des systèmes (superviseur de groupe, tours de parole, workflow...) ont l'inconvénient de rendre explicite ce qui n'est qu'un ensemble de conventions sociales et sont donc perçus comme très contraignants. Remarquons qu'on retrouve des problèmes similaires quand la technique inhibe les interactions entre participants (par exemple à cause de délais de transmission [Tang92], voir 1.1.1), ces conventions sociales ne pouvant plus s'appliquer. Le modèle spatial propose un nouveau mécanisme de médiation qui réduit la surcharge d'information et qui de plus laisse aux individus leur liberté d'action, met sur un même pied d'égalité le locuteur et l'auditeur, évite de préétablir des contraintes fortes et permet au moins d'évoluer dans un contexte social.

Le premier élément de ce modèle est l'*aura*. Chaque objet possède une aura qui lui est attachée et qui se déplace avec lui. Elle constitue un sous-espace traduisant la présence de l'objet. Elle est spécifique à un médium particulier, et un objet peut en posséder autant qu'il existe de médias. La forme de l'aura, sa taille peuvent être quelconques. Elle n'est pas nécessairement autour de l'objet auquel elle appartient. Ces auras sont le premier moyen grâce auquel le nombre d'interactions est limité. Elles permettent de définir une partition de l'espace. Dans le modèle spatial, deux objets ne peuvent entrer en interaction (dans un médium précis) que si leurs auras se recoupent. On peut alors imaginer un monde très étendu et très peuplé, le nombre d'interactions effectives étant en fait très limité grâce aux auras. Si mon aura visuelle est plus grande que mon aura sonore, alors je serai vu avant de pouvoir être entendu.

Une fois que la possibilité d'interaction entre deux objets est déterminée, celle-ci va dépendre de la *conscience* que chacun a de l'autre. Afin de la quantifier, on fait appel au *focus* et au *nimbus* de chacun. Le focus est un sous-espace définissant une zone sur laquelle est portée l'attention d'une personne. Le nimbus est, encore, un sous-espace, dans lequel la personne projette sa présence. Il est également possible d'avoir un nimbus ou un focus occupant tout l'espace, mais dont la valeur n'est pas constante. Deux règles régissent le calcul du niveau de conscience :

- plus un objet est dans mon focus, plus je suis conscient de lui ;
- plus un objet est dans mon nimbus, plus il est conscient de moi.

Le niveau de conscience qu'un objet A a d'un objet B dans un médium M est une fonction du focus de A sur B dans M et du nimbus de B sur A dans M. Ces champs peuvent être manipulés explicitement, par exemple pour concentrer son attention sur un objet particulier (on « se met des lunettes » dans un médium particulier pour ne pas être distrait par l'environnement) ou au contraire pour chercher, explorer dans une large région. Le focus et le nimbus ont des propriétés qui donnent autant de pouvoir à la personne qui « émet » qu'à celle qui reçoit. Chaque personne contrôle sa conscience en s'approchant ou en s'éloignant et en se tournant. Cette conscience n'est pas symétrique (je peux voir une personne, i.e. en être conscient, sans qu'elle me voie nécessairement). Je ne suis pas obligé d'écouter quelqu'un qui me parle. Par contre, même si je désire ne pas écouter, je ne peux pas vous empêcher de parler.

La fonction à appliquer pour calculer la conscience en fonction des valeurs de focus et de nimbus dépend de l'application. En considérant que ces deux champs sont compris entre 0 et 1, on peut additionner les deux valeurs pour obtenir une conscience entre 0 et 2 [Benford93a]. À 0, c'est la conscience périphérique : je sais juste que la personne est là et où elle se trouve. À 1, la demi-conscience, je suis informé quand elle cherche à me parler. À 2, j'ai une conscience totale d'elle et j'entends ce qu'elle me dit. Si on multiplie ces deux valeurs [Greenhalgh94], l'observateur et l'observé influent également sur le niveau de conscience (entre 0 et 1). Chacun peut décider de forcer la conscience à être nulle, par contre il n'est pas possible de l'augmenter au-delà de ce que l'autre souhaite. La valeur de la conscience peut être utilisée comme une valeur continue pour le contrôle d'un paramètre (niveau sonore) ou interprétée par rapport à des seuils au-delà desquels certains comportements (envoi de messages...) seront déclenchés.

Pour que ce modèle soit utilisable en pratique, il est préférable de cacher les auras, focus et nimbus, voire de cacher leur existence même. Il vaut mieux suggérer ce qu'ils sont censés représenter par des mécanismes naturels. Par exemple les yeux d'une personne bougeront dans la direction où elle regarde, qui correspond à son focus. Leur manipulation doit se faire par des actions spatiales naturelles. Implicitement, l'aura, le focus et le nimbus sont modifiés quand on se déplace ou que l'on se tourne. Explicitement, l'intérêt peut être concentré sur une zone étroite en changeant la forme du focus. Enfin des outils « *adaptateurs* » auront pour fonction de modifier ces champs. Un porte-voix étendra mon nimbus, de nombreuses personnes pourront alors m'entendre et m'écouter.

La notion de *limite* a été introduite assez tardivement. Elle vient s'ajouter aux possibilités de modification des auras, focus et nimbus. Ces limites définissent des zones de l'espace. Tout comme les adaptateurs, elles vont changer les propriétés d'interaction. Une porte peut bloquer le passage à une personne et l'empêcher de percevoir ce qui se trouve au-delà si elle ne possède pas la clé. Une fenêtre ne peut pas être traversée mais ne bloque pas l'aura, le focus et le nimbus. On peut délimiter un territoire, contrôler le mouvement, modifier les propriétés de l'espace avec ces limites.

Ces avatars minimalistes sont des « points de repère » pour l'organisation spatiale. Dans DIVE, plate-forme d'expérimentation au champ d'application plus large que MASSIVE, on trouve des corps plus luxueux, certains étant même anthropomorphes. La différence ne s'arrête pas là. Les actions de chacun sont matérialisées par des couleurs et, surtout, une ligne est dessinée qui part du clone pour arriver à son point d'action [Benford95]. Cette aide visuelle, bien utile pour la compréhension de l'activité, est similaire à celles de Teledesign et de Desktop-MAJIC. Il est étonnant de ne pas la trouver dans MASSIVE ! L'explication vient probablement du fait que MASSIVE se concentre sur la communication plutôt que sur la production et le partage de documents. Nous verrons un peu plus loin que les utilisateurs ont quand même parfois du mal à comprendre l'activité, à savoir ce que font leurs collègues, ce qu'ils regardent.

L'observation de réunions tenues via MASSIVE tend à valider les choix de conception qui ont été faits. Deux sessions ont principalement été étudiées [Greenhalgh95]. La première rassemblait six personnes dotées de stations graphiques, sonores et texte sur réseau local. La deuxième, à cinq personnes, était distribuée sur plusieurs sites et accueillait à la fois des utilisateurs texte+audio et graphiques+texte.

Les résultats purement techniques sont convaincants. L'installation de MASSIVE sur chaque site, la puissance suffisante des machines, le réseau qui ne s'est étranglé que rarement ont contribué à la bonne impression que tous les utilisateurs ont ressentie. Toutefois il y a eu des problèmes et des effets assez inattendus.

L'angle de vue limité de l'image sur écran ou, pire, dans les casques d'immersion était insuffisant pour que les utilisateurs aient la sensation de vision périphérique. Des tâches simples, comme se mettre en cercle, en étaient rendues très difficiles. L'élargissement de l'angle de vue déforme la perspective. La solution appliquée a été d'offrir la possibilité de zoomer ainsi que plusieurs vues différentes : subjective, de derrière, latérales, du ciel... Il peut être intéressant d'automatiser les changements de point de vue dans certaines situations (vue du ciel quand on s'assied à une table par exemple).

Les vues de haut, à l'instar des *WIM* (Figure 56), ont l'intérêt d'offrir une vue *globale* sur l'environnement. Sans elle, le repérage, la compréhension générale du monde, la navigation sont très difficile dans les environnements de Réalité Virtuelle. Dans MASSIVE, la navigation posait effectivement problème, surtout pour les débutants. Il était courant de repasser la porte de téléportation par laquelle on venait d'arriver. La précision et la rapidité d'exécution des déplacements et des « mouvements de tête » (i.e. du corps entier pour les blockies) faussaient les interactions entre personnes, résultant en une impossibilité de coordonner l'activité, les prises de parole par les attitudes corporelles. Un problème semblable avait été expérimenté avec des outils vidéo générant un retard dans la transmission [Tang92]. Des périphériques mieux adaptés, éventuellement faits sur mesure, restent à trouver.

Les utilisateurs n'étaient pas sûrs d'être toujours bien entendus. Ces doutes ne provenaient pas uniquement des problèmes épisodiques de transmission du son. Le manque de feed-back des autres, l'incertitude quant à ce qu'ils regardaient, le manque d'indices visuels réduisent les interactions entre les participants. Pour aider un peu, une bouche apparaissait sur les avatars dont le propriétaire se mettait à parler. Il est probable qu'un clonage plus fin des personnes réelles serait bénéfique sur la qualité des interactions, mais MASSIVE y perdrait en simplicité, en efficacité réseau...

Il est arrivé que des utilisateurs se soient momentanément désintéressés de MASSIVE, laissant leur avatar à l'abandon. Contrairement à un outil vidéo, il n'était pas ici possible de savoir si la personne distante était effectivement à l'écoute ou si elle s'était absentée. S'il est aisé de représenter graphiquement si une personne est présente ou non, le passage d'un état à un autre doit être fait explicitement à moins d'inventer un système de détection... Cela implique une certaine discipline de la part des utilisateurs. Le mieux est de pouvoir « titiller » un avatar pour guetter sa réaction. Le problème s'était déjà posé aux premiers temps de DIVE : en cas de plantage du programme client, des avatars « morts » étaient abandonnés, sans que rien dans leur aspect ne signale leur vacuité. L'astuce était d'attraper un avatar pour le secouer et voir s'il réagissait. Les « cadavres » avérés étaient enfoncés dans le sol... Citons aussi l'outil de communication ICQ qui signale l'absence momentanée de son utilisateur quand l'économiseur d'écran se déclenche. Là aussi c'est une astuce simple mais qui ne manque pas de bon sens.

La notion d'espace personnel était étrangère aux utilisateurs d'interfaces en mode texte. Ils se plaçaient juste devant leur interlocuteur, quand ils ne passaient pas à travers. Ceux qui disposaient de machines graphiques, par contre, se tenaient à distance. Deux explications possibles sont données. Ils cherchaient à éviter l'obstruction de leur champ visuel, déjà limité, par un objet trop proche. Ou alors ils adoptaient les conventions sociales habituelles, la représentation graphique étant suffisamment éloquente pour les y inciter.

Pour terminer, citons une étude menée sur MASSIVE par des psychologues et des sociologues [Bowers96] dont l'originalité est d'avoir appliqué des techniques d'analyse de la conversation à celle des interactions entre les participants. Nous ne reprendrons ici que leurs conclusions. Les avatars observés ont profité des améliorations proposées précédemment : ils peuvent sommeiller quand leur maître s'absente, bouger les oreilles indépendamment, et ils sont dotés d'une bouche qui s'ouvre quand le propriétaire parle (détection à seuil). Un point intéressant est que les utilisateurs essaient d'employer les possibilités très limitées de démonstration des blockies pour retrouver des méthodes d'interaction qui leur sont familières. La négociation de la prise de parole favorise ces comportements. Les participants s'échangent simultanément des coups d'œil (en faisant pivoter leur blockie tout entier). Ils réagissent en fonction de ce que dit le locuteur, de ses silences. Dans l'ensemble, les comportements observés sont clairement similaires à des comportements naturels et habituels. Ce genre d'observation avait déjà été faite pour le téléphone, laissant à penser que quelque soit l'outil, les utilisateurs l'amèneront à reproduire des schémas sociaux déjà connus. Trois recommandations importantes sont formulées. Premièrement, pour que tout se passe bien, les auteurs proposent d'introduire dans le système une détection des pannes et problèmes techniques pouvant survenir, afin

d'informer les personnes concernées et éventuellement de réparer. Deuxièmement, les mondes virtuels devraient être construits autour des moyens d'interaction entre utilisateurs qu'ils veulent offrir. La navigation, la présentation d'informations sont moins importants qu'une table de réunion qui, même représentée très simplement, suggère l'existence du groupe en tant que tel, permet d'organiser la discussion, de placer les gens, de faire des tours de table pour que tous puissent s'exprimer. Troisièmement, il ne faut pas oublier la complexité du corps humain et des interactions qu'il peut engendrer. Un blockie ne permet pas de regarder derrière soi sans tourner le dos à son interlocuteur, c'est tout son corps qui bouge. Articuler le corps (le cou, dans l'exemple précédent) pourrait améliorer la situation. Des actions de haut niveau, à forte signification sociale (s'approcher, jeter un œil, se mettre en cercle) pourraient être prévues, cette voie reste à explorer. Le plus important semble être d'améliorer et de raffiner les avatars uniquement si cela s'avère nécessaire et sur un point précis à chaque fois. Il est inutile, comme la plupart des systèmes de Réalité Virtuelle cherchent à le faire, de tout simuler *a priori*, de montrer un corps humanoïde complet, de capter le moindre mouvement. Les ressources que cela représente en développement sont bien mal employés car déconnectées des préoccupations sociales primordiales pour un environnement partagé. Notre approche personnelle a été d'étudier ce qui importait dans une activité collaborative et de construire notre interface à partir de ces observations. Ainsi, nous avons porté directement nos efforts sur des éléments que nous savions être importants dans le cadre du travail de groupe assisté par ordinateur.

2.3 Nos recommandations

Nous allons maintenant lister, sous formes de recommandations, les points que les travaux étudiés indiquent comme étant importants à respecter. Le travail coopératif synchrone fait apparaître des contraintes dues à la nature des interactions homme-homme sur l'organisation de la scène et ce qu'elle doit contenir. Les interfaces évoluées permettent le travail de groupe, utilisent la 3D et mettent à profit les capacités de perception et de compréhension par l'observation de l'être humain. La Réalité Virtuelle présente de nombreuses techniques intéressantes, notamment pour représenter des personnes et les intégrer dans un monde spécialisé.

2.3.1 Travail coopératif synchrone

Au-delà de la technique, le travail coopératif impose de respecter des contraintes humaines sous peine d'échec. Elles découlent des capacités de perception, d'analyse et de compréhension de l'homme, en particulier lors des échanges langagiers. L'un des défis qu'impose la médiatisation d'une activité synchrone consiste à détecter les indices significatifs et, après transmission, à les rendre de manière utilisable. Les travaux déjà réalisés dans ce domaine montrent qu'il faut :

- respecter le *contact visuel* et permettre la *conscience du regard*. D'une manière ou d'une autre, il faut montrer ce qu'une personne est en train de regarder. L'idéal est évidemment que la

personne que l'on voit regarde effectivement l'objet concerné, plutôt que de passer par des artifices comme des couleurs, des flèches, etc. ;

- rendre les *gestes* :
 - ◊ pour la *désignation* : quand quelqu'un veut diriger l'attention des autres sur un objet particulier, il le montre par gestes ;
 - ◊ pour leur *expressivité* : ils aident, pendant la discussion, à traduire l'état d'esprit du locuteur, à donner de l'emphase à son discours...
- rendre les *expressions du visage* : tout comme les gestes, ils aident à mieux comprendre ce que pense l'interlocuteur ;
- respecter le principe de *continuité*. D'après les recommandations précédentes, il apparaît clairement que personnes et objets doivent être placés dans un même espace si l'on veut pouvoir montrer les interactions entre personnes et documents, ou entre les personnes uniquement ;
- assurer une *grande qualité du canal sonore*. Les voix ne doivent pas être distordues, incompréhensibles. Les « trous » doivent être évités : celui qui écoute ne comprend rien et celui qui a parlé ne comprend pas pourquoi l'autre n'a rien compris ! Enfin le délai de transmission doit être minimal. Les interactions sont gravement endommagées quand le décalage dépasse la demi-seconde ;
- *positionner l'image et le son* (la voix) au même endroit. La distinction spatiale permet de déterminer plus facilement qui parle. La « spatialisation » du son apporte une aide réelle à l'utilisateur.

2.3.2 Interfaces

Nous pouvons tirer d'autres enseignements des recherches menées sur les nouvelles interfaces. En particulier, le passage à la 3D, et son application pour le TCAO synchrone qui nous préoccupe, doit se faire moyennant quelques précautions.

- La *perception directe* doit être favorisée pour réduire la charge cognitive de l'utilisateur. Il faut autant que possible que l'interface soit explicite plutôt que forcer l'utilisateur à recréer mentalement les liens qui existent entre différents objets. L'animation temps réel, que nous prônons, peut être un moyen de montrer un événement de manière continue, sans rupture. Comme contre-exemple, on trouve des produits de TCAO qui, pour partager des applications, attribuent à chacun une couleur. Plusieurs pointeurs colorés sont visibles. C'est l'utilisateur qui doit recréer le lien entre pointeur et son propriétaire, en s'aidant de la couleur ;
- Il est possible d'*augmenter la densité d'information* à l'écran en utilisant un affichage graphique 3D. Cela suppose de fournir à l'utilisateur les outils (métaphores, widgets, périphériques d'entrée adaptés) pour naviguer et se repérer facilement et instantanément, sous peine de le

noyer sous une quantité de données qu'il ne peut appréhender dans son entièreté et de rendre l'interface inutilisable ;

- L'*animation interactive temps réel* est en fait la fusion de deux concepts distincts. Tout d'abord, l'animation temps réel est un moyen naturel pour montrer les changements survenant dans l'interface. Cela permet d'éviter les sauts d'un état à un autre, parfois incompréhensibles, en décomposant visuellement la transition. Un exemple est, lors de l'icônification d'une fenêtre, la réduction rapide de la fenêtre en l'icône lui correspondant. Ces préoccupations peuvent sembler simples mais ne sont hélas pas considérées de manière systématique. En conséquence, l'animation reste sous-exploitée. Pour s'en persuader, on peut prendre l'exemple des menus déroulants, dont l'apparition à l'écran ne s'est animée dans Windows que très récemment... Ensuite, en rendant cette animation interactive, on donne à l'utilisateur la possibilité d'explorer par lui-même des données complexes en gardant des points de repère, un contexte. La densité d'information à laquelle il est confronté n'est alors plus un obstacle ;
- Une activité de TCAO doit apparaître dans une *scène globale*. Il est crucial de représenter les interactions diverses entre composantes de l'activité. On peut en effet confondre l'activité et l'ensemble des interactions homme/homme et homme/documents. Nous avons déjà évoqué le principe de *continuité*, qui recommande d'intégrer du mieux possible les différents objets, notamment les personnes et les documents, dans un même espace. Nous recommandons d'effectuer cette intégration de manière à pouvoir montrer les liens entre les personnes, et entre chaque personne et les documents. La perception directe de l'activité en sera grandement favorisée. Le ClearBoard par exemple répond pleinement à cette contrainte, tandis que ClearFace reste, sur ce point, inachevé. Il pêche par manque de réalisme dans l'organisation spatiale des différents objets de l'activité. En faisant une comparaison imagée, le mélange entre personnes et documents ressemble à une émulsion où les ingrédients sont mélangés mais restent distincts, séparés. Desktop-MAJIC (voir 2.2.1) « globalise » les systèmes à fenêtres en montrant la direction du regard et en permettant la désignation, mais le fait au prix d'approximations très artificielles. Ce système effectue pourtant un premier pas prometteur vers la méthode « analyse/resynthèse de la scène », abordée également dans [Seligmann95] et que nous-mêmes avons explorée différemment ;
- La *connaissance du point de vue* des autres participants est nécessaire. Sans adopter forcément la rigueur du WYSIWIS strict pour tous (i.e. avoir la même vue pour tout le monde), il faut donner les moyens à l'utilisateur de percevoir le centre d'intérêt d'un de ses collègues éloignés, pour comprendre de quoi il parle, à quoi il fait référence et donner un sens aux expressions déictiques courantes dans le langage parlé ;
- Les *télépointeurs* semblent inévitables. Ils ont pour caractéristique d'être simples tout en étant explicites et informatifs. Ils apportent de la précision dans la simulation de la téléconférence, laquelle est difficile à obtenir dans un environnement synthétique. En plus de matérialiser le *point d'action* d'un utilisateur, ils peuvent apporter des renseignements sur le type de l'action effectuée ;

- *Tout doit être visible* (ou du moins perceptible) à tout moment. L'une des principales différences entre le travail individuel et en groupe, c'est que dans ce dernier un utilisateur n'est pas la cause de tous les événements qui ont lieu. Il devient alors primordial, pour pouvoir suivre dans de bonnes conditions le déroulement de l'activité, que chacun ait connaissance des diverses actions des uns et des autres ainsi que des changements pouvant avoir lieu. Dans le cas contraire, l'utilisateur risque de perdre le fil, pour finir par ne plus comprendre de quoi l'on parle, pourquoi certains choix ont été faits, etc. Une conséquence est qu'il faut éviter autant que possible les occlusions entre objets (par exemple le recouvrement de fenêtres 2D) qui masquent une partie de l'activité. Rappelons la gêne qu'éprouvaient les utilisateurs de MASSIVE à cause du champ visuel restreint. Dans [Seligmann95] par exemple, le son est utilisé pour rendre certains événements et actions. C'est une idée déjà ancienne dans les interfaces mono-utilisateurs [Gaver89] [Conversy97] ;
- L'interface doit être *immédiatement attirante*. Les critères esthétiques sont difficiles à énoncer, car en bonne partie subjectifs. Les produits interactifs 3D, animés, rencontrent toujours un vif succès. Les images de synthèse 3D ne sont pourtant plus une nouveauté, elles sont quotidiennes au cinéma, dans la publicité et l'attrait de la nouveauté ne suffit plus à expliquer le succès des images de synthèse. Les jeux notamment ont retrouvé un nouveau souffle avec la généralisation des moteurs 3D. Le réalisme (visuel) et la facilité d'accès (manipulations simplistes non réalistes, au clavier par exemple), la fluidité des animations, l'adéquation avec la perception humaine doivent être mis en avant.

2.3.3 Réalité Virtuelle

La Réalité Virtuelle a fait naître de nombreuses techniques, matérielles ou logicielles, pour représenter un monde réaliste. Nous en avons retenu quelques-unes (surtout logicielles, étant donné que nous voulons nous affranchir des périphériques complexes) dont nous pourrions profiter directement pour la réalisation d'outils de travail coopératif synchrone :

- les *abstractions*. Bien choisies, elles ne nuisent pas à la compréhension. Par contre, elles apportent des possibilités que le respect absolu de la réalité ne permet pas. Par exemple, l'épaisseur de l'icône d'un fichier sera proportionnelle à la taille de celui-ci [Staples93]. Ces abstractions doivent correspondre aux métaphores évoquées par l'interface ;
- ne pas tout reproduire, *simplifier et spécialiser*. Au-delà de l'impossibilité de retrouver à distance une situation réelle, on voit que, dans la mesure où un outil de travail coopératif synchrone constitue un cas particulier de la réalité, il est inutile et même inefficace qu'il offre un trop grand nombre de libertés. Au contraire, il sera plus efficace s'il optimise la communication et la production en groupe, éventuellement aux dépens d'une plus grande polyvalence ;
- se limiter à un *haut niveau de contrôle* des avatars, mais aussi des interactions. Les actions doivent être interprétées et « comprises » par le système pour leur attribuer un sens, une

intention, une finalité. Le volume des communications est faible (seul l'important est transmis), et le système peut représenter des événements complexes du mieux possible ;

- exploiter un *nombre réduit de schémas d'interaction*, et de manipulation en particulier. Par exemple, les widgets permettent de déclencher simplement une suite d'action complexes. De plus l'utilisation systématique des mêmes widgets permet à l'utilisateur de s'habituer rapidement à une interface ;
- les clones, par leur réalisme et leur versatilité, constituent une représentation de choix des participants à l'activité ;
- les accélérateurs 3D : ils sont maintenant largement disponibles, évolués et leur coût est devenu très raisonnable. Ils intègrent toutes les primitives classiques en synthèse d'images temps réel : éclairage, placage de textures, calculs matriciels, interpolations. C'est la porte ouverte à une nouvelle génération d'interfaces entièrement 3D.

Sur d'autres points, on peut critiquer la Réalité Virtuelle classique, en particulier quand il s'agit de l'appliquer au travail coopératif et au travail de bureau.

- l'*immersion* semble malvenue. Il est raisonnable de penser qu'une personne utilisant un outil de travail coopératif le fera de son bureau. Elle sera amenée à discuter avec ses collègues, à répondre au téléphone, à être dérangée pendant la réunion distante. Il est préférable de ne pas la couper du monde réel. De plus, le matériel est trop contraignant et il est illusoire de vouloir faire revêtir casques, gants et combinaisons intégrales aux utilisateurs visés ;
- les *périphériques d'entrée* sont complexes. Si la simulation de la réalité justifie de capter le maximum d'informations, cela complique l'interaction. A l'instar des interfaces 2D, qui utilisaient la souris, nous nous orientons plutôt vers des périphériques très simples en soi mais qui permettraient de tout faire dans notre interface ;
- il est impensable de faire du *retour d'effort tactilo-proprio-kinesthésique*. Les contraintes techniques sont encore beaucoup trop lourdes et coûteuses pour que les mécanismes actuels soient compatibles avec l'utilisation que nous voulons en faire.



3. Notre proposition

Après avoir retiré de notre étude bibliographique un ensemble de contraintes, d'idées fortes et de conseils, nous allons élaborer un modèle de présentation d'interface respectant les recommandations énoncées à la fin du chapitre précédent. C'est la partie interface du projet SPACE.

3.1 Le projet SPACE

Nous allons maintenant présenter nos travaux. Le projet SPACE vise à mettre au point un système de TCAO général, sur micro-ordinateur. Les résultats présentés dans cette thèse concernent principalement *Spln*, la partie visible de l'interface. Les autres éléments propres au projet SPACE font l'objet de deux autres thèses : le module ScenIC, qui interface *Spln* et le serveur d'activité ODESCA, sera décrit par Patrice Carlier, tandis que Cédric Dumas détaillera les interactions (homme/machine et homme/homme) au sein de *Spln* ainsi que l'intégration des clones.

Les principes que nous mettons en œuvre sont généraux et constituent un modèle d'interface. L'instanciation de ce modèle pour des applications particulières nécessite de le spécialiser. Selon le nombre de participants attendus, l'importance des critères sociaux (hiérarchie), le type d'activité (design participatif, négociation, réunion informelle, cours magistral...), le rôle de chacun, le modèle

général sera bridé pour répondre aux contraintes spécifiques. Le modèle doit être le plus large, le plus général et complet possible pour faire face à tous les besoins.

3.1.1 Architecture du projet SPACE

L'activité coopérative a ceci de particulier qu'elle existe potentiellement à tout moment. Les personnes susceptibles d'y participer, les outils dont elles auront besoin, leurs fonctions respectives sont connus à l'avance. Seule une entité centralisée permanente peut assurer la conservation de ces informations. L'interface en elle-même est temporaire. Elle permet à son utilisateur de percevoir l'activité, d'y participer et de faire progresser la tâche commune. Un autre élément doit fournir à l'interface les données centralisées sous une forme exploitable, faire la part entre les actions locales et celles ayant une incidence sur l'activité de groupe, se charger des communications et des mises à jour.

Trois entités distinctes ont été identifiées dans SPACE. La première est un serveur d'activité centralisé, ODESCA. La seconde, répliquée sur chaque machine cliente, est l'interface visible par l'utilisateur : c'est Spln. Entre les deux vient se placer ScenIC, dont là aussi une instance existe par machine.

ODESCA est une version remise à jour du serveur d'activité utilisé dans le projet Co-Learn [Viéville95]. Il centralise les informations sur l'activité (identités et rôles des participants, droits...) et permet la persistance de l'activité et des documents en fin de session. La gestion de l'activité est assurée par le module, appelé ScenIC, chargé de faire le lien entre chaque instance de Spln et le serveur d'activité ODESCA (Figure 63). ScenIC est chargé de fournir à Spln, la partie chargée de la présentation et de l'interaction, une description de l'activité (métaphore de la mise en scène) éventuellement adaptée à chaque utilisateur. Cette description est respectée par l'interface au mieux de ses possibilités, détachant ainsi le noyau fonctionnel collaboratif de la présentation. La description indique un type de décor, quels sont les objets visibles, quelles sont les contraintes spatiales (par exemple deux objets doivent rester côte à côte, tous les objets ont une place fixe...), quels sont les droits de l'utilisateur sur chaque objet (éditer, sauvegarder...), etc. Elle concerne aussi bien l'environnement sonore que visuel. Elle s'effectue par envois de messages et peut évoluer au fur et à mesure que l'activité se déroule. Spln construit la scène représentant l'activité sur laquelle l'utilisateur vient agir, ses actions étant répercutées via ScenIC et ODESCA vers les autres participants.

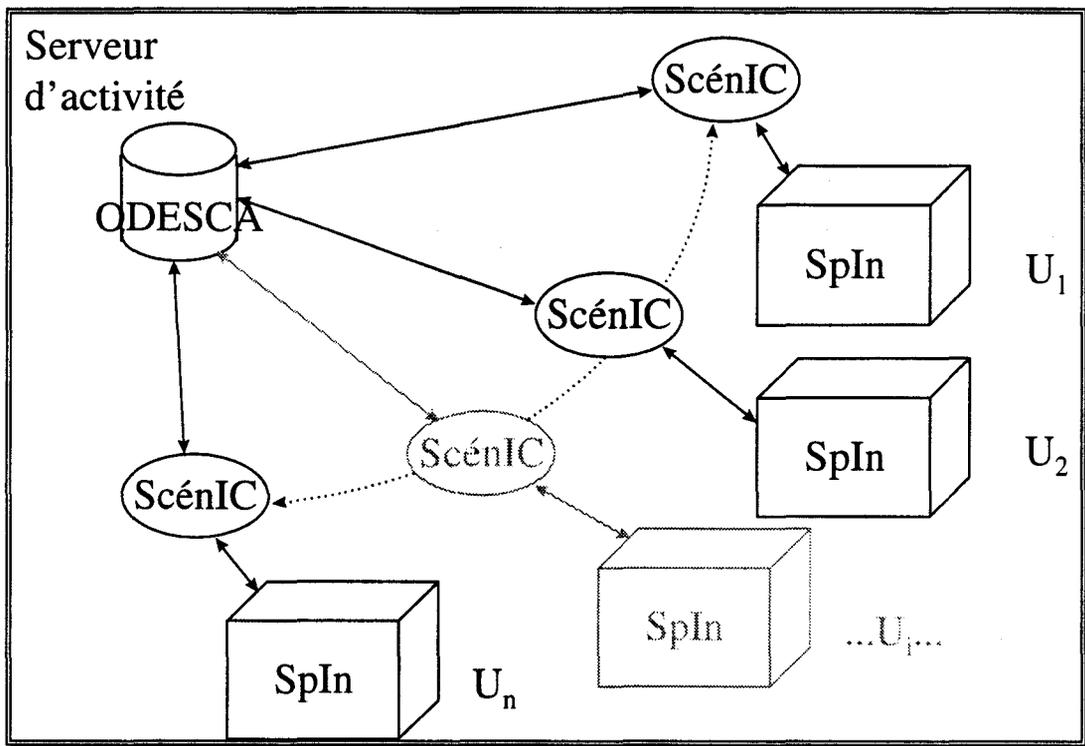


Figure 63 : l'architecture du projet SPACE

Les informations suivantes sur ODESCA et ScénIC sont adaptées d'un document de travail de nos collaborateurs du laboratoire Trigone [Hoogstoel97].

3.1.2 ODESCA : gestion de l'activité



Les rôles principaux d'ODESCA sont au nombre de quatre :

Connaissance des participants, des documents et des outils

Pour décrire complètement une tâche, il est nécessaire de décrire qui participe à cette tâche et sur quoi les participants vont travailler. ODESCA n'identifie pas des individus en tant que tels mais des rôles dans une tâche. Par exemple, dans une conférence, on pourrait définir 4 rôles : le conférencier, le maître de conférences, les panelistes et le public. Logiquement, l'individu qui est conférencier le reste durant toute l'activité. Cependant il existe des activités durant lesquelles les rôles passent d'un individu à l'autre. Par exemple, pour un vote, une personne peut avoir tout d'abord le rôle d'initiateur du vote, puis avoir le rôle de votant, puis avoir le rôle d'annonceur du résultat du vote.

Les documents sont identifiés par leurs noms et leurs adresses. Les outils sont identifiés par un type ; on parle d'outils de vote, d'outils de dessin, d'outils de « chat », et non pas l'outil Netmeeting ou Cooltalk.

Connaissance du découpage de la tâche

Chaque tâche est découpée en phases séquentielles. Une phase correspond à une étape dans la tâche. Chaque phase a une durée connue ou non ; ainsi, il est possible de changer de phase à la demande.

Connaissance des droits de chacun

Une phase regroupe des participants, des documents et des outils. Pour chaque phase, des règles décrivent les droits qu'ont les participants sur les documents et les outils. Par exemple, dans la phase « Exposé du conférencier », le public a le droit d'utiliser un outil de communication (le *chat*). Dans cette même phase, le maître de conférences n'a pas le droit à la parole, n'a pas accès au document du conférencier, etc. Ces droits reflètent le rôle d'un participant dans l'activité.

Maintenance de la persistance

Dans le cadre d'*activités asynchrones* : entre deux sessions, ODESCA garde en mémoire l'état de l'activité, c'est-à-dire l'état des outils et des documents (au sens large du terme) utilisés, mais aussi dans quelle phase est l'activité ainsi que l'état des participants dans cette phase. Par exemple, pour un vote sur une semaine, ODESCA sait qui a déjà voté et qui ne l'a pas fait ; il connaît également le vote de chacun.

Dans le cadre d'*activités synchrones* : ODESCA maintient en permanence la cohérence entre les postes de travail. Par exemple, la description de la tâche peut indiquer que lorsque le conférencier parle, les autres membres du groupe de travail ne peuvent pas prendre la parole. Étant donné qu'ODESCA connaît ce découpage, il sait quelle est la phase en cours et donc quels sont les droits associés aux participants. Ainsi, lorsque l'activité entre en phase « exposé du conférencier », ODESCA informe chaque poste de travail de modifier les droits des participants en fonction de cette phase.

En résumé, ODESCA connaît :

- la tâche et son découpage en phases ;
- les participants et leurs droits dans chaque phase ;
- les outils ;
- les documents ;
- la salle virtuelle²⁰ dans laquelle se déroule l'activité. Cette connaissance induit :
 - ◊ la connaissance du type d'outils qu'on y trouve ainsi que les rôles mis en jeu ;

²⁰ Il existe des types de salles (téléconférence, téléassistance, atelier de travail, forum, etc.).

- ◇ la nature de la tâche (réunion informelle, cours magistral, etc.) ;
- ◇ le type de coopération associé (synchrone/asynchrone).

ODESCA ne connaît pas :

- les contraintes appliquées sur la tâche, par exemple :
 - ◇ le caractère obligatoire ou facultatif d'un document ;
 - ◇ les positions respectives des participants ;
 - ◇ les types de média à utiliser en priorité ;
- les attributs de la salle virtuelle associés à la tâche tels que :
 - ◇ décor ;
 - ◇ résonance ;
 - ◇ espace fermé ou non...

3.1.3 ScenIC : représentation abstraite de la scène

ScenIC est l'interface entre Spln et ODESCA. Il contient une représentation abstraite de la scène et de la tâche qui s'y déroule. Par cette représentation abstraite, ScenIC est donc indépendant du rendu visuel exécuté par Spln ; il peut supporter différentes représentations visuelles pour une même activité. Par contre, ScenIC est fortement dépendant d'ODESCA puisque c'est de ce dernier que proviennent des informations essentielles pour l'activité.

Les rôles principaux de ScenIC sont au nombre de cinq.

Représentation abstraite de la scène et de la tâche associée.

ScenIC reprend les informations associées à la tâche fournies par ODESCA (phases, rôles, documents, outils, état de l'activité, salle virtuelle) et les transforme en arbre. Les feuilles de l'arbre sont tous les « intervenants » de la tâche (participants, documents, outils), et les nœuds sont des opérations sur ces feuilles ou sur des branches de l'arbre. Tout comme dans DIVE, on a un arbre par « média » différent (visuel, sonore).

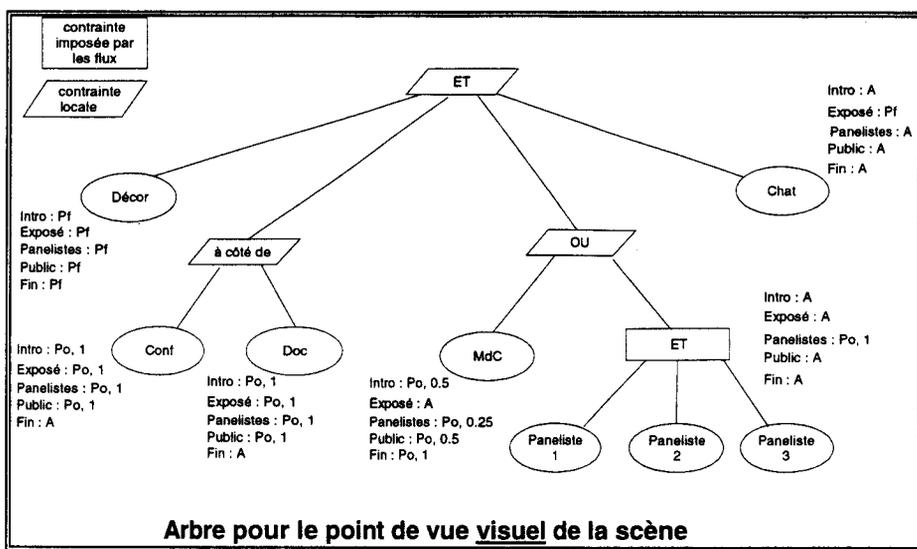


Figure 64 : un exemple d'arbre décrivant la scène

La Figure 64 contient un exemple d'arbre de description d'une scène. Tout d'abord, un décor est défini. Deux objets ayant une importance maximum, le conférencier et son document, sont ajoutés. Il est recommandé de les représenter proches l'un de l'autre. La branche suivante indique que le maître de conférence sera montré alternativement avec une vue d'ensemble des auditeurs. C'est le changement de phase qui déterminera le choix de la branche à montrer. Ce procédé est fréquent, par exemple, dans les émissions télévisées. Enfin, un outil de bavardage sera disponible.

D'une manière générale, chaque personne d'un groupe a un rôle qui lui est propre. Il peut dépendre de multiples paramètres : position hiérarchique, orateur ou simple spectateur... Ce rôle, et aussi ce qui en découle, est très dépendant du type d'activité à laquelle on participe.

A partir de ce rôle et du type d'activité des droits sont définis : accès à certains documents, possibilité de prise de parole, annotations sur un document, ouverture et fermeture d'une session de travail. Ils conditionnent également dans une certaine mesure le point de vue de l'observateur.

Une bonne définition a priori de ces droits, qui sont le reflet de la place d'une personne dans l'activité, permet de gérer efficacement des ressources matérielles (suppression de canaux sonores indésirables) et de restituer visuellement cette place en modifiant les outils disponibles (par exemple microphone, rétroprojecteur et marqueur dans une réunion) ou en adaptant le point de vue. Le professeur est libre de montrer les transparents qu'il a préparés à sa classe, de parler à tout le monde. Il n'a par contre pas besoin de distinguer précisément chaque élève. De leur côté, les élèves peuvent consulter les transparents précédents, poser une question.

La vue sur l'activité fournie dans l'interface par SpIn doit être le reflet du rôle de l'observateur. En lui indiquant sa place et ses droits, il lui est plus facile de se sentir concerné et de s'investir dans le travail du groupe. C'est ScenIC qui gère les droits, dynamiquement, selon une métaphore de « mise en scène ». Il avertit SpIn de toute modification de ceux-ci.

Stockage des contraintes spatio-temporelles.

ScenIC ajoute à la description fournie par ODESCA des contraintes spatio-temporelles destinées à fournir des indications pour la représentation visuelle. Ces contraintes sont stockées dans les nœuds de l'arbre constituant la représentation abstraite de la scène. À terme, les relations exprimées dans ces nœuds devront permettre une description suffisamment détaillée pour que Spln puisse en faire un rendu adéquat. Cependant l'objectif n'est pas de donner une description totale et précise de la scène. Seules des indications de positions, de simultanéité, d'exclusion, etc., seront utilisées. Voici quelques exemples de contraintes :

- contraintes spatiales :
 - ◊ le conférencier et son document doivent être côte à côte ;
 - ◊ dans la phase de questions, le conférencier doit être à côté du maître de conférences ;
 - ◊ les panelistes sont tous situés au même endroit et ne peuvent pas être déplacés.
- contraintes temporelles :
 - ◊ un seul paneliste est visible à la fois ;
 - ◊ on affiche soit les panelistes, soit le maître de conférences.

Point de passage obligatoire entre SpIn et ODESCA.

Ce point de passage concerne évidemment les échanges d'ODESCA vers Spln et de Spln vers ODESCA.

- ODESCA vers Spln : ce sont des informations liées à la tâche, par exemple un changement de phase, un changement de rôle, un changement de droit, etc. Ces informations sont filtrées par ScenIC qui d'une part en extrait les données pour modifier son arbre afin de maintenir la cohérence avec ODESCA, et d'autre part transforme ces informations pour les transmettre à Spln, en les surchargeant éventuellement d'informations vitales pour le rendu visuel et sonore ;
- Spln vers ODESCA : ce sont des informations liées à l'interaction, par exemple le fait de regarder vers un objet, ou de désigner un objet à l'écran. Si l'interaction est purement locale, ScenIC ne fait rien. Si l'interaction a des répercussions sur les autres postes, mais sans modifier l'activité, alors ScenIC avertit les autres ScenIC de cette modification. Si l'interaction doit modifier l'activité (par exemple un changement de phase), alors ScenIC fait remonter l'information à ODESCA qui retourne ensuite de nouvelles informations à ScenIC, qui peut à son tour les retransmettre à Spln (cf. cas ODESCA → Spln).

Stockage des propriétés des différents composants de la tâche.

Nous avons vu qu'ODESCA sait quels sont les participants, les outils et les documents intervenant dans une tâche. Pour gérer la persistance, il conserve également d'autres informations. Ces

informations sont utilisées par ScenIC et Spln, mais pas nécessairement par ODESCA. Ces informations décrivent précisément chaque intervenant. Par exemple :

- un participant est décrit par :
 - ◊ un identificateur qui est propre au système et qui n'est connu que de lui (donnée issue d'ODESCA) ;
 - ◊ un nom qui l'identifie auprès des autres participants (son nom réel ou un pseudonyme, donnée issue d'ODESCA) ;
 - ◊ des représentations visuelles et/ou sonores classées par type (vidéo ou photo ou dessin) associées soit à des flux soit à des adresses (URL, nom de fichier, ...) ;
 - ◊ des renseignements complémentaires (sa carte de visite, son URL, etc. - données issue d'ODESCA).

- un document est décrit par :
 - ◊ un nom de document (par exemple « rapport de stage de Patrick Martin », donnée issue d'ODESCA) ;
 - ◊ un type (texte, vidéo, image) ;
 - ◊ des représentations visuelles et/ou sonores classées par type (vidéo ou photo ou dessin) associées soit à des flux, soit à des adresses (URL, nom de fichier, ...).

- un outil est décrit par :
 - ◊ un type (dessin, retouche d'image, traitement de texte, PréAO²¹, etc. - donnée issue d'ODESCA) ;
 - ◊ une localisation (URL, disque dur, etc. - donnée issue d'ODESCA).

Gestion améliorée de la qualité de service.

Les informations vues au point précédent sont utilisées lors de l'activité d'une part pour le rendu visuel et d'autre part pour essayer de fournir la meilleure qualité de service (QOS). Pour chaque phase de la tâche, ScenIC sait quelles sont les « données » obligatoires, les « données » facultatives et donc disponibles et les « données » interdites. Par exemple, dans la phase « exposé du conférencier », ScenIC va indiquer à Spln que la représentation visuelle du conférencier est obligatoire, les renseignements complémentaires sur le conférencier sont facultatifs, la représentation visuelle du maître de conférences est interdite, etc. Comme nous le verrons plus loin, c'est Spln qui en fonction de ces indications va essayer de donner la meilleur qualité possible pour le rendu.

Le filtrage des informations par ScenIC a une incidence sur l'efficacité de l'interface. L'un des problèmes en travail coopératif est que, contrairement à une application mono-utilisateur, les documents utilisés sont nombreux et, de plus, d'autres personnes sont présentes et doivent apparaître

²¹ Présentation Assistée par Ordinateur, ou mise en page

à l'écran. Pour minimiser l'encombrement, ScenIC attribue une importance aux documents dont Spln va se servir pour réduire si besoin est la quantité d'information à l'écran (nombre ou niveau de détail) et ainsi éviter un encombrement face auquel l'utilisateur serait désarmé. D'un côté, ScenIC va choisir quelles sont les informations pertinentes à un moment donné (les documents « actifs ») tandis que Spln va organiser au mieux les objets visibles. Les personnes ne faisant qu'écouter, les transparents déjà présentés sont des exemples d'objets ayant un faible intérêt. Spln a donc carte blanche pour les représenter comme tels, en les groupant, en les mettant à l'arrière-plan etc. Si l'un des auditeurs passifs désire intervenir, il le signale à ScenIC en agissant dans Spln. Il est alors considéré comme actif et est mis en avant dans la présentation que Spln construit.

3.2 Nos choix essentiels

À partir de maintenant, nous allons nous focaliser sur Spln, la partie du projet SPACE dont nous avons la responsabilité. Les concepts généraux, nos solutions pour développer l'outil envisagé, la mise en œuvre du prototype d'interface 3D seront successivement abordés.

3.2.1 Terminal de communication individuel

Nombreux sont les systèmes de visioconférence qui nécessitent du matériel spécialisé (caméras, écrans, ordinateur de contrôle...), voire des salles équipées à demeure [Okada94]. Si leur efficacité est indéniable, on peut leur reprocher à la fois d'être chers, de monopoliser une grande partie de l'espace de travail. L'obligation de gérer un planning pour partager les ressources matérielles dans le temps, lesquelles en conséquence ne peuvent être utilisables spontanément par n'importe qui, est une contrainte lourde, mal vécue par les utilisateurs. [Tang92].

La solution que nous voulons privilégier est l'utilisation de l'ordinateur personnel comme poste individuel de travail de groupe. C'est un matériel maintenant accessible à tous. Sa puissance, ainsi que l'amélioration des réseaux et l'expansion de leur couverture géographique, permet d'envisager son utilisation comme outil de communication intégré pratique et performant. Le fait de banaliser le terminal permet de se connecter quasiment de n'importe où, donc n'importe quand, dès lors que le besoin s'en fait sentir. L'expansion du « multimédia » grand public a amené les constructeurs à doter de base les machines de périphériques fort utiles (son, caméras, cartes graphiques performantes...). Le matériel dont nous avons besoin est standard et rend notre outil accessible à tous.

La fusion entre l'outil de communication et l'outil de travail permet également de conserver le confort et les possibilités des applications informatiques actuelles tout en améliorant la production en permettant le travail synchrone à plusieurs. Les produits actuels sont généralement des compromis. Certains privilégient la qualité de communication, d'autres le partage d'applications et de documents. Parfois, l'utilisateur dispose de deux outils distincts [Okada94], l'un pour discuter et l'autre contenant

l'espace de travail. Ce n'est pas une hérésie dans la mesure où cette séparation correspond à des phases dans l'activité [Buxton92].

Par la suite, rien n'empêchera d'utiliser notre interface en liaison avec d'autres types de matériels (par exemple, à des amphithéâtres comme ceux de l'ENIC, voir Figure 4), la gestion de l'activité et les informations transmises étant indépendantes de la partie visible de l'interface, c'est à dire la représentation de l'activité et les processus d'interaction.

3.2.2 Téléprésence et perception de l'activité

Autant que possible, l'outil de travail doit rendre la communication transparente à l'utilisateur. Celui-ci doit être aussi à l'aise dans la situation médiatisée que dans la réalité. L'observateur doit ressentir l'impression que d'autres personnes sont effectivement présentes et qu'il fait partie de leur groupe. Cela implique que ce qu'il voit soit compréhensible, que ce qui se passe soit explicitement montré. La charge cognitive étant moindre, l'utilisateur peut travailler efficacement sans être perturbé par les défauts de l'outil. D'autres indices comme la taille des représentations des personnes distantes ont leur importance.

La voix est loin d'être le seul vecteur de communication lors d'une conversation [Bauml97] [Feyereisen91] [Morris77] [Cadoz94]. On obtient de nombreuses informations en regardant une personne. Il est facile de voir quelle est son humeur, si elle approuve, quelle est son opinion sur un sujet, en grande partie par l'expression de son visage. De même, lorsqu'elle s'exprime, ses gestes accompagnent ses paroles. Ils peuvent indiquer l'enthousiasme de l'orateur... Certains gestes, tout comme certaines expressions du visage d'ailleurs, ont une signification bien précise que tout le monde comprend. Ils aident à faire passer le message oral. Leur importance est indéniable, même si l'on s'est adapté à des moyens de communication artificiels comme le téléphone (jusqu'à tirer parti de ses limitations). Les indices visuels ne sont pas anarchiques malgré leur apparente complexité. Ils sont fortement liés entre eux et possèdent un sens précis, suffisamment pour que l'on puisse concevoir de les générer à partir de certaines règles [Cassel94]. Il faut donc les respecter, ainsi que leurs interdépendance, leur synchronisation : les divers moyens d'expressions forment un tout indissociable. La reproduction de toutes ces informations et leur synchronisation nécessite une machine puissante et des capteurs adaptés.

Les gestes ont une autre fonction bien distincte, qui est de montrer, de désigner. On emploie constamment des références à ce à quoi le geste est associé : « celui-ci est plus beau que celui-là ». Un bon système doit supporter la désignation par gestes et les habitudes de langage qui s'y rattachent.

Enfin, le regard ne doit pas être négligé. D'une manière générale, en observant le regard de quelqu'un d'autre, on est capable de savoir ce qu'il regarde, à quoi il fait référence, à qui une question s'adresse. Le fait de regarder quelqu'un dans les yeux ou, au contraire, de fuir le regard de l'autre possède un poids social très fort, d'ailleurs variable selon les cultures. Un outil coopératif devrait en tenir compte. Rappelons que la majorité des systèmes de visioconférence commercialisés utilisent une

caméra placée au-dessus de l'écran, vidant de sens la direction du regard de la personne distante [Buxton92].

Nous avons vu qu'un bon outil de communication doit donc être à même de restituer, et donc de capter, ces informations :

- contact visuel ;
- direction du regard ;
- expressions du visage ;
- gestes expressifs ;
- gestes de désignation.

L'important est de donner une impression cohérente et exploitable, que l'utilisateur du système peut immédiatement appréhender. Ces indices s'ajoutent, se complètent, se répètent. S'ils ne sont pas supportés directement, les utilisateurs cherchent à détourner l'outil pour retomber dans des schémas qu'ils connaissent et qui sont compréhensibles par tous. Une « gestuelle des oreilles » s'est mise en place naturellement dans MASSIVE. L'absence de télépointeur dans Teledesign a été résolue par les utilisateurs en agitant l'objet dont ils parlaient, ou en se construisant un pointeur à partir d'un objet « bidon » temporaire qu'ils plaçaient près de l'objet à désigner.

Un effet de bord de ces informations est de restituer le *degré de présence* d'une personne. Savoir si une personne est attentive, si elle fait autre chose en même temps ou si elle s'est absentée, facilite les interactions.

3.2.3 Environnement 3D : métaphore de table de réunion

Nous avons choisi dès le début du projet de faire usage des possibilités graphiques 3D temps réel des machines récentes. L'idée principale était d'explorer ce que la 3D pouvait apporter par rapport aux représentations 2D classiques. Celles-ci, nous l'avons vu, présentent bien des inconvénients dans le cadre du travail coopératif synchrone que nous nous sommes fixé. Les recherches en représentation 3D sont prometteuses. Les outils de Réalité Virtuelle se développent rapidement et montrent la viabilité de mondes 3D multi-utilisateurs. Il ne faut d'ailleurs pas négliger l'attrait que cette forme de représentation exerce. En particulier, dans le domaine des jeux vidéo, l'animation interactive est largement employée. L'engouement suscité par les « Doom-like » est révélateur. Ces jeux allient une représentation graphique réaliste de qualité à un maniement intuitif et direct (Figure 65).

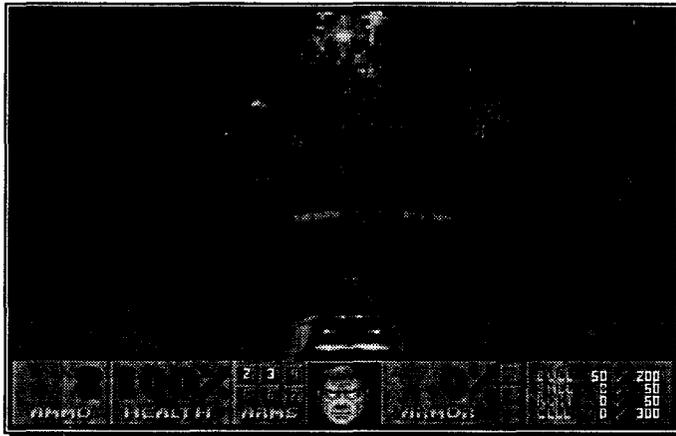


Figure 65 : le jeu Doom II

Nous avons recensé les problèmes propres au travail coopératif synchrone, par l'étude de travaux déjà réalisés dans ce domaine. Nous construisons un environnement 3D global qui résout ou évite ces problèmes après avoir recherché et analysé leurs causes. L'apport principal se situe probablement dans la possibilité d'avoir une scène cohérente en face de soi, et non une collection de vues sur des personnes et des documents. Les outils utilisant les interfaces à fenêtres classiques souffrent tous de ce problème. Les très rares propositions pour y remédier, comme Desktop-MAJIC [Tanaka96], sont justement des détournements, des « perversions » du fenêtrage. Nous pensons que la 3D est *intrinsèquement* à même de bien représenter les interactions.

Évidemment, ce choix nous contraint à repenser complètement les métaphores, celles des interfaces WIMP (bureau, documents) n'étant pas adaptées à des activités de groupe, à plusieurs utilisateurs.

Les interfaces 2D WIMP exploitent originellement la métaphore du bureau individuel. Des documents (textes, tableaux...) sont disposés sur la surface de travail. On peut placer deux feuilles côte à côte, faire des piles. L'élément central de ces interfaces est la surface de travail, et donc les documents posés dessus. Le travail coopératif synchrone nécessite de voir dans l'espace de travail non seulement des documents, mais également les autres personnes avec qui l'activité se déroule.

Nous avons donc pris une nouvelle métaphore pour représenter la scène : la table de réunion. L'avantage est de retrouver un cadre de réunion habituel et matérialisé visuellement. Son importance provient non pas d'une représentation réaliste du mobilier, mais plutôt de ce que suggère la présence de cette table : bureau autour duquel a lieu une discussion, estrade à partir de laquelle on s'adresse à une large audience, salle de réunion... Cet élément de décor est suffisamment abstrait pour pouvoir accueillir toutes sortes d'activités. Il aide l'utilisateur à retrouver ses repères et à identifier rapidement la situation dans laquelle il se trouve. Nous nous efforçons donc, à défaut d'être réalistes, de proposer à l'utilisateur par une organisation globale de l'espace de travail.

3.2.4 Une interface centrée sur l'activité

En travail coopératif synchrone, on ne peut vraiment dire qu'une activité coopérative existe qu'à partir du moment où des échanges ont lieu entre deux personnes ou plus, en plus des travaux personnels que chacun peut effectuer de son côté. C'est ce que décrit Schmidt [Schmidt91] :

« (...) le travail coopératif apparaît lorsque plusieurs acteurs sont concernés pour réaliser un travail et sont mutuellement dépendants dans leur travail et doivent coordonner et assembler leurs activités individuelles pour effectivement réaliser le travail qui leur est demandé. »

Les communications sont donc fortement liées à la notion d'activité, qui plus est si celle-ci se déroule de manière synchrone. Il nous a semblé raisonnable de postuler qu'une personne n'est donc amenée à participer qu'à une seule activité à la fois. C'est pourquoi nous avons décidé que notre interface représente une unique activité.

Ce choix est consolidé par des études de Xerox [Henderson86] qui ont montré que, dans les interfaces WIMP, un utilisateur n'utilise que quelques fenêtres (ou applications) à la fois, même si son écran en contient beaucoup. Ce petit sous-ensemble est constitué des outils nécessaires à une tâche donnée. L'idée de Xerox a été de créer des « écrans virtuels », assimilés à des pièces, chacun regroupant les quelques éléments utiles à un travail bien déterminé. Nous conservons ce principe de séparation des tâches mais au niveau du groupe, pas de l'individu. Il faut noter que leurs préoccupations diffèrent des nôtres. Leur but initial était de lutter contre l'encombrement de l'écran par de trop nombreuses fenêtres. Nous nous intéressons plutôt au moyen de renforcer la sensation de présence, de travail commun en prenant le parti de ne représenter qu'une activité à la fois. Ce cloisonnement renforce la sensation d'être impliqué dans l'activité de groupe, en étant déconnecté de l'extérieur.

3.2.5 Tout visible

La participation à une activité de groupe implique de pouvoir suivre ce qui s'y passe. Ainsi il faut avoir connaissance d'événements comme l'arrivée d'une personne, la fin d'une session, la modification d'un document ou la prise de parole. Si cette surveillance n'est pas possible, l'utilisateur risque de ne plus savoir reconstituer le déroulement de l'activité, de ne pas arriver à déterminer pourquoi et comment l'état actuel du travail a été atteint. Le passage d'une situation naturelle à son équivalent médiatisé, par un système informatique ou non, ne doit pas être un handicap. Le matériel ne doit pas imposer de limitations mais au contraire il doit respecter les capacités naturelles de perception de l'observateur et éventuellement construire des informations supplémentaires directement perceptibles. Dans une situation naturelle, des indices, surtout visuels et auditifs, viennent constamment informer les participants : mouvements, prise de parole, bruits. L'observation peut aussi être active, on peut jeter un œil au public pour avoir une idée de sa réaction.

Le système doit respecter ces « messages », la possibilité de se focaliser sur quelque chose tout en étant averti de changements importants. Le côté artificiel de la réunion médiatisée peut être utilisé pour exacerber ces indices d'avertissement, les détailler explicitement, ou pour montrer les choses de manière plus efficace. Tout ce qui se passe doit être visible, perceptible.

3.2.6 Adaptation de la Réalité Virtuelle : simplifications et abstractions

De notre point de vue, les périphériques habituels de la Réalité Virtuelle sont trop inconfortables pour supporter une activité de travail à plusieurs longue et fréquente. Nous préférons que la personne, pendant qu'elle travaille, ne soit pas coupée de son environnement réel immédiat, qu'elle garde la possibilité de chercher un livre, de demander l'aide d'un collègue. Il est impensable d'attacher les gens à leur ordinateur par des périphériques complexes comme un casque immersif ou un gant.

D'autre part, nous ne cherchons pas à imiter la réalité, nous cherchons plutôt à fournir un outil de bureau efficace, spécialisé et adapté au travail coopératif. En transposant notre approche à une salle de cinéma, nous cherchons à donner à chacun la meilleure place, alors que dans une salle normale seule une partie du public est satisfaite, les autres voyant ou entendant mal. Nous avons étudié comment se passaient des réunions, ce qui existait comme outils et ce qu'on pouvait leur reprocher ou ce qu'ils avaient comme avantages. Nous avons cherché à établir une liste de ce qu'il est intéressant de conserver d'une vraie réunion et de ce qui n'est pas indispensable et qui alourdit la tâche. En particulier, la possibilité de se déplacer devient inutile dès lors que l'on est impliqué dans une activité bien définie. Les outils de travail et techniques d'interaction, par contre, sont primordiaux. Nous avons donc travaillé à l'adaptation des techniques de Réalité Virtuelle à nos besoins par l'introduction d'abstractions.

Le premier niveau d'abstraction consiste donc à élaguer la réalité pour supprimer ce qui n'est pas directement utile, voire ce qui peut constituer une gêne. C'est le cas des déplacements. Ceux-ci sont peu utiles lors d'une réunion. Ils n'interviennent que ponctuellement (utilisation d'un projecteur vidéo, d'un tableau par exemple) pour des raisons purement physiques. Le déplacement permet de changer la place d'une personne par rapport à l'activité (changement de rôle) ou simplement de prendre possession d'un outil. Les deux sont souvent liés. Le déplacement est une fonctionnalité de base dans les environnements 3D de Réalité Virtuelle. Les interfaces WIMP l'ont supprimé depuis longtemps, l'utilisateur a tout sous la main (outils, armoires à documents). Nous voulons atteindre ce niveau de simplicité dans une interface 3D : c'est nécessaire pour un produit à vocation bureautique.

Un second niveau est l'amélioration, l'augmentation de la réalité. La Réalité Augmentée tire profit, entre autres, des techniques de Réalité Virtuelle pour surimposer à une vue réelle des informations artificielles. Dans notre cas, la simulation d'une réunion va être augmentée par l'ajout d'informations à l'écran (par exemple les télépointeurs que nous décrivons plus loin) et de nouvelles propriétés physiques (par exemple l'ubiquité).

Le déplacement est un problème difficile en Réalité Virtuelle. La solution idéale serait de reproduire les déplacements réels de la personne. C'est difficilement réalisable : il faudrait avoir la place de le faire. De plus, les objets présents dans le monde virtuel (murs, meubles, autres personnes...) doivent influencer sur la trajectoire suivie : comment simuler une collision avec un objet n'ayant pas d'existence physique ? Une autre approche du problème consiste à se déplacer indirectement, par l'intermédiaire d'un périphérique ou d'un widget adapté. C'est la solution qui donne de meilleurs résultats, même si elle risque d'alourdir le système. Les grands environnements (DIVE par exemple) fournissent parfois des véhicules mais leur maniement est peu pratique et c'est un moyen difficilement transposable dans une réunion.

En fait, beaucoup de problèmes dans la réalité proviennent de la place qu'occupent physiquement les objets. Ces places sont figées pendant la réunion, soit parce que le matériel est installé à demeure (un projecteur vidéo vissé au plafond) ou parce que la configuration adoptée convient pour la réunion (placement des tables, d'un tableau pliant). Elles induisent donc des déplacements. Une autre contrainte du monde réel est qu'à un endroit donné ne peut exister qu'un seul objet. Si un grand groupe veut se réunir, il faut un lieu adapté pouvant contenir tout le monde dans de bonnes conditions. L'accès aux documents peut devenir malaisé. Pour que toutes les personnes présentes prennent connaissance d'un rapport papier, il faut que chacune d'elles le lise à tour de rôle. Nous proposons de relaxer la disposition spatiale géométrique classique pour améliorer l'accès aux objets présents dans l'activité. En quelque sorte, nous voulons que chacun se sente à la meilleure place pour travailler. Pour supprimer tous ces problèmes, nous voulons proposer une nouvelle métaphore qui ne nécessitera pas de déplacement, comme dans les interfaces WIMP. Cependant, cela nous empêche de conserver la cohérence spatiale, c'est-à-dire que le placement relatif des objets n'est plus le même pour tous. Par contre nous conservons une cohérence globale au niveau, plus élevé, des relations entre objets.

3.2.7 Animation et fluidité

Nous avons cherché à préserver la continuité de la scène, que ce soit lors des actions de l'utilisateur ou des événements ayant des causes extérieures (typiquement les actions des autres participants). Nous avons donc fait usage d'animations partout où c'était possible. Comme il ne faut pas perturber l'observateur en attirant inutilement son attention, nous nous sommes efforcés de rendre fluides tous les mouvements, apparitions, transformations, changements d'état, modifications du point de vue, désignations, développements d'arborescences, ouvertures de documents, déplacements d'objets. Les animations doivent se faire continûment, sans sauts incompréhensibles. Il est possible de s'inspirer des jeux vidéo qui font un usage intensif de l'animation, tout d'abord pour les déplacements (courir, piloter...) mais aussi pour l'affichage de messages (bonus, fin de la partie, alerte). Les cartes ou écrans radar, les métamorphoses, les mouvements de caméra sont fréquemment employés. Les événements visuels sont systématiquement associés à des bruitages, des rappels sonores.

3.3 La réalisation

3.3.1 La table de réunion

Notre présentation s'organise autour d'une table de réunion métaphorique, un espace d'activité partagé. L'organisation doit respecter un degré suffisant de réalisme, permettre une représentation globale de l'activité et autoriser une navigation et des manipulations pratiques.

Nous considérons l'espace autour de la table comme un présentoir. Il accueille bien entendu les autres participants. Pour simplifier l'organisation et homogénéiser l'interface (à la fois la présentation, la navigation et les possibilités d'interaction), nous y plaçons également les documents utilisés lors de la réunion. Nous profitons également de ce fait des techniques mises en place pour la navigation dans l'interface. Cet espace autour de la table a été baptisé « *bandeau* » en raison de son aspect lorsque, dans les premiers temps [Saugis95], il accueillait des objets 2D (Figure 66).

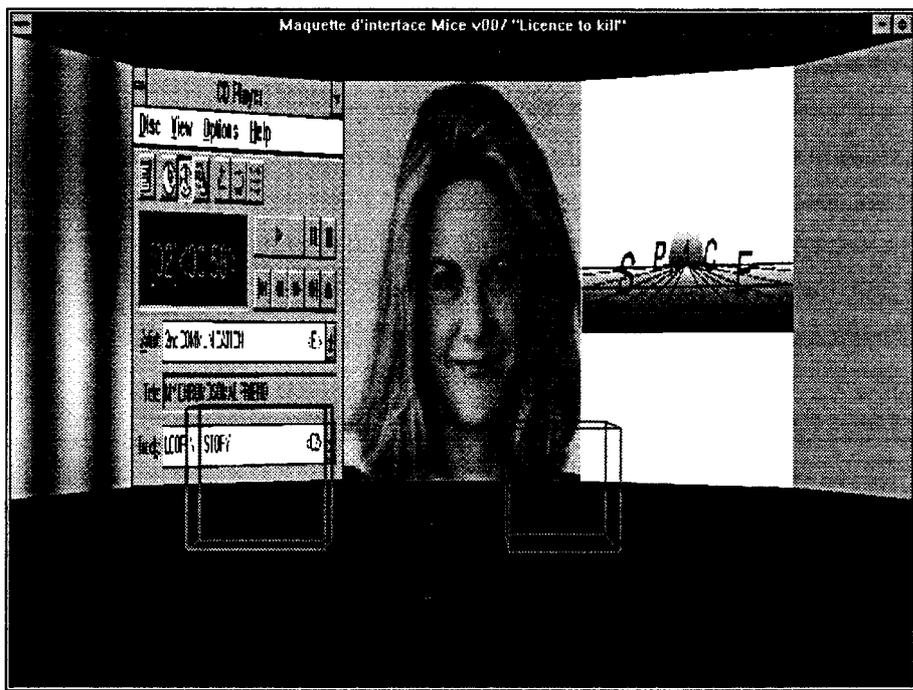


Figure 66 : le premier « bandeau »

Après tâtonnements et raffinements, nous avons déterminé que la forme la plus pratique était une table ronde (les premières maquettes plaçaient l'activité dans une « pièce carrée à trois murs » [Chaillou94]). Cela permet de ne pas avoir de places « discrètes », prédéfinies, déterminées par la géométrie de la table, son nombre de côtés. Les personnes, et les documents, peuvent donc être disposées librement autour de la table.

En particulier, cette disposition se fait sans a priori quant au nombre d'objets présents. Dans notre interface, ils occupent au mieux la place disponible. Quand un objet vient prendre place autour de la table (arrivée d'une personne, nouveau document...), ceux qui sont déjà présents « se serrent » pour lui faire une place, comme cela se passe autour d'une vraie table (Figure 67). Ce processus peut être répété autant de fois que nécessaire, dynamiquement. A l'inverse, quand un objet est retiré, les objets restants réoccupent l'espace libéré.

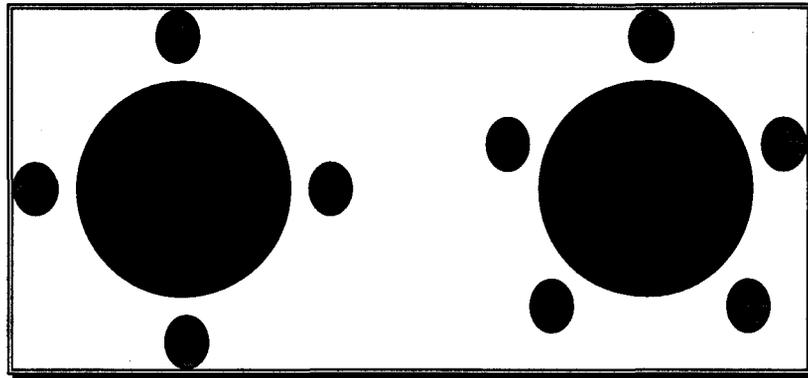


Figure 67 : 4 puis 5 personnes réparties autour de la table

La suppression des déplacements et l'organisation « en rond » a un effet de bord intéressant : tous les participants se voient directement. Dans MASSIVE, les gens se tournent parfois le dos pour utiliser un outil ou dire quelques mots à une autre personne. Cela n'arrive jamais dans notre interface. La localisation des personnes est facilitée, il n'est plus besoin d'explorer les lieux avec un champ visuel réduit pour trouver celui ou celle à qui l'on désire parler. Nous verrons également comment la table ronde rend la navigation aisée, alors qu'une autre disposition pourrait poser problème [Chaillou94].

3.3.2 Navigation

Nous désignons par navigation les actions de déplacement d'une personne dans la scène, de changement de point de vue, par opposition à la manipulation. Notre parti pris double, qui est d'une part de simplifier l'interface et d'autre part de se centrer sur l'activité, nous permet de définir une navigation volontairement limitée.

Nous adoptons le principe de séparation des tâches de *room* [Henderson86]. L'activité se déroule dans un « lieu » qui en est la matérialisation. Les utilisateurs sont plongés dans un espace dédié, avec un but précis : la tâche commune. Dans ces conditions, les seules possibilités de navigation qu'il nous faut assurer sont celles nécessaires dans la « pièce ». Les situations nécessitant, dans le réel, un déplacement seront pris en compte par des outils abstraits simples.

Les cas obligeant à se déplacer sont rares : accès à du matériel (estrade, tableau, microphone, projecteur), distribution de documents, aparté avec une personne particulière. Le reste du temps, un participant reste à sa place, d'où il peut suivre ce qui se passe, regarder les documents visibles par tous ou qu'il possède personnellement, parler aux uns et aux autres.

Nous avons cherché à réaliser notre interface en privilégiant l'utilisateur, en le considérant comme le « centre du monde », le « repère absolu » de la scène. Visuellement, l'utilisateur sera toujours à la même place par rapport à la table de réunion. Dans ces conditions, la meilleure solution consiste à fixer le regard de l'utilisateur vers la table. Pour lui permettre de regarder un objet en particulier, au lieu de changer sa direction de regard, nous préférons amener l'objet en question dans son champ de vision : ce sont les objets qui bougent par rapport à l'observateur. En pratique, les objets tournent autour de la table (ronde) jusqu'à ce que l'objet désiré soit face à l'observateur. Sur la Figure 69, l'observateur voit les objets 1, 2 et 3. Une rotation des objets autour de la table revient à bouger cet observateur. Dans l'exemple, il voit alors les objets 2, 3 et 4, comme s'il s'était déplacé.

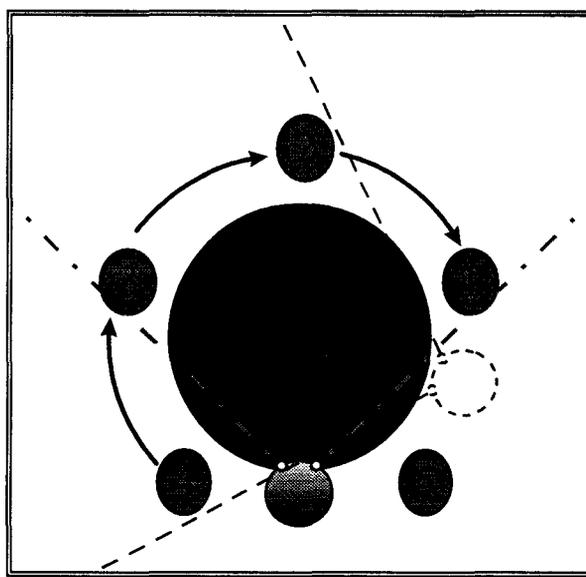


Figure 69 : Observateur fixe et objets venant dans son champ de vision

Nous avons également simulé l'approche d'un objet par l'observateur, les déplacements réels, géométriques dans la scène étant proscrits. Pour ce faire, nous grossissons l'objet regardé (celui en face de l'observateur) et nous réduisons d'autant les autres (Figure 72). Ce « zoom » produit un effet similaire à un déplacement réel vers l'objet. Nous avons défini et mis en œuvre une distorsion qui produit l'effet recherché (voir 3.3.4).

La navigation étant ainsi simplifiée, avec seulement deux degrés de liberté, un périphérique simple suffit à naviguer dans notre interface. Nous avons utilisé une souris classique. L'avantage est de pouvoir accomplir la rotation des objets et le zoom d'un seul mouvement. L'inconvénient est qu'une rotation de grande amplitude nécessite plusieurs allers et retours de la souris (amplitude limitée), une molette sans fin serait plus adaptée. Le zoom étant par contre limité, un contrôleur linéaire (avec butées en bouts de course) pourrait être plus intuitif. Nous avons aussi fait des essais avec une Spaceball isotonique. Celle-ci a l'inconvénient d'être très sensible et rend la manipulation délicate, bien qu'elle soit très efficace (mouvements rapides et précis) pour la rotation.

3.3.3 Tout perceptible

Deux choses sont nécessaires pour pouvoir suivre toute l'activité :

- l'interface doit être capable de rendre compte des événements (regard, gestes, actions...);
- tous les objets doivent rester visibles en permanence.

Le premier point, dépendant de l'interaction et de la navigation, sera abordé par la suite.

Pour que tous les objets restent visibles, ils ne doivent pas se masquer entre eux, être l'un derrière l'autre. Ils ne doivent pas non plus sortir du champ de vision de l'observateur ou, en d'autres termes, passer derrière lui. La première condition est immédiatement remplie par notre disposition qui consiste à placer les objets l'un à côté de l'autre. Pour la seconde, nous imposons l'angle de vue de l'observateur de manière à ce qu'il contienne la moitié de la table. Les objets vus seront groupés autour de cette demi-table (Figure 70).

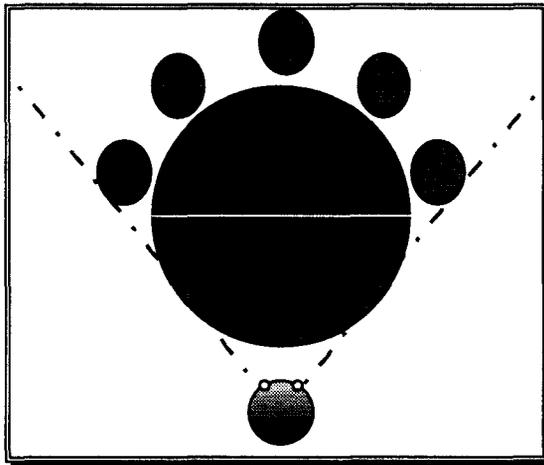


Figure 70 : objets placés autour d'une demi-table, dans le champ de vision

Cela pose problème dans le cas où l'observateur « tourne la tête », c'est-à-dire pour nous que les objets tournent autour de la table : dans ce cas, des objets disparaissent du champ de vision. La solution que nous proposons est simple : nous faisons réapparaître l'objet « sortant » d'un côté, de l'autre côté. Il est dessiné deux fois, l'une à gauche et l'autre à droite (Figure 71). Ce mécanisme induit la perte de cohérence spatiale absolue des scènes vues par les différents intervenants. De plus, il implique l'ubiquité d'un objet. Il ne nous a pas encore été possible d'évaluer la gêne occasionnée par cette réplication. D'après notre propre expérience, l'utilisation de la rotation se fait facilement. La réplication d'un objet se fait d'une part passagèrement et d'autre part à la périphérie du champ de vision, ce qui peut expliquer qu'elle ne provoque pas la sensation de « malaise » qui était à craindre. Une évaluation précise est souhaitable mais n'a pas encore été menée.

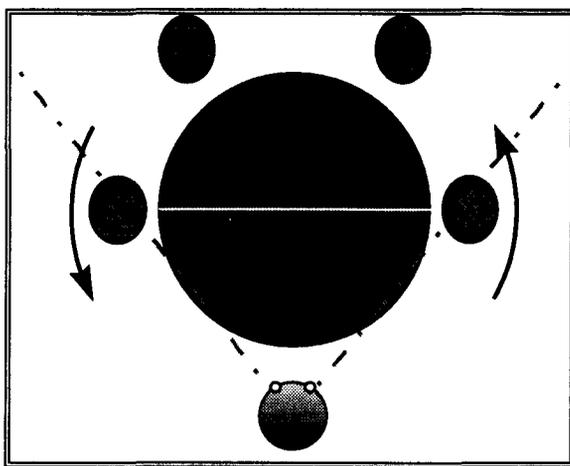


Figure 71 : disparition, à gauche, et réapparition simultanée, à droite, de l'objet X

Les objets présents doivent être exploitables. Si un trop grand nombre d'objets est affiché, ils n'occuperont chacun qu'une petite partie de l'écran. C'est ce qui rend la partie « approche/recul » de notre navigation nécessaire : le zoom, ou plus exactement la distorsion, grossit un objet précis en gardant tous les objets sur l'écran, préservant ainsi le contexte de l'activité. La qualité de représentation des objets éloignés du centre est sacrifiée, mais si une modification survient, nous savons que l'œil humain est à même de la détecter. L'utilisateur peut alors faire en sorte de porter son attention sur l'objet qui vient de changer. Dans ces conditions, nous préférons parler de « tout perceptible » plutôt que de tout visible strict. L'animation interactive permet de pallier à la perte de détail temporaire d'un objet.

3.3.4 Distorsion

Nous proposons une vue déformée « fisheye » de la scène avec un seul centre d'intérêt, ou focus, gardant tous les objets à l'écran pour préserver le contexte de l'activité. Nous avons choisi une distorsion de l'image plutôt qu'un filtrage sur le niveau de détail [Furnas86]. La fonction de déformation que nous avons mise en œuvre répond à des critères bien définis :

- distorsion horizontale uniquement. Nos objets étant placés côte à côte, ils constituent une série de « données » consécutives. Tout comme le Perspective Wall, notre déformation 1D est bien adaptée à nos besoins ;
- action limitée à la scène visible. Elle se distingue en cela d'autres distorsions prévues pour représenter un nombre « infini » de données. Elle n'a d'effet que dans un sous espace limité ;
- fonction d'un coefficient supplémentaire α . Ce paramètre, modifié par l'utilisateur à l'aide du périphérique d'entrée employé pour naviguer, permet de contrôler le degré de déformation pour passer d'une vue normale à une vue fish-eye plus ou moins zoomée. Dans notre navigation, il correspond grosso modo à l'inverse de la distance entre l'observateur et l'objet auquel il fait face ;

- dérivable selon x (la coordonnée horizontale du point subissant la distorsion) et α . Nous nous sommes imposé cette condition pour obtenir des animations fluides dans la scène. En effet, la déformation ne doit pas introduire d'à-coups ou de ruptures dans la scène. La possibilité de dériver notre fonction selon x nous garantit que, quand les objets tournent autour de la table, leur vitesse est continue. De même, si la formule est dérivable par rapport à α , un changement de zoom se fera de manière fluide. Cette contrainte de dérivation nous est apparue empiriquement : nous avons fait des essais avec des formules définies par morceaux, continues mais non dérivables par rapport à x . Les changements abrupts de vitesse des objets lors de la rotation étaient perceptibles, voire perturbants, et attiraient inutilement l'attention ;
- grossissement du centre de l'écran (Figure 72). Fixer la place du centre de la déformation nous permet de déterminer à quoi un observateur s'intéresse. En effet, *notre déformation l'incite à placer l'objet constituant son centre d'intérêt à l'endroit où le fish-eye fournit l'image la plus détaillée*. Il aurait été tentant d'utiliser des dispositifs de suivi du regard²² mais ceux-ci sont encore encombrants et gênants (Figure 26) et vont à l'encontre de notre politique de simplification de l'outil.

Notons que cette déformation peut également nous servir à changer le niveau de détail des objets. Par exemple, pour améliorer les performances, le niveau de détail (LOD, *level of detail*) des objets 3D annexes, c'est-à-dire le nombre de facettes ou la précision des textures, peut être réduit suivant l'intérêt qui leur est porté par l'utilisateur.

Notre scène s'inscrit entre les plans $x = -1$ et $x = 1$. La formule de notre déformation pour $0 \leq x \leq 1$ est la suivante :

$$x' = 1 - (1 - x)^\alpha$$

elle est symétrique dans $-1 \leq x \leq 0$. La distance (entre 0 et 1) du point à déplacer au bord droit est élevée à la puissance $\alpha \geq 1$. Cette distance diminue un peu pour les points près du bord, beaucoup plus pour ceux du centre.

²² Eye tracking

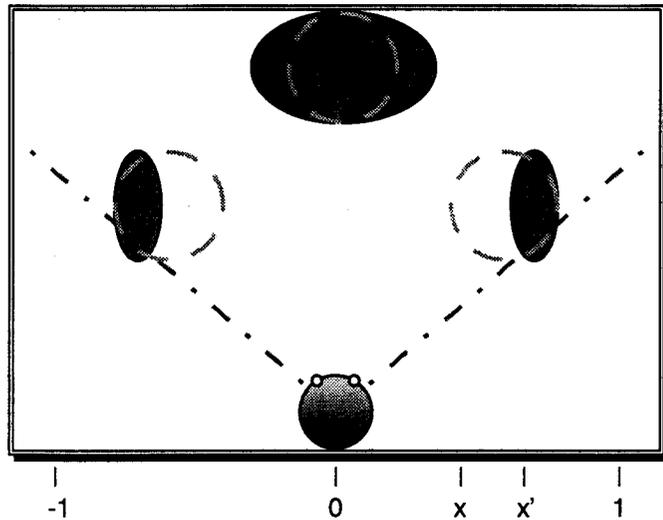


Figure 72 : la distorsion, schématisée

La Figure 73 montre un bandeau à sept « murs » (objets 2D) subissant une distorsion, pour plusieurs valeurs du coefficient α .

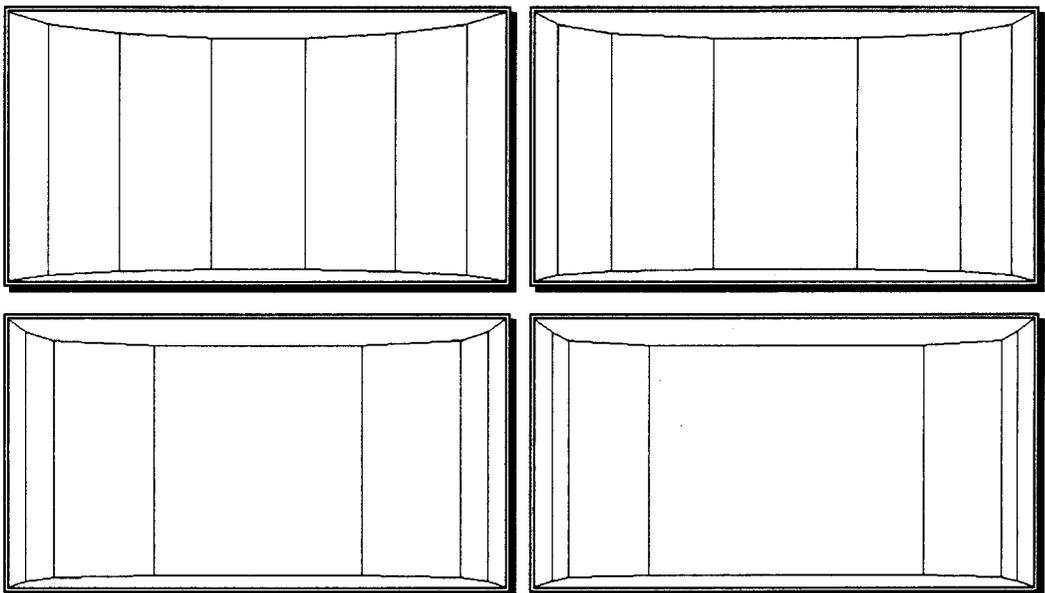


Figure 73 : le bandeau subissant une distorsion

3.3.5 Manipulation

Nous privilégions un mode d'action direct. L'utilisateur doit pouvoir accéder aux objets et à leurs fonctions de manière immédiate (par exemple, il ne doit pas avoir besoin du clavier). C'est pourquoi la définition et la mise en place des interactions est importante dans SPACE et fait l'objet d'une autre thèse dans l'équipe Graphix du LIFL.

L'action la plus simple est le pointage d'un objet dans l'espace 3D. Cela suppose d'avoir un périphérique adéquat (au moins trois degrés de liberté) et de visualiser le point d'action. L'utilisateur de Spln possède donc au moins un périphérique isométrique 3D. A l'heure actuelle nous utilisons un capteur à ultrasons OWL de Pegasus, qui se présente comme une bague munie de deux boutons. Un pointeur 3D, une flèche, se déplace dans l'espace virtuel en même temps que la main de l'utilisateur. Le pointeur permet de désigner et de sélectionner un objet et d'agir dessus.

Nous désirons développer les interactions à deux mains [Kabbash94]. Le principe général est d'avoir une main servant de point de repère tandis que l'autre effectue l'action. Par exemple, quand quelqu'un écrit sur une feuille de papier, une main tient le stylo et forme les lettres tandis que l'autre constitue un point fixe : on s'aperçoit en effet que la main dominante trace les caractères dans une zone réduite et que l'autre main déplace le papier au fur et à mesure. Nous voulons également mettre à profit des périphériques de types différents, typiquement un premier isométrique et un deuxième isotonique. Nous travaillons actuellement au couplage d'une OWL (isométrique) et d'une Spaceball (isotonique). Les actions à deux mains générales doivent être définies (par exemple, sélection par l'OWL, rotation avec la SpaceBall). Les actions de base [Poupyrev97] avec des manipulations 3D devront être réalisables facilement.

Nous ne développerons pas plus la manipulation dans Spln ici. Elles n'est en effet pas entièrement finalisée et doit encore évoluer pour permettre la mise en place d'applications complètes.

3.3.6 Clones

Représentation des personnes

Plusieurs solutions étaient envisageables pour représenter les personnes distantes dans notre interface.

- la *photographie* : c'est la méthode la plus simple techniquement. Sa diffusion se fait une fois pour toutes sur le réseau et le placage de texture est réalisé par le matériel, en l'occurrence l'accélérateur graphique 3D. Une simple image fixe permet d'identifier rapidement et naturellement une personne. Par contre, par sa nature figée, elle n'est pas directement à même de traduire certains événements de l'activité. Par exemple, on ne voit pas qui parle. On est forcé de reconnaître la personne par le timbre de sa voix. C'est particulièrement gênant quand la réunion se passe entre personnes ne se connaissant pas encore. On retombe sur les problèmes liés aux réunions purement sonores (par téléphone). Le système pourrait indiquer explicitement cette personne, pourvu qu'il soit capable de détecter le locuteur à coup sûr. Desktop-MAJIC emploie plusieurs images indiquant ce qu'une personne regarde et si elle parle. La personne peut être détournée à l'avance pour s'intégrer dans une scène *continue* ;
- la *vidéo* : beaucoup plus intéressante qu'une simple photographie, elle retransmet le visage et les gestes. Elle a malheureusement de nombreux inconvénients. Elle nécessite, pour son transport, des réseaux à large bande passante. Le détournement dynamique, s'il est possible, est

un processus consommant une partie de la puissance de calcul de l'ordinateur. L'utilisation classique d'un fond bleu facilite grandement la tâche mais est une contrainte pour l'utilisateur qui doit pouvoir se connecter de n'importe où. La cohérence spatiale n'étant pas assurée dans notre interface, le contact visuel ne peut être préservé de la même manière que dans MAJIC ;

- les *clones* : on désigne par ce terme la représentation synthétique 3D réaliste d'une personne. Les premiers travaux ont cherché à analyser l'expression d'un visage. Dès le début, il s'est agi de repérer la position de points clés (coins de la bouche par exemple) à l'aide de pastilles aisément visibles. Les systèmes actuels, comme celui de l'INA [Viaud96] ou de VLNet [Capin97] effectuent l'analyse en temps réel à partir d'une simple image numérisée. Ces points de contrôle servent ensuite à modifier un modèle 3D pour lui appliquer l'expression de la personne réelle. Il est possible d'avoir des clones très réalistes : le visage de la personne réelle est transformé pour donner une texture 2D et un modèle 3D qui seront utilisés pour générer son double. Pour utiliser un clone, il faut d'abord transmettre le modèle original. Pendant la communication seuls seront transmises les coordonnées des quelques points de contrôle. Une autre méthode de clonage est de plaquer un flux vidéo sur un modèle 3D. L'avantage est de conserver de manière simple les expressions du visage. Par contre, les inconvénients de la vidéo demeurent (transferts réseau importants, mauvaise capture de la direction du regard et difficulté à s'intégrer dans une scène virtuelle donc différente de l'environnement réel de la personne filmée).

Intérêt des clones

Même si nous n'écartons pas la possibilité d'utiliser d'autres modes de représentation, les clones fournissent une bonne solution pour incarner une personne à distance. En tant que vrais objets 3D, ils sont mieux à même de figurer dans notre scène que des supports 2D. La charge réseau qu'ils occasionnent est insignifiante [Quéau93], à tel point que l'animation faciale sera incluse dans la norme MPEG-4 [Doenges97]. Ils sont faciles à manipuler et se mélangent facilement à des objets virtuels que des images issues directement du monde réel. Le clone peut éventuellement être manipulé pour faire plus qu'une reproduction fidèle de la réalité, par exemple adopter une expression caricaturale.

Pour des raisons de puissance de calcul, l'acquisition vidéo et l'analyse du visage (partie développée par l'INA) fournissant l'expression de la personne sont déportées sur une deuxième machine (SGI Indy). Dans SPACE, le point d'intérêt et le point d'action sont distincts. Notre système utilise la possibilité de modifier la direction du regard d'un clone. La désignation d'un objet par un clone se fait par l'intermédiaire d'un télépointeur. Dans l'interface actuelle, l'animation faciale de l'INA n'est pas utilisée dans SpIn. L'aspect réaliste, les yeux, les télépointeurs rendent les clones vivants. L'ajout de nouvelles fonctionnalités ne se fera que si elles apparaissent comme nécessaires.

i) Yeux

Le focus de chacun étant détecté grâce à notre distorsion, nous faisons en sorte que cette information soit visible par les autres utilisateurs. Dans la réalité, ce sont les yeux des autres qui nous renseignent. Tout naturellement, nous forçons les clones à regarder vers l'objet que son maître regarde, i.e. celui qu'il a placé au centre de son bandeau. Ainsi nous sommes capables de restituer le

contact visuel et de permettre aux utilisateurs d'avoir conscience du centre d'intérêt d'un autre participant. La tête tourne aussi, accompagnant le mouvement des yeux, ce qui ajoute au naturel et évite des angles de vue irréalistes.

Ce mécanisme nous autorise à ne transmettre que des informations de haut niveau (qui regarde quoi), ayant une signification sémantique, plutôt que les positions spatiales « brutes » des objets.

Cette manipulation n'est pas exclusive au clone. Les photographies pourraient en profiter : il suffirait de redessiner les yeux sur l'image 2D de la personne. Avec de la vidéo, par contre, cela implique de repérer dynamiquement où sont les yeux de la personne filmée pour rediriger son regard. Les algorithmes à mettre en œuvre ne sont pas simples et consomment de la puissance de calcul.

ii) Télépointeur

Un utilisateur peut sélectionner un objet dans son interface. Son clone doit désigner cet objet pour matérialiser le point d'action de son maître. Pour ce faire, chaque clone est doté d'un télépointeur. Il s'agit d'un lien visuel direct, immédiatement observable, entre le clone et l'objet manipulé par son maître.

Le télépointeur a l'avantage d'être très explicite, précis. Il peut remplacer avantageusement les gestes de désignation. Pour plus de naturel et de dynamique, le télépointeur peut être considéré comme un outil prolongeant le bras du clone. Dans ce cas, le bras accompagne le télépointeur dans ses déplacements.

3.3.7 Vue subjective

Les premières applications partagées, en 2D, étaient WYSIWIS. Les bibliothèques graphiques comme Xlib étaient adaptées pour que les messages (clic souris...) soient retransmis vers les machines des autres participants. Tout le monde était forcé d'adopter les mêmes fenêtres en raison de ces méthodes primitives de réplication de l'application, sauf celle qui avait le contrôle de l'application. Il n'était pas possible aux autres utilisateurs de visualiser un autre document, ou de changer le mode de représentation. Le WYSIWIS strict est très contraignant et impose des méthodes de travail rigides.

Très naturellement, les CVE ont « relâché » le WYSIWIS. Le monde partagé est le même pour tout le monde, mais chacun a un point de vue qui lui est propre. La vue résulte donc de la liberté d'exploration que permet l'environnement. Il est important de noter que la collaboration n'est possible que parce que les personnes sont représentées. Si ce n'était pas le cas, chaque utilisateur aurait l'impression d'évoluer dans un monde statique ou qui changerait sans raison apparente. Un projet comme MASSIVE est une bonne illustration de l'importance des représentations des utilisateurs dans un environnement virtuel partagé : elles assurent la coordination et les interactions homme-homme. Le système peut donc donner plus de liberté aux utilisateurs.

Le WYSIWIS relâché est bien sûr possible en 2D mais dans ce cas une adaptation des données distribuées est nécessaire. Prenons pour exemple un long texte. Si une personne commente un paragraphe particulier, elle va entourer, souligner, désigner certains mots ou phrases. En WYSIWIS strict, chacun voit le fragment de texte concerné. Il peut être intéressant pour un utilisateur de momentanément retourner lire un passage précédent, qui éclaire le paragraphe courant. Sa vue va donc changer par rapport à celle du locuteur. Il ne suffit plus de passer des informations de bas niveau (curseur en (x,y) , clic gauche souris...) qui permettraient, dans le WYSIWIS strict, de reproduire les actions du locuteur privilégié. Une phase d'analyse et d'interprétation est nécessaire. La position du pointeur est transformée en un emplacement dans le texte (tel mot dans tel paragraphe). Cette information de plus haut niveau est transmise. Les autres utilisateurs voient la portion de texte désigné changer, et ceci quelles que soient la taille et la position de la fenêtre partagée sur leur machine. De tels mécanismes peuvent être étendus en 3D [Smith97]. Il est même possible de jouer sur la manière de représenter les données plutôt que de changer juste le point de vue ou l'organisation, par exemple en jouant sur la transparence, le niveau de détail.

Nous proposons d'aller plus loin encore dans le relâchement des vues de chacun. Notre organisation des objets de l'activité en bandeau autour de la table est intrinsèquement abstraite et ne se conforme pas à une métrique spatiale particulière. Nous perdons la cohérence spatiale mais nous gagnons en souplesse. Le placement des objets dans chaque vue est libre : un utilisateur peut, dans sa vue, ordonner les objets présents comme bon lui semble : face à lui le document qui l'intéresse, sur les côtés des personnes passives qui ne font qu'écouter... La scène est construite de manière complètement synthétique pour chaque utilisateur (c'était déjà un peu le cas dans [Seligmann95]). La cohérence d'une vue à l'autre existe pourtant, mais pas dans l'organisation spatiale. Des informations de haut niveau sont partagées. Spln les utilise au moment de reconstruire la vue. Ces informations concernent en premier lieu les documents. Quand un utilisateur effectue un changement sur un document, cette action est décrite, transmise, interprétée chez les utilisateurs distants. Les documents restent donc cohérents d'une machine à l'autre. De la même manière, on peut considérer qu'un acteur est une collection d'attributs décrivant son état et son comportement. L'exemple le plus simple est la direction du regard. Là où MASSIVE ne fait état que de la position et de l'orientation d'une personne, nous transmettons un renseignement du type « la personne X regarde l'objet Y ». Rappelons que nous extrayons cette information par le biais de notre système de navigation : nous assimilons l'objet regardé à l'objet placé face à l'utilisateur.

Spln construit donc sa scène en fonction des données de haut niveau reçues (actions sur les documents, direction du regard des personnes distantes, objets qu'elles désignent etc.). Nous profitons pleinement des capacités de manipulation de l'apparence d'un clone 3D. Nous faisons effectivement regarder le clone vers l'objet qui intéresse son maître et ceci quelles que soient les positions relatives de (la copie locale de) l'objet et de (la représentation sous forme de clone de) la personne.

3.4 Éléments sur l'implémentation

3.4.1 Plate-forme d'accueil

Notre interface a entièrement été développée sur et pour compatible PC avec le système d'exploitation Windows NT. Nous avons utilisé l'environnement logiciel Microsoft Visual C++. Nous avons beaucoup profité des possibilités offertes par ce langage orienté objet. De plus, Visual C++ a très tôt supporté l'utilisation de la bibliothèque graphique 3D temps réel OpenGL. Celle-ci a été choisie pour les nombreuses fonctions qu'elle offre (calculs matriciels, transparence, texture etc.), pour sa présence sur de nombreuses plates-formes (Windows 95/NT, SGI, Sun...) et surtout parce qu'elle est fournie en standard avec Windows 95 et Windows NT. C'est l'assurance de l'apparition rapide de cartes accélératrices très performantes à des coûts dérisoires pour les PC, notamment pour les jeux (citons Quake, Quake II et Hexen II par exemple). Pour toutes ces raisons, la bibliothèque OpenGL a été préférée à ses concurrentes (Direct3D, RenderWare...).

Les essais ont été réalisés sur des machines Intergraph GLZ et HP Vectra, munies d'accélérateurs 3D OpenGL. Si elles pouvaient être considérées comme du haut de gamme lors de leur achat pourtant pas très ancien, l'évolution technique a répondu à nos attentes. Les machines grand public actuelles offrent des performances presque comparables à des prix accessibles.

Nous avons essayé de rendre SPACE indépendant de la plate-forme d'accueil. Le langage (C++) et la librairie graphique (OpenGL) sont standard et portables. Les adaptations à effectuer pour le portage sur un système autre que PC/Windows concernent les échanges avec l'extérieur (gestion des périphériques par exemple) et quelques fonctions du système de fenêtres (typiquement, ouvrir une fenêtre pour SPACE et déclarer qu'elle va recevoir des images issues d'OpenGL). Le portage sous Unix serait très simple et très rapide.

Précisons que l'analyse des visages développée par l'INA, employée lors du clonage, s'effectue (acquisition + traitement) à part sur une SGI Indy. Les informations résultant de cette analyse sont transmises par réseau au Spln de la personne filmée puis répercutées chez les autres participants (via ScenIC et le serveur central). Notre prototype est prévu pour pouvoir fonctionner même sans l'animation faciale, facultative.

3.4.2 Objets graphiques

Chaque objet graphique de Spln est indépendant. Il doit se charger lui-même de sa représentation : un texte, un clone, un objet 3D... Tous gèrent leur affichage propre. Ces différents objets sont placés, assemblés pour constituer la scène. Nous ne faisons qu'une supposition sur ces briques de base : elles doivent s'inscrire dans un cube de côté 1, « pose » sur le plan $y=0$, dont la *face avant* est dans le plan $z=0$ et tournée vers $z>0$ (Figure 74). La « face avant » est importante pour orienter convenablement certains objets : un clone, une image 2D... Les objets placés dans le bandeau sont

également mieux visibles (plus proches, dans le champ de vision) lorsqu'ils sont tout contre la table. La boîte englobante ainsi définie est utilisée pour la composition d'objets (par exemple pour placer des visages côte à côte en « mosaïque ») et pour le pointage.

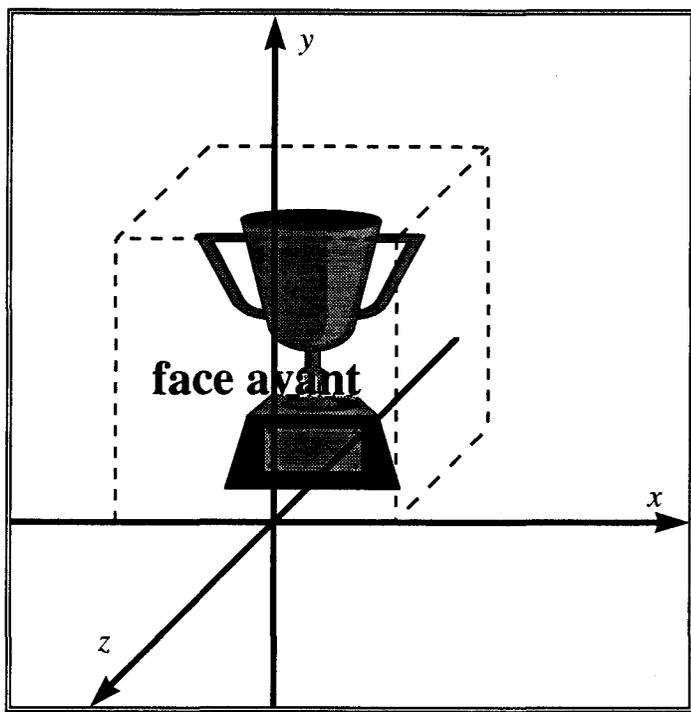


Figure 74 : boîte englobante de chaque objet graphique

Pour être compatible avec les standards existants, Spln construit un arbre VRML de la scène. Tout l'affichage se fait à l'aide des fonctions OpenGL. À terme, cette conformité permettra d'inclure Spln dans les browsers type Netscape Navigator, sous forme de plug-in. Nous souhaitons également respecter la norme MPEG-4 sur les transports d'informations multimédia, les clones par exemple s'y prêtent bien.

Notre système n'a pas pour but de se substituer à un système d'exploitation complet mais plutôt de valider nos concepts et idées de mise en œuvre. Tout le prototype est compilé d'un bloc et contient les applications dont les utilisateurs vont pouvoir se servir. Dans le cas d'une mise en œuvre à grande échelle, les applications devraient bien sûr être séparées du noyau de SPACE et lancées dynamiquement. Une bibliothèque gérant l'affichage devrait alors être développée et serait à Spln ce qu'est Xlib à X-Window.

3.4.3 Apparition d'un nouvel objet dans le bandeau

Comme décrit précédemment, les objets placés en demi-cercle prennent toute la place disponible. L'arrivée d'un nouveau participant ou d'un nouveau document dans le bandeau se fait progressivement, en animation. L'objet vient s'insérer et grossit jusqu'à ce qu'il ait la même taille que

ses voisins. La taille (angulaire) des objets est diminuée pour que tous puissent tenir dans le demi-cercle (Figure 83).

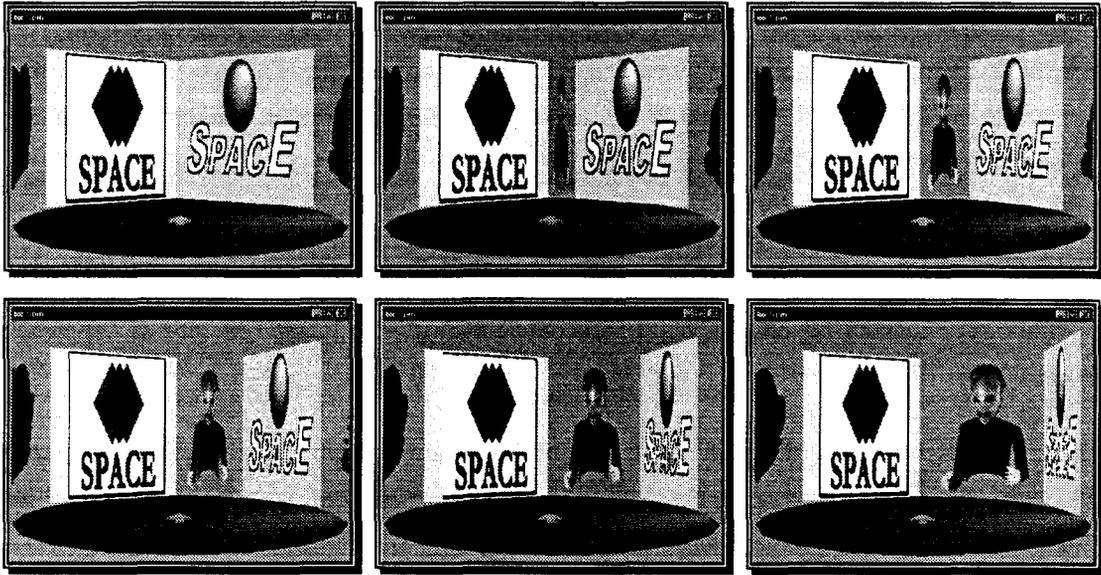


Figure 75 : arrivée d'un participant

Lorsqu'un objet quitte le bandeau, c'est exactement l'inverse : l'objet rétrécit jusqu'à ne plus exister.

3.4.4 Pointage et objets composés

Les manipulations, en particulier la sélection, se font dans notre interface par l'intermédiaire d'un pointeur 3D. Celui-ci se déplace avec un périphérique d'entrée 3D isométrique et reproduit donc les évolutions de la main de l'utilisateur dans l'espace. Il apparaît initialement comme une flèche. Cette forme n'est pas figée et peut se modifier comme les pointeurs 2D se changent parfois en sablier, en panneau d'interdiction et montrent visuellement l'action en cours. Pour faciliter son repérage [Wanger92], le pointeur et les objets projettent une ombre sur la table. L'orientation du pointeur est modifiée en fonction de sa position pour accentuer la sensation de déplacement [Venolia93].

Nous avons implémenté un système simple et générique de détection de « collision » entre le pointeur et un objet 3D (éventuellement composé de sous-objets). Ce problème, formulé autrement, revient à déterminer si un point (la pointe du curseur) est à l'intérieur de la boîte englobante (un parallélépipède rectangle) de l'objet.

La méthode que nous employons est la suivante : pour chaque opération de déplacement, rotation ou dimensionnement subie par l'objet (laquelle vient s'accumuler à la matrice M de transformation de l'objet), nous mémorisons l'opération inverse et construisons au fur et à mesure une matrice inverse (M^{-1}). Quand un objet est constitué de sous-objets, la hiérarchie d'objets est mémorisée et les matrices de transformations inverses des sous-objets sont également calculées (en partant de la matrice

inverse de l'objet les contenant) et mémorisées. Le bandeau est un exemple (un peu particulier) d'objet composé : pointer dans le bandeau revient à pointer dans l'un des objets qui le composent.

La détection se fait en transformant les coordonnées du pointeur dans le repère global à l'aide de cette matrice inverse. Si le point obtenu est dans la boîte englobante de base (Figure 74) alors le curseur pointe sur l'objet. L'arbre des sous-objets peut être parcouru récursivement jusqu'à déterminer le plus petit objet désigné par le pointeur (Figure 76).

Remarquons que la matrice inverse détermine en fait les coordonnées du pointeur dans le repère local de l'objet. La méthode de détection, pas toujours suffisamment précise, par boîte cubique étant une « méthode virtuelle » de chaque objet, elle peut être raffinée au cas par cas si l'application le demande.

Cette méthode est d'une efficacité correcte et en tout cas suffisante pour nos besoins. Elle utilise les fonctions de calcul géométrique d'OpenGL qui peuvent être câblées dans l'accélérateur graphique. De plus, les boîtes sont peu nombreuses car chaque objet ou sous-objet est une primitive d'assez haut niveau (le but n'est pas de descendre au niveau de la facette). Elle est simple à utiliser et s'adapte facilement à des utilisation pointues spécifiques.

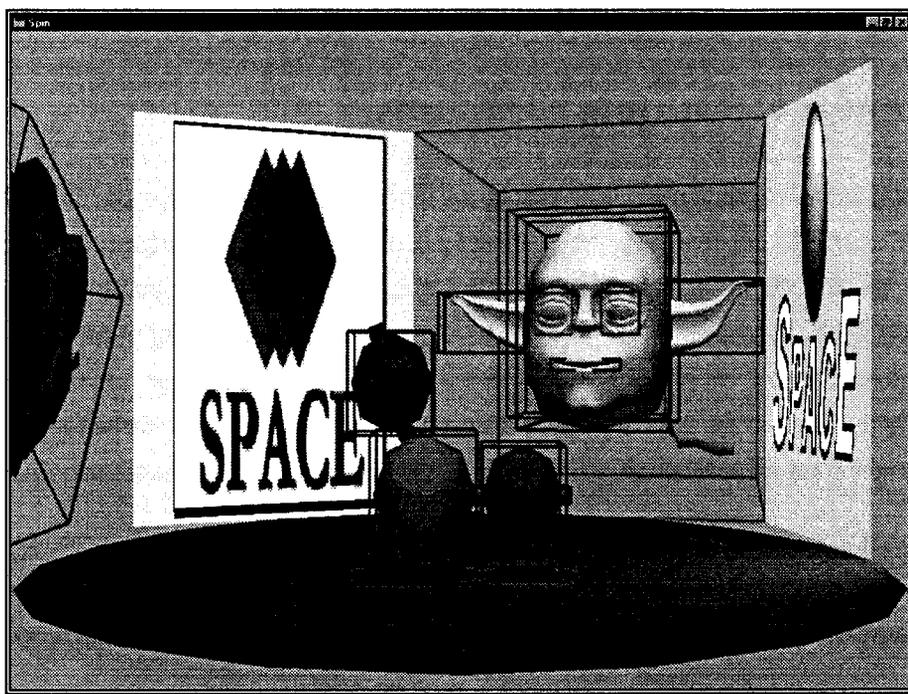


Figure 76 : objets composés et boîtes englobantes

3.4.5 Messages

L'interaction fonctionne à l'aide d'un système d'envoi de messages classique (programmation événementielle). Celui-ci a été développé spécifiquement pour Spln et est sans rapport avec les files

d'événements propres à Windows. Seuls les périphériques sont gérés, à bas niveau, par des messages Windows, lesquels sont renvoyés après traduction à Spln.

Une file d'attente reçoit les messages. Elle est vidée au fur et à mesure, les messages qui y sont stockés étant communiqués à l'objet récepteur, généralement l'objet sélectionné par le pointeur. Chaque objet contient un ensemble de fonctions (ou méthodes) réagissant aux divers messages reçus.

Une fonction « *idle* » a été implémentée, elle est appelée lorsque la file de messages est vide, i.e. lorsqu'il n'y a rien à faire. Elle est très utile pour gérer, par exemple, les animations ou toute autre action devant se poursuivre de manière autonome.

3.4.6 L'échec de l'intégration directe d'applications Windows

Nous avons longtemps eu l'espoir d'intégrer des applications 2D déjà écrites dans notre interface. L'intérêt est évident : dès sa création, notre système aurait disposé d'une énorme bibliothèque de logiciels. Nous avons donc étudié la viabilité technique de cette inclusion.

Les essais ont été menés sur compatible PC sous Windows NT, une plate-forme qui cumule les avantages : diffusion très large, prix relativement bas et l'existence d'une pléthore de logiciels variés. Nous nous sommes penchés sur les possibilités de ce système (intercepter des routines d'affichage, simuler la présence d'un périphérique), sur les mécanismes utilisés par les applications (déroulement d'un programme, comportement réactif aux messages reçus etc.). Notre but était de récupérer en temps réel des fenêtres sous forme de textures (i.e. d'images bitmap 2D) pour les afficher dans Spln, puis de piloter les applications à partir d'actions effectuées dans Spln sur ces images. Typiquement, nous voulions qu'en plaçant notre pointeur 3D sur une image de fenêtre et en cliquant, l'application Windows correspondante reçoive un message l'informant que l'utilisateur avait cliqué dans la vraie fenêtre.

Les difficultés purement techniques ont été nombreuses. Windows est un système fermé, prévu pour fonctionner d'une certaine manière et dont les API reflètent la « philosophie ». En conséquence, nous n'avons pas réussi à faire dessiner une application dans une fenêtre « virtuelle », en mémoire. L'affichage se faisait forcément directement sur l'écran. Cela impliquait que, pour pouvoir récupérer une fenêtre, celle-ci devait être complètement visible, elle ne devait ni sortir de l'écran ni être recouverte par d'autres fenêtres. La place sur l'écran devait être partagée entre l'application à récupérer et la fenêtre contenant SPACE.

Si cette solution était bien sûr décevante, ce n'était qu'à cause des possibilités d'un système particulier et non sur le concept même de l'intégration de fenêtres. Il aurait pu être envisageable (en toute théorie !) de réécrire une partie de Windows et/ou des applications que nous voulions capturer pour que l'affichage se fasse selon des contraintes que nous aurions imposées. Dans le même ordre

d'idée, en portant SPACE sous X-Window, beaucoup plus ouvert, nous ne nous serions plus heurtés aux limites du système. Des serveurs X-Window ont déjà été réalisés dans le but de faire afficher des fenêtres dans un environnement 3D.

Nous savions que la technique pouvait être adaptée pour servir nos besoins. De même, simuler des interactions (souris, clavier...) avec une application était possible. La vraie difficulté tient au déroulement des applications dans un système comme Windows. Les fenêtres sont réellement les « atomes » qui composent ces programmes. De nouvelles sont constamment créées puis détruites : menus, boîtes de dialogues (où chaque bouton est lui-même une fenêtre), widgets etc. L'inclusion d'une application existante implique d'être informé de l'existence de toutes ces fenêtres secondaires et surtout de savoir quoi en faire. Dans Windows, seule l'application reconnaît les messages déclenchant l'apparition d'une nouvelle sous-fenêtre. Si l'apparition d'un menu sur une fenêtre est adapté à une représentation 2D, comment la représenter dans une interface 3D comme la nôtre, avec notamment les contraintes de placement que nous avons fixées ?

Malgré nos efforts, nous avons été finalement contraints de renoncer à la récupération directe d'applications déjà écrites. En supposant que l'on dispose des fichiers sources, disons, de Word, la partie à réécrire pour l'adapter à notre type d'interface serait énorme étant donné que *rien ne correspond directement à une fenêtre 2D* dans SPACE.

3.4.7 La distorsion horizontale

Nous avons eu un choix à faire sur la mise en œuvre de la distorsion des objets placés autour de la table (Figure 77).

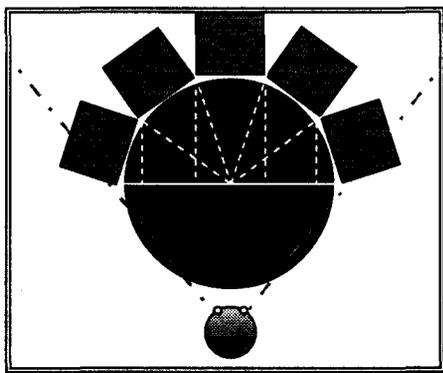


Figure 77 : bandeau à cinq objets, sans distorsion

Deux solutions étaient envisageables (NB : la formule de la section 3.3.4. peut s'appliquer dans les deux cas). La première était de jouer sur la taille angulaire des objets placés en rond autour de la table (Figure 78). L'objet le plus au centre aurait représenté une proportion importante du périmètre de la table.

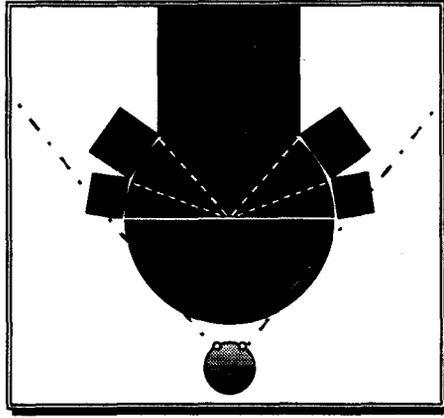


Figure 78 : distorsion angulaire

La seconde étape d'étirer horizontalement ces objets après qu'ils aient été placés (Figure 79).

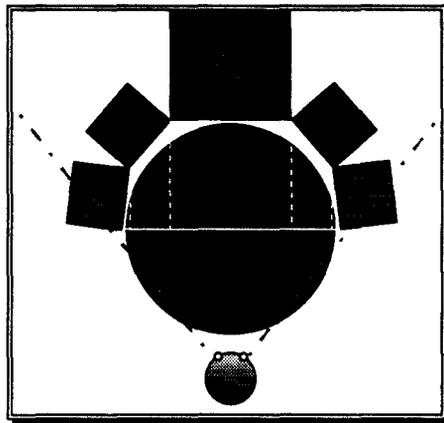


Figure 79 : distorsion horizontale

Nous avons gardé la solution de l'étirement horizontal. La distorsion angulaire a dû être rejetée car l'objet central empiétait sur la table lors de déformations prononcées.

La distorsion s'effectue comme suit. Les positions des coins de la face avant (x'_1, z'_1) et (x'_2, z'_2) de l'objet est d'abord calculée par la formule vue à la section 3.3.4. En sont déduits les nouvelles position (x', y') , taille d et orientation α . L'objet peut alors être affiché.

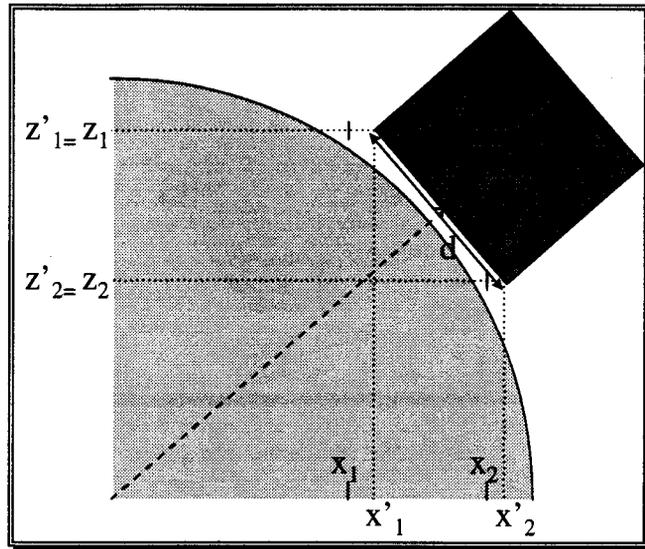


Figure 80 : transformations à appliquer à l'objet pour la distorsion

3.4.8 Écrasement latéral des objets du bandeau

Dans un premier temps, la disposition des objets autour de la table en bandeau a été mise au point avec des images 2D, de simples rectangles texturés. Quand nous avons introduit les objets 3D, nous avons eu un problème : les objets placés à gauche et à droite sortaient du champ de vision, i.e. de l'écran (Figure 81).

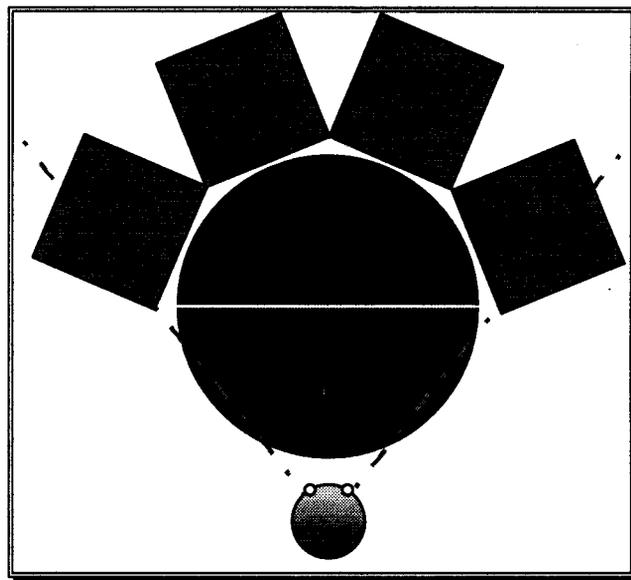


Figure 81 : les objets 3D périphériques sortent de l'écran

Nous aurions pu élargir l'angle de vision horizontal ou reculer l'observateur par rapport à la table. Dans les deux cas, les objets auraient bien tenu dans le champ de vision mais leur taille, en particulier celle de l'objet central, aurait été moindre, d'où une perte de détail.

La solution que nous avons appliquée est d'aplatir les objets vers leur face avant au fur et à mesure qu'ils s'éloignent de la position centrale (Figure 82). Ainsi les objets restent visibles dans leur intégralité, bien qu'étant déformés. Un clone 3D se transforme sur le bord en « photographie synthétique » 2D. Ce procédé laisse intact l'objet central.

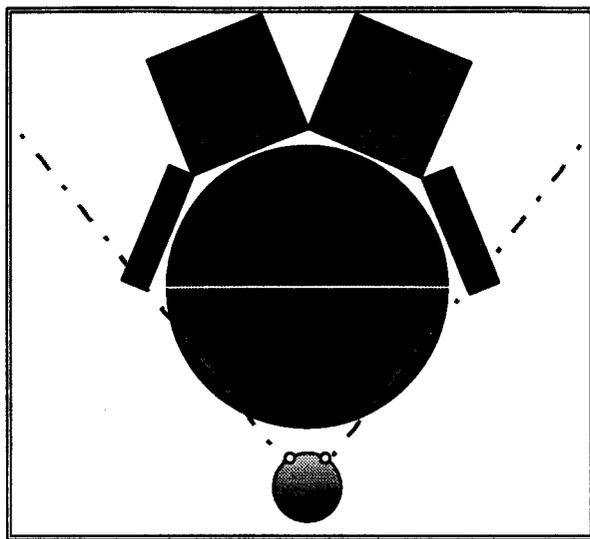


Figure 82 : aplatissement des objets périphériques

Les vues en perspective de la Figure 83 et de la Figure 84 illustrent l'avantage de notre méthode du point de vue normal de l'utilisateur.

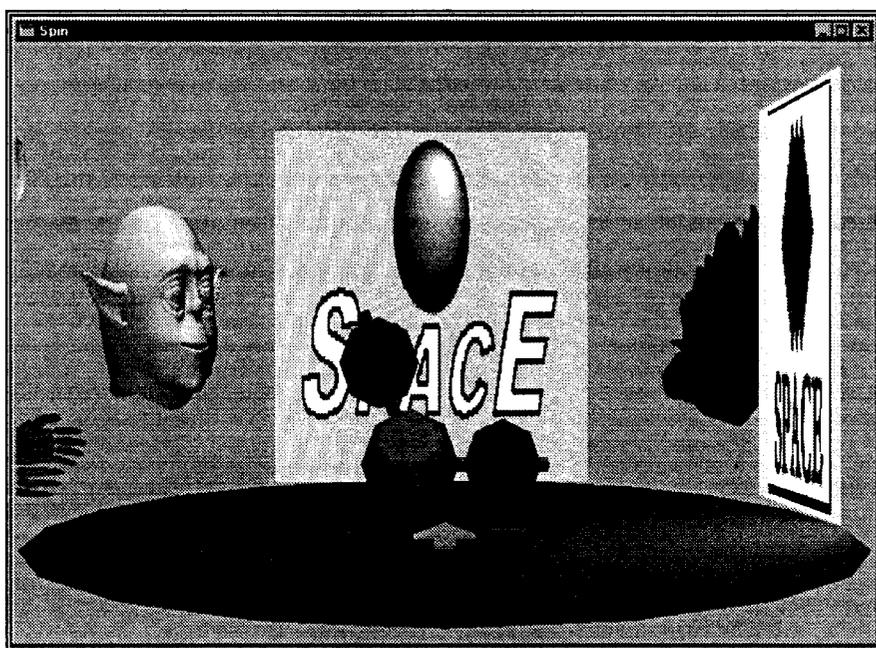


Figure 83 : pas d'aplatissement latéral, le clone de gauche disparaît

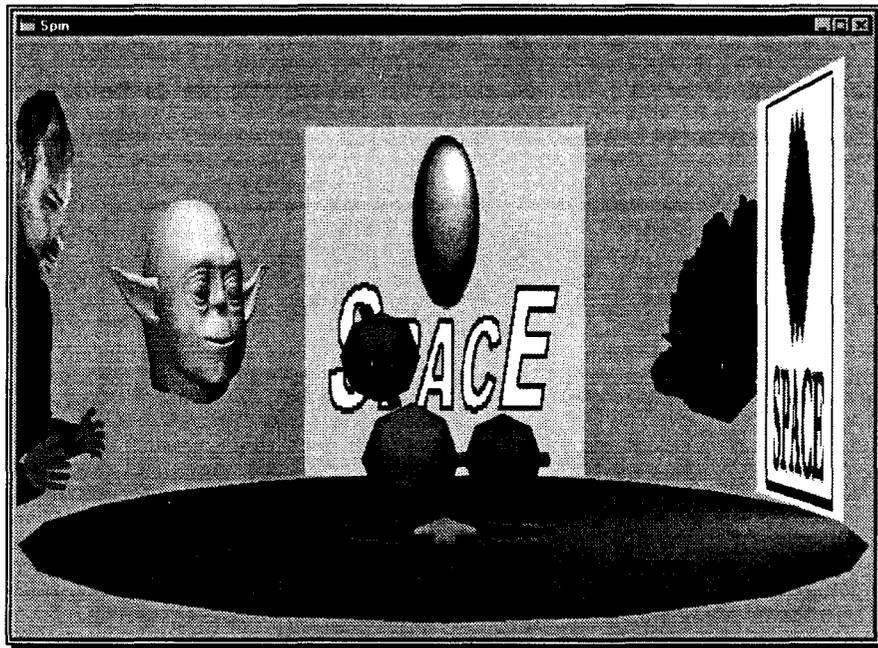


Figure 84 : le clone reste visible grâce à l'aplatissement latéral

3.4.9 Regard

L'angle des yeux avec l'axe de la tête étant limité, nous avons également fait tourner la tête du clone pour le faire regarder là où il fallait. D'après [Hideaki92], l'angle de la tête avec le corps et celui des yeux avec la tête sont proportionnels. Nous avons fait en sorte que nos clones tournent la tête et les yeux d'un même angle (Figure 85). C'est visuellement réaliste. Il reste quand même parfois difficile de savoir exactement quel est le centre d'intérêt d'un clone (certaines positions ne sont pas très « lisibles »). Il est possible, empiriquement, d'améliorer la conscience du regard en jouant sur le rapport entre les deux angles ou en exagérant l'angle de vue dans certaines positions, mais il est techniquement difficile de restituer la subtilité d'un regard humain réel (taille des yeux à l'écran, différences légères dans l'orientation des deux yeux, accommodation proche/lointaine...)

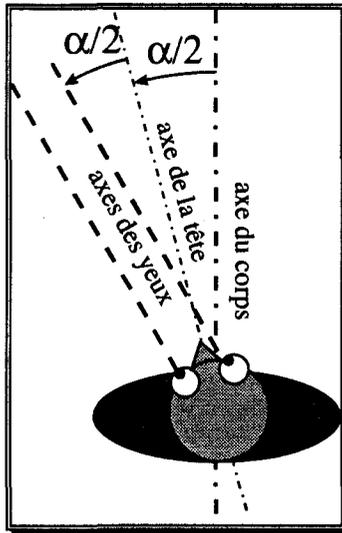


Figure 85 : rotation de la tête et des yeux

Si nous employons maintenant les clones réalistes de l'INA, les premières expériences sur le regard ont été faites sur des avatars caricaturaux [Saugis97a]. L'ajout d'yeux mobiles a été d'autant plus remarquable qu'il a insufflé un caractère presque vivant à ces masques figés.



Figure 86 : un clone expérimental

3.4.10 Télépointeurs

Dans Spln, les télépointeurs sont des flèches courbes dont la pointe est placée sur l'objet montré, à l'endroit précis qu'a sélectionné l'utilisateur distant (Figure 87). Nous avons utilisé des splines pour représenter le lien entre clone et objet, ainsi que pour décrire la trajectoire de la pointe du télépointeur d'une cible à une autre. Lorsque l'objet sélectionné change, le télépointeur se déplace en un

mouvement souple vers la nouvelle cible. La transition se fait sans rupture dans la continuité de la scène.

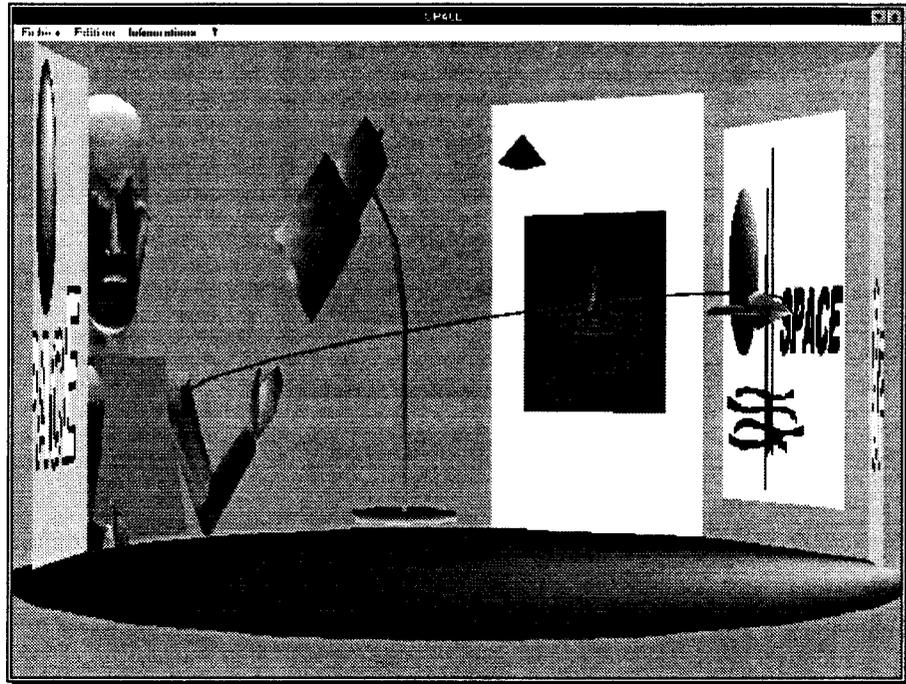


Figure 87 : un télépointeur

L'inconvénient est que l'espace de travail intérieur (le dessus de la table) est occupé par les « fils » des télépointeurs, c'est pourquoi nous travaillons actuellement sur l'utilisation des gestes des clones (mouvements de bras) pour la désignation, éventuellement couplés à un curseur placé sur l'objet concerné. Le « fil » liant le clone à l'objet désigné, redondant avec le geste du bras, est alors supprimé (c'est le cas dans les illustrations de la section 0) ou, à la rigueur, être rendu transparent et donc prendre moins de place, sémantiquement parlant, dans la scène. Le pointeur local et les télépointeurs ont des formes différentes, ce qui évite de les confondre.

3.5 Le prototype actuel de SpIn

Pour terminer la présentation de SPACE et de SpIn, nous allons illustrer quelques situations pouvant se présenter par des captures d'écran de notre prototype.

Deux personnes, Cédric et Grégory, vont se réunir pour discuter du choix d'un logo pour le projet SPACE (cette situation a fait l'objet de deux vidéos, l'une en langue anglaise [Saugis97b], l'autre en français [Saugis97c]). Cédric, qui arrive le premier en réunion, ne voit au départ que les documents disponibles :

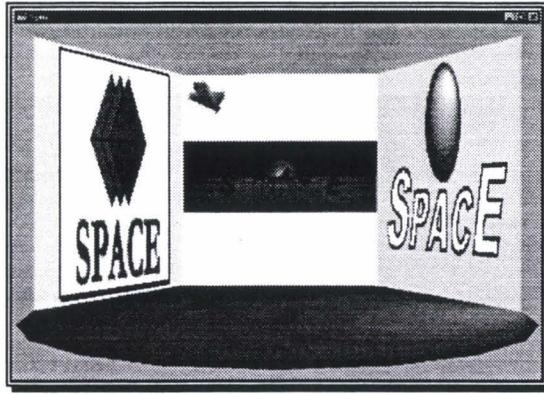


Figure 88 : la table et le bandeau contenant des documents

Grégory arrive dans la salle :

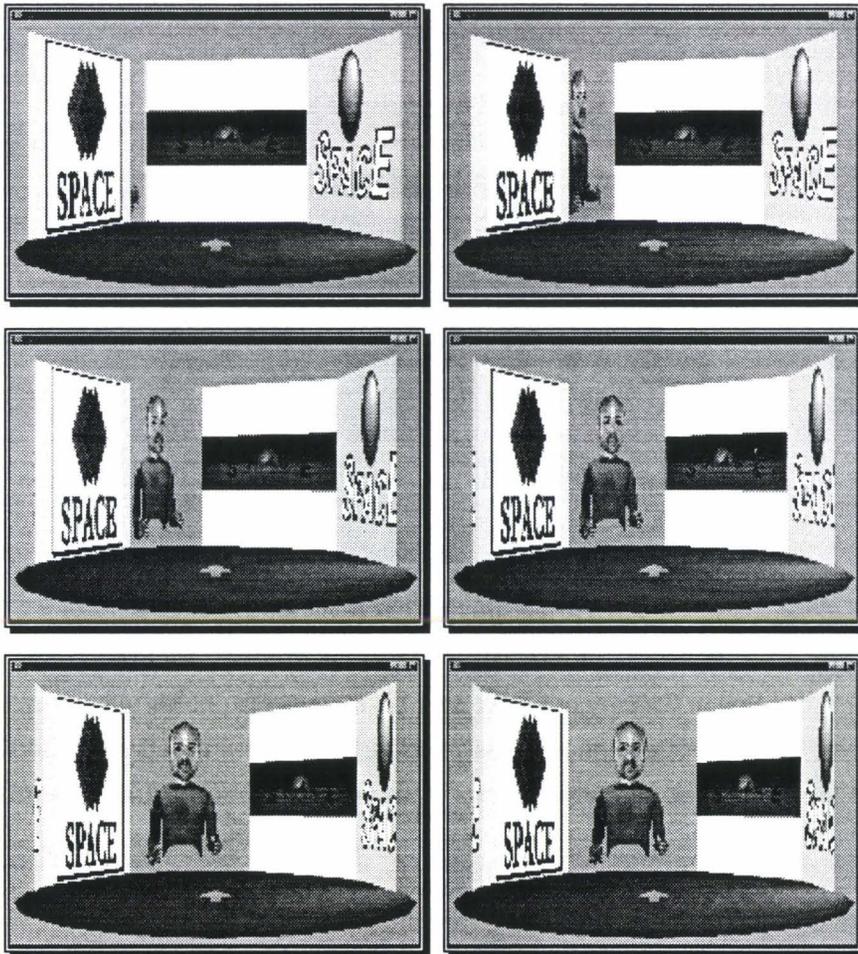


Figure 89 : Grégory prend place autour de la table

Cédric : « Bonjour Grégory ! »

Cédric désigne le document qu'il est en train de regarder à l'aide de son pointeur 3D :

Cédric : « Ce logo est trop simple, pas assez évocateur. »

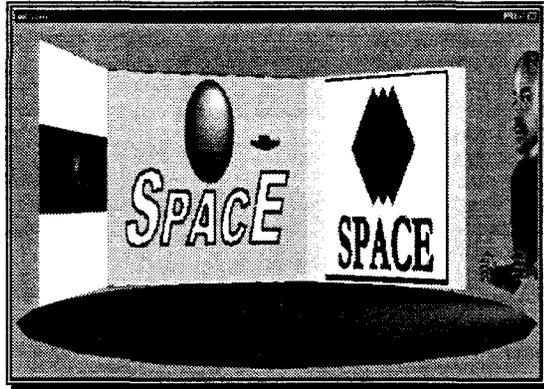


Figure 90 : Cédric montre un logo à Grégory

Pour Grégory, le clone de Cédric regarde et désigne également ce document :

Grégory : « Oui, je suis d'accord. »

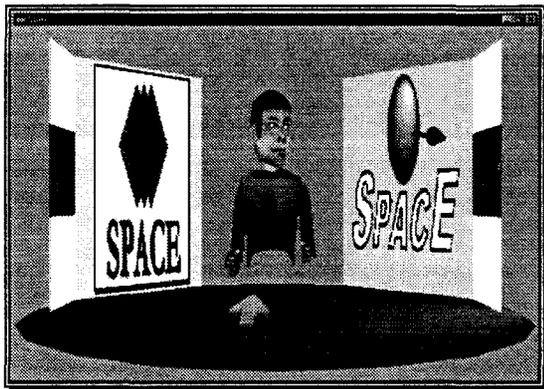


Figure 91 : le logo désigné par le clone de Cédric

Les documents que voient Grégory et Cédric sont les mêmes (ils ont les mêmes droits), par contre leur placement est décidé localement. C'est le principe de la vue subjective. Si Cédric avait été le possesseur des logos, il aurait pu choisir de les montrer au fur et à mesure qu'il les commentait à Grégory.

Cédric montre un autre document.

Cédric : « Je préfère de loin celui-ci ! »

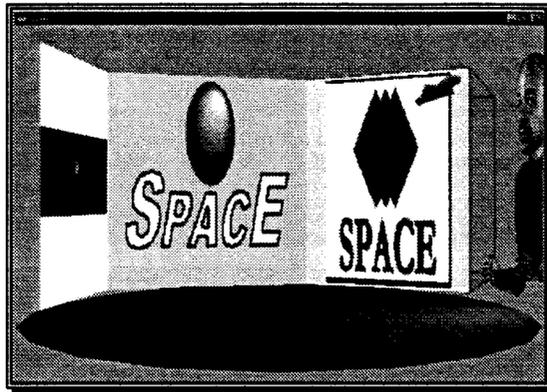


Figure 92 : pointage sur un document

Le télépointeur de Cédric désigne ce logo à Grégory.

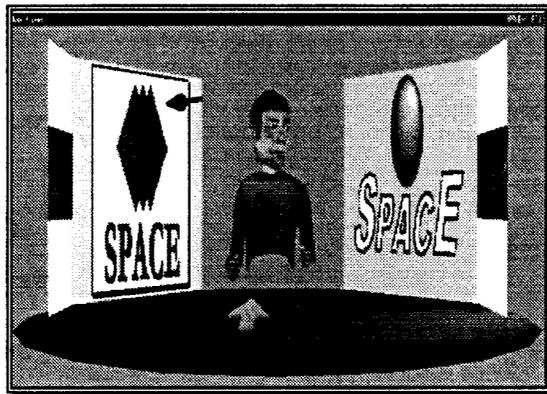


Figure 93 : le télépointeur de Cédric a bougé

Grégory tourne la tête vers ce document dont il est maintenant question. Il fait tourner le bandeau pour être face à l'objet.

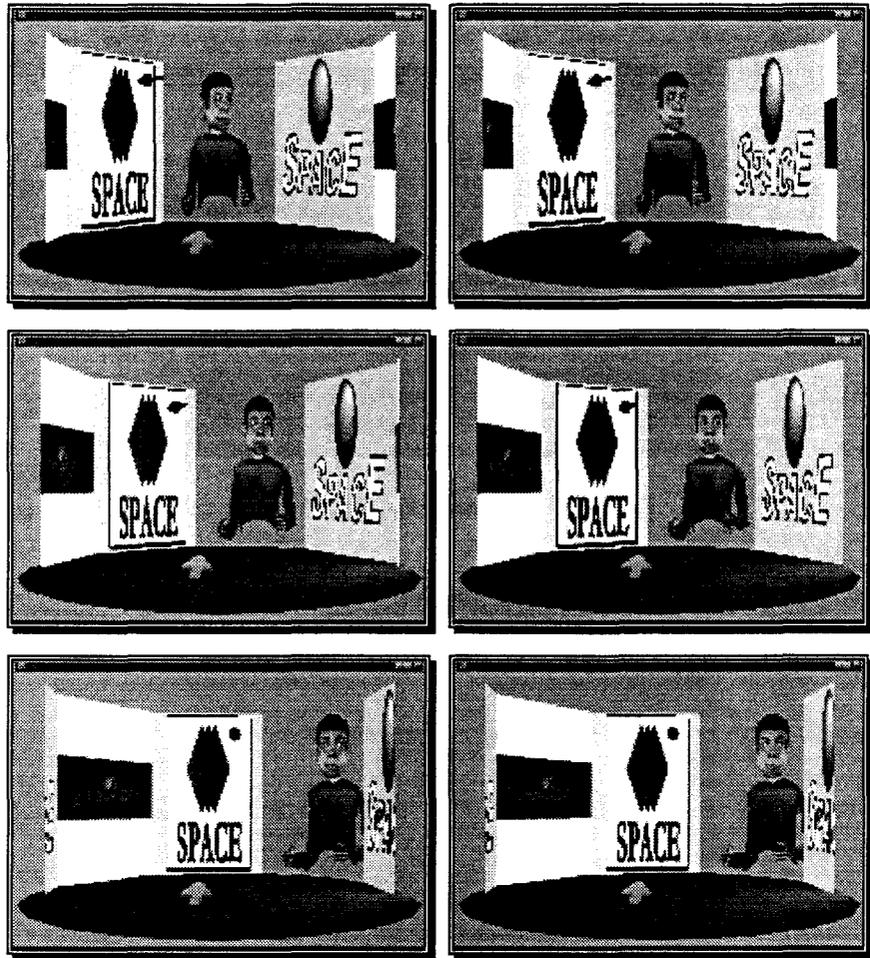


Figure 94 : Grégory porte son regard sur le logo en question

Dans la vue de Cédric, le clone de Grégory va regarder ce même logo.

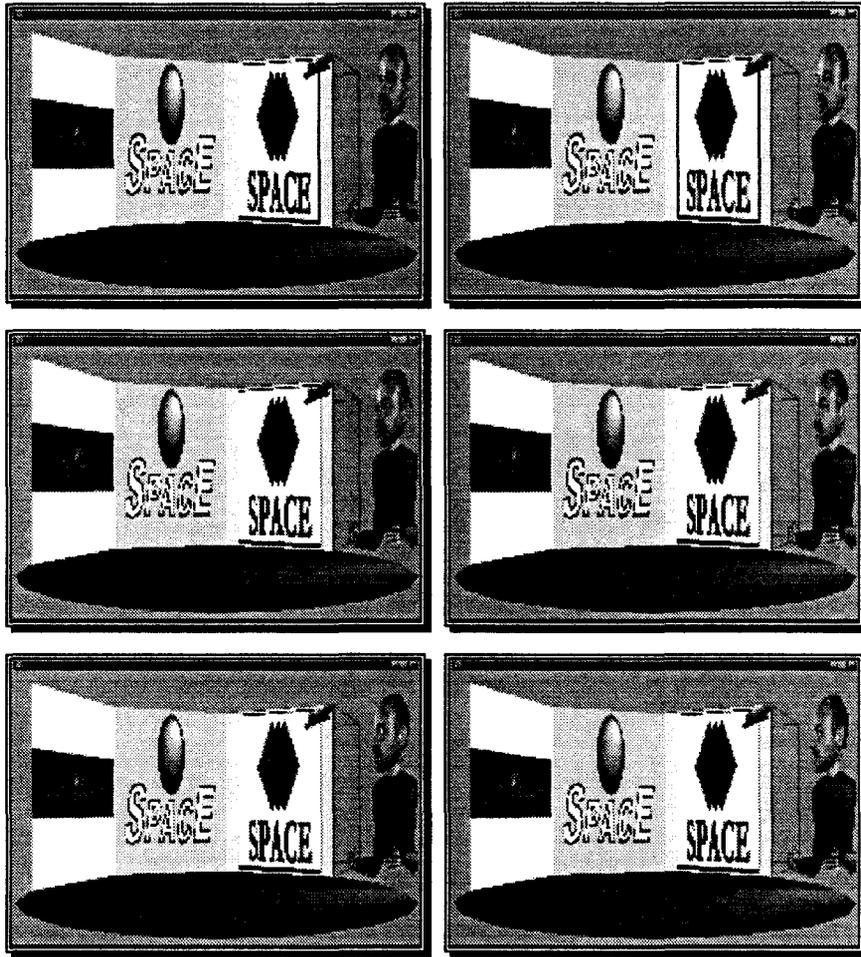


Figure 95 : Cédric voit Grégory regarder l'autre logo

Grégory regarde avec attention le logo proposé. Il utilise la distorsion.

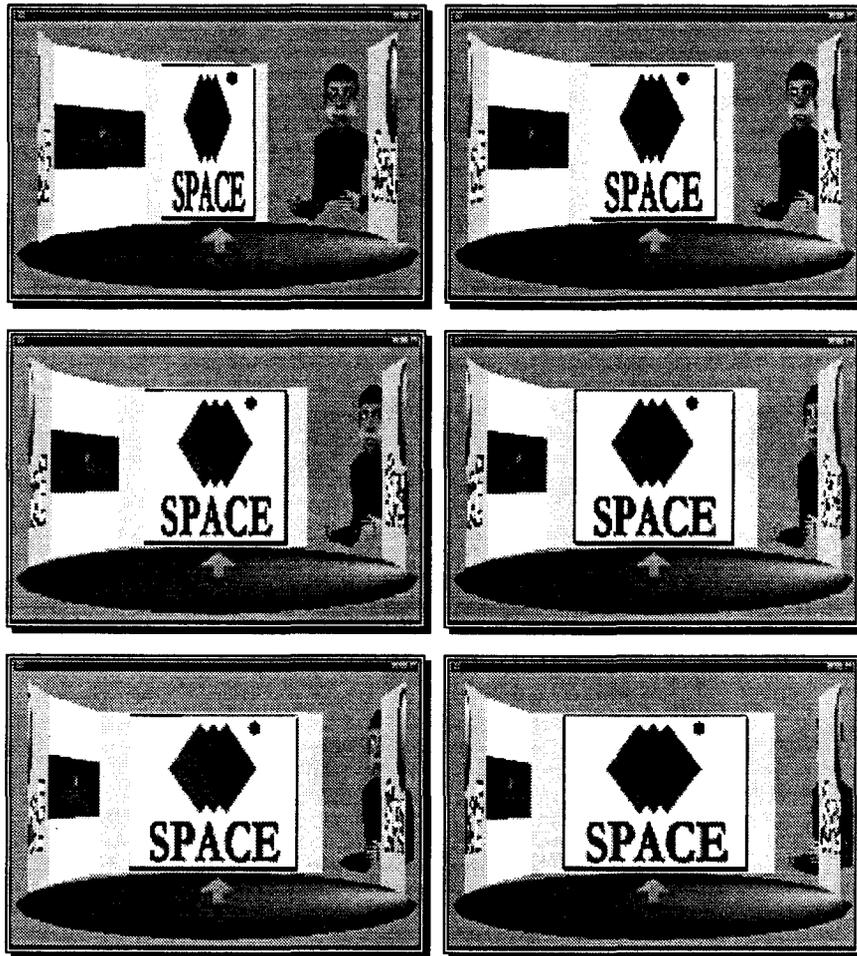


Figure 96 : la distorsion agrandit le centre d'intérêt

Grégory : « Hmm... en effet il est très bien. »

Cédric : « OK, on garde celui-ci alors. »

Les échanges entre Cédric et Grégory utilisent les possibilités de notre interface : regarder vers un objet, le grossir comme si on s'en approchait pour le voir plus détaillé, voir son interlocuteur sous forme de clone, savoir ce qu'il regarde et où il pointe. Chacun a un point de vue, une organisation de la scène, une vue sur l'activité subjectifs. Les composantes sont les mêmes mais leur placement est libre.

3.6 Travaux en cours

Le modèle que nous avons défini doit être complété et étendu pour s'adapter au plus grand nombre possible de situations. Nous avons entrepris l'exploration de plusieurs pistes.

Le développement de la manipulation 3D dans Spln est en voie d'achèvement. Une architecture logicielle permettant de développer des applications par programmation événementielle existe. La Figure 97 présente un cas de déplacement d'un atome dans une molécule.

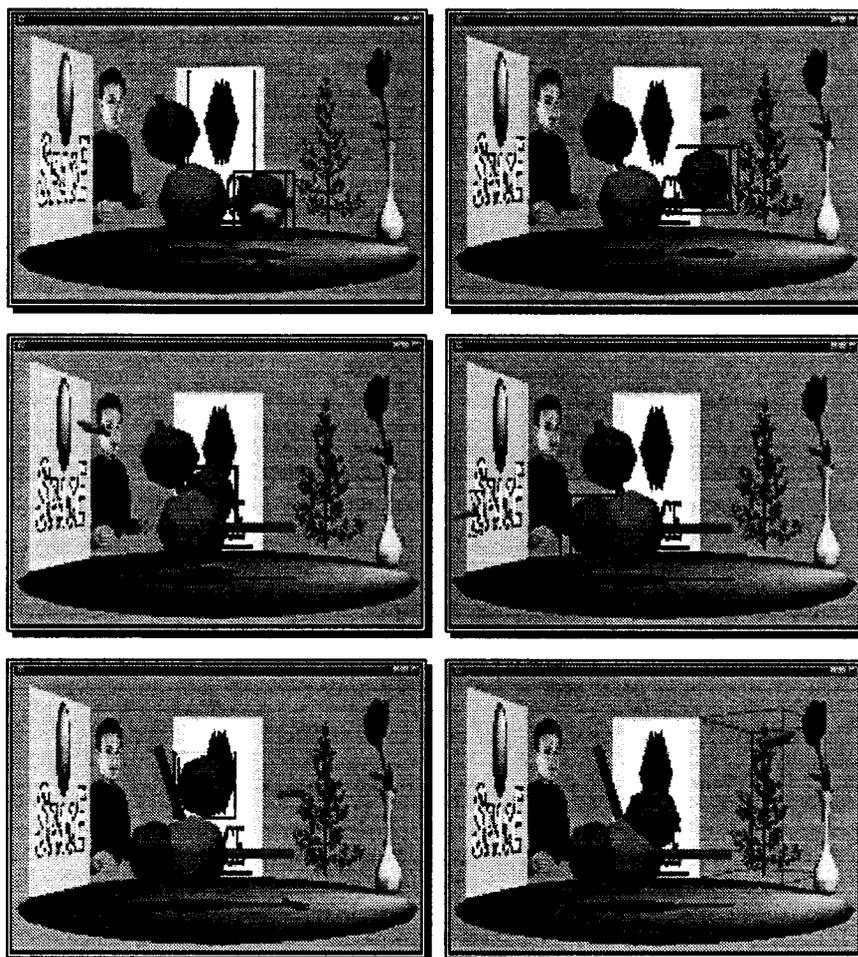


Figure 97 : manipulation 3D d'une molécule

Un document a un caractère privé (visible seulement par son propriétaire), public (accessible par tous) ou éventuellement partagé par certains utilisateurs particuliers du groupe. Cela pose deux problèmes dans l'interface. La première est de représenter l'état des documents. Comment faire la distinction entre un document public ou privé ? Plus difficile encore, comment savoir qui connaît un document partiellement partagé ? La Figure 98 montre un clone dont le propriétaire tourne son regard d'un objet public (le logo de droite) vers un objet privé, invisible pour l'observateur. Le clone présenté baisse donc la tête vers l'espace qui est devant lui.

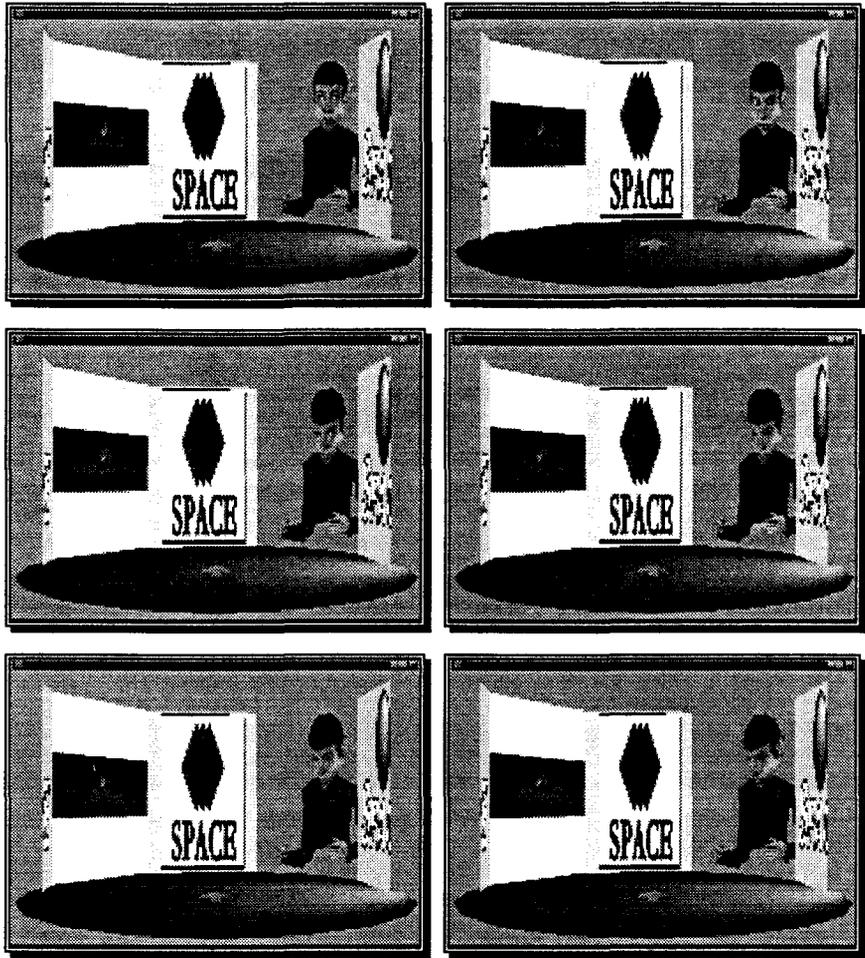


Figure 98 : le clone regarde dans son espace privé

4. Conclusion et perspectives

4.1 Conclusion

Nous avons décrit comment Spln organise la présentation d'une activité synchrone de groupe. Pour situer ses capacités par rapport à d'autres projets, nous allons reprendre la classification et les critères proposés par Steve Benford [Benford95] :

- présence – les utilisateurs distants sont représentés sous forme de clones 3D réalistes. Seule le haut du corps est visible, ce qui ne nuit pas dans le contexte que nous nous sommes fixé ;
- localisation – en grande partie contrainte (autour de la table), le placement par l'utilisateur reste libre. Des contraintes supplémentaires (comme « être à côté de ») peuvent éventuellement être demandées par ScenIC ;
- identité – sans problème grâce aux clones réalistes. Dans le futur, le son spatialisé asservi à la scène visuelle permettra de situer les interlocuteurs dans l'espace. Cela renforcera la présentation visuelle et facilitera encore l'identification ;
- point de vue – déterminé grâce à la navigation (« se tourner vers », « s'approcher de »), il est rendu par les clones (mouvement des yeux et de la tête) ;

- point d'action – visible très précisément grâce aux télépointeurs ;
- activité et disponibilité – le système de manipulation étant en cours de réalisation, cette rubrique est pour l'instant sans objet ;
- gestes – pas encore pris en compte. Dans le futur, les clones se perfectionneront, leurs bras seront animés et reproduiront les gestes de leur maître ;
- expressions faciales – elles exploitent le moteur d'analyse/synthèse de l'INA. Pour l'instant (puissance de calcul, problèmes d'adaptation pour PC), l'acquisition vidéo et l'analyse sont faites sur une machine SGI Indy à part ;
- degré de présence – fort grâce à l'emploi la construction d'une scène *continue* homogène et de clones réalistes animés ;
- confiance – le modèle de clonage étant adaptable à n'importe quel modèle 3D, rien n'empêche de se forger une image falsifiée, ou d'adopter une apparence « bande dessinée ». Ce problème concerne plutôt la gestion de l'activité et des participants ;
- efficacité – à la portée d'un PC récent équipé d'une carte accélératrice OpenGL. La puissance requise par l'analyse des expressions faciales est pour l'instant un problème mais n'est pas hors de portée.

La proposition que nous avons faite est avant tout exploratoire. Nous sommes partis d'une analyse des contraintes du TCAO pour voir ce que la synthèse d'images 3D pouvait apporter comme solutions.

Le concept de « tout visible » est sans équivalent. Sa réalisation est exigeante et a conduit à la perte d'une cohérence spatiale globale, commune à tous les utilisateurs. Par contre, la cohérence du contenu, à un niveau plus élevé, est préservée : documents, personnes, actions sont perçus par tous.

Le bandeau actuel convient pour des sessions en petit comité, dans lesquelles il n'accueille pas plus d'une dizaine d'objets. Il organise le placement des objets tout en facilitant la navigation. Les possibilités de densification de l'information à l'écran sont néanmoins limitées et Spln devra forcément faire des concessions au principe du tout visible pour s'adapter aux grands groupes. Par exemple, les nombreux auditeurs d'un conférencier seront rassemblés sur une unique vue d'ensemble du public.

Si nous avons largement simplifié les interactions réalistes, notamment la navigation, le contrôle des indices visuels traduisant l'activité (les yeux des clones et surtout les télépointeurs) constituent une « augmentation » de la réalité.

Nos propositions de simplification de la réalité, de description de haut niveau permettant une reconstruction synthétique subjective, de navigation et d'organisation contraintes autour de la métaphore de table de réunion sont plus poussées que dans d'autres projets, comme Desktop-MAJIC [Tanaka96] ou encore celui d'AT&T par exemple [Seligmann95]. Nous avons abouti à une présentation qui prend en compte les capacités d'un ordinateur de bureau multimédia, la facilité d'utilisation (navigation à deux degrés de liberté, manipulation directe et intuitive), le cadre de travail du TCAO

synchrone, les standards réseau (Internet, VRML...), la perception humaine (table familière, clones à l'apparence humaine...).

La vision subjective et la « virtualisation » qui lui est liée (simplification et augmentation de la réalité) sont très prometteuses. Elles doivent être réalisées avec soin pour ne pas perdre d'information, pour être facilement compréhensibles, pour être bénéfiques à l'activité en apportant de nouveaux renseignements et pour tirer parti des possibilités des systèmes informatiques. Le passage d'une situation réelle à une communication médiatisée, à travers un poste de travail informatique, se fait forcément avec des concessions. L'utilisateur les acceptera d'autant plus facilement qu'elles respecteront ses habitudes, ses attentes et sa compréhension. Les recommandations que nous avons énoncées sont des références permettant d'évaluer a priori la qualité d'un outil coopératif de groupe.

4.2 Perspectives

Dans cette thèse nous avons présenté les concepts de base d'un mécanisme de présentation et d'interaction pour le travail coopératif synchrone.

Nous en sommes maintenant arrivés à développer les concepts relatifs à la place des objets dans l'interface. En particulier, nous devons préciser quel usage sera fait de la table. La table est un espace de travail, de manipulation bien plus pratique que le bandeau : l'objet que l'on y place est plus grand, ne subit aucune déformation. Nous pensons qu'un seul objet devrait être posé sur la table. Nous espérons beaucoup de l'animation 3D pour offrir des métaphores visuelles compréhensibles dans ce genre de situations. Le passage d'un objet de la table au bandeau, et réciproquement, doit se faire naturellement. Là encore, l'interaction et l'animation doivent se compléter. Nous travaillons à la réalisation de méthodes pour afficher des objets temporaires (par exemple un sélecteur de fichier) sur la table, en plus de l'objet qui y serait déjà présent.

La Figure 99 montre une session à laquelle participent cinq personnes. Trois documents sont visibles, dont l'un occupe la table. L'organisation spatiale que nous avons définie pour SpIn supporte bien un tel groupe limité. Si en théorie la capacité du bandeau est infinie, en pratique la qualité se dégrade quand de nombreux objets sont présents. La limite du nombre d'objets que le bandeau peut contenir tout en restant exploitable doit être évaluée précisément mais elle ne dépasse probablement pas la dizaine, voire moins. Des méthodes spécifiques de composition de la scène doivent être mises au point pour gérer les grands groupes (comme dans ARENA [Benford97]) ou les quantités importantes de documents. La Figure 100 présente une solution possible : une étagère accueille plusieurs objets. Le choix des objets qui y sont rangés est soit laissé à l'appréciation de l'utilisateur (notamment pour ses documents privés) soit automatique en fonction des informations fournies par ScenIC.

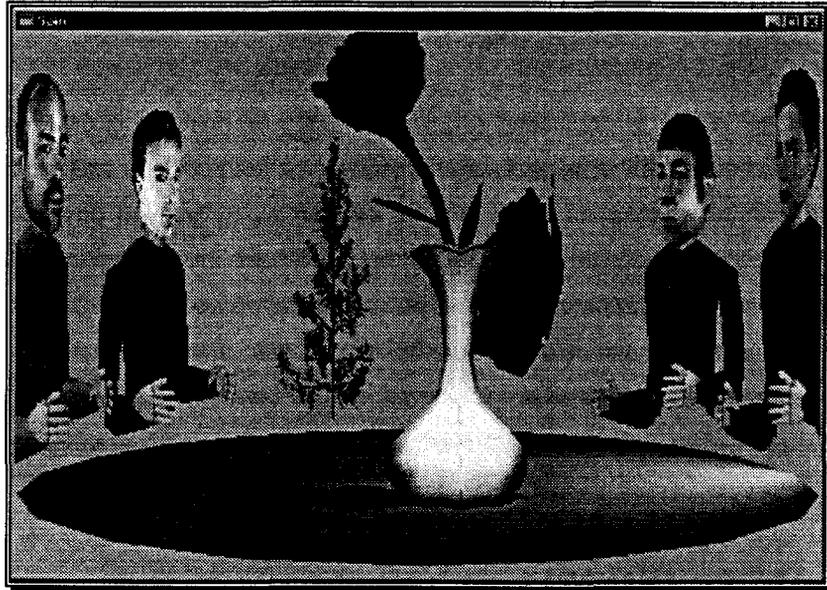


Figure 99 : une réunion à cinq

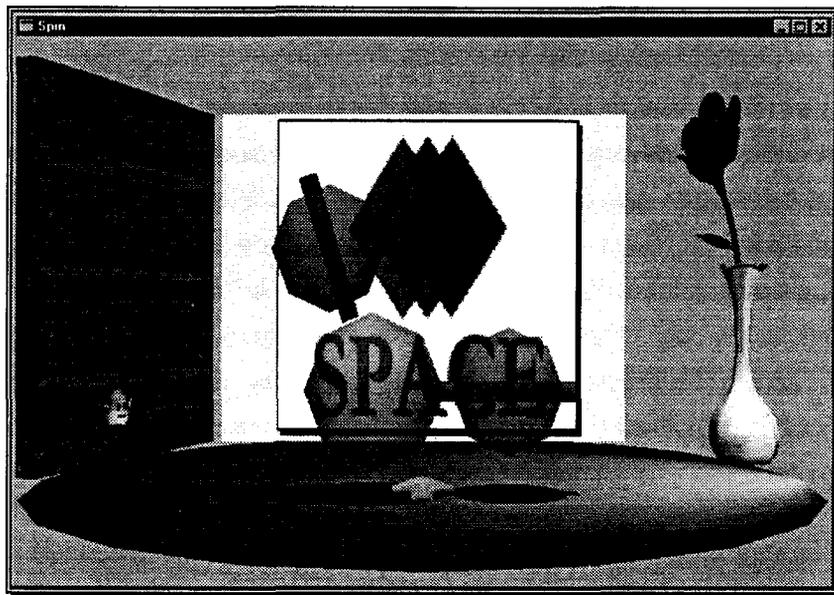


Figure 100 : une armoire accueille les objets « inactifs »

La table est un espace de travail bien plus exploitable que le bandeau. Nous la destinons à recevoir un objet à la fois, celui qui sera manipulé par l'utilisateur. L'inconvénient est que l'objet ramené au centre masque les objets du bandeau qui se trouvent derrière lui. L'utilisation de la transparence (Figure 101) est une solution simple à ce problème. La transparence est une technique graphique maîtrisée, disponible dans la bibliothèque OpenGL. En allant plus loin, la transparence peut être mise à profit pour densifier l'information et augmenter le nombre d'objets visibles.



Figure 101 : l'objet sur la table est transparent

On peut également exploiter la transparence pour faciliter l'interaction, comme c'est le cas avec le Silk Cursor [Zhai94]. Subordonnée à la description de la scène fournie par SceneC, elle peut rendre le degré de présence ou d'importance d'un objet.

Nous avons développé en premier lieu l'interaction 3D au sein de Spln. Pour poursuivre l'orientation bureautique que nous donnons à notre interface, il faudra fatalement étudier comment il est possible de manipuler des documents 2D. En effet, si beaucoup d'applications peuvent bénéficier du passage à la 3D (représentation de données, classement, etc.), certains documents sont intrinsèquement 2D (textes, images, tableaux...). Nous devons trouver les bonnes méthodes de travail sur ces documents à l'aide des périphériques utilisés dans notre proposition.

Un important travail de développement doit être réalisé sur les clones. Nous sommes partis de clones très simples que nous avons complétés au fur et à mesure : les yeux et tête sont mobiles, le visage est réaliste. L'animation faciale temps réelle est également présente mais nécessite des ressources importantes (une machine dédiée à l'analyse). Nous comptons travailler sur la gestuelle des clones, dans les deux modes (désignation et expressivité) et sur le passage de l'un à l'autre de ces modes. Les gestes d'expression qui accompagnent la parole sont locaux à la personne qui parle, tandis que la désignation concerne la scène entière. Nous devons les traiter séparément tout en étant capables de passer de l'un à l'autre. D'autres raffinements peuvent encore être imaginés pour les clones. Nous voulons évaluer précisément l'apport de chacun de ces éléments. Cela explique que nous n'utilisons pas encore l'animation faciale, bien qu'elle soit implémentée : son importance pour Spln ne sera évaluable que quand l'interface sera plus complète et utilisable au minimum sur une application donnée. Notre but est de rendre les clones vivants, porteurs d'informations. Nous avons vu que notre mécanisme d'yeux, pourtant simple, modifiait la perception que l'on avait des personnes distantes à partir du moment où elles s'animaient. En fin de compte, nous recherchons une

représentation efficace des êtres humains, adaptée à notre outil bureautique. Si par exemple nous déterminons qu'une animation automatique très simple de la bouche est bien perçue par les utilisateurs, nous saurons qu'il n'est pas nécessaire de passer par une analyse du visage visant à rendre les mouvements réels.

Quand nous disposerons d'une version pleinement utilisable de notre interface, nous développerons un « *browser Spln* », exploitant le format VRML, qui viendra se greffer sur des outils de communication répandus comme Netscape Navigator.

Pour présenter et évaluer tout le projet SPACE, un démonstrateur est en cours de développement. Au point de vue matériel, nous envisageons de brancher trois stations utilisateur sur un réseau local, de même qu'un serveur d'activité. Une application coopérative spécifique, que nous définissons avec le laboratoire Trigone, sera développée. Les utilisateurs devront communiquer et travailler ensemble pour mener leur tâche à bien.

Références bibliographiques

- [Barme96] Laurent Barme, « **Le son dans les systèmes informatiques pour le travail coopératif** », *Thèse de Doctorat de l'Université de Lille I*, 30 avril 1996
- [Bauml97] Betty J. Bauml, « **Dictionary of worldwide gestures** ». 2nd ed., Lanham, Md. : Scarecrow Press, 1997
- [Beaudouin-Lafon96] Michel Beaudouin-Lafon, « **Les habits neufs du travail en équipe** », spécial *La Recherche*, No. 285, pp. 48-52, mars 1996
- [Benford93a] S. D. Benford, A. N. Bullock, N. Cook, P. W. Harvey, R. Ingram, O. Lee, « **A spatial model of cooperation for virtual worlds** », *Proceedings of Informatique '93 - Interfaces to Real and Virtual Worlds*, mars 1993
- [Benford93b] S. Benford, A. Bullock, N. Cook, P. Harvey, R. Ingram, Ok-Ki Lee, « **A model of conversation management in virtual rooms** », *Telepresence 93, Applica 93*, 22-25 mars 1993, Lille, France
- [Benford93c] Steve Benford, Lennart Fahlén, « **A spatial model of interaction in large virtual environments** », *Third European Conference on Computer Supported Cooperative Work ECSCW'93*, Milan, septembre 1993.
- [Benford95] Steve Benford, Chris Greenhalgh, John Bowers, Dave Snowdon et Lennart E. Fahlén, « **User Embodiment in Collaborative Virtual Environments** », *Proceedings of CHI'95*, pp. 242-248, 7-11 mai 1995, Denver, Colorado, USA
- [Benford97] Steve Benford, Chris Greenhalgh, David Lloyd, « **Crowded collaborative virtual environments** », *CHI 97*, 22-27 mars 1997, Atlanta, Géorgie, USA
- [Bier93] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, Tony D. DeRose, « **Toolglass and Magic Lenses: the see-through interface** », *Computer Graphics proceedings*, pp. 73-80, ACM, 1993
- [Blinn82] James F. Blinn, « **A generalization of algebraic surface drawing** », *ACM Transactions on Graphics*, Vol. 1, No. 3, pp. 235-256, juillet 1982
- [Bowers96] John Bowers, James Pycok, Jon O'Brien, « **Talk and embodiment in collaborative virtual environments** », *CHI 96*, pp. 58-65, 13-18 avril 1996, Vancouver, Canada

- [Bregler97] Christoph Bregler, Michele Covell, Malcolm Slaney, « **Video rewrite: driving visual speech with audio** », *Proceedings of SIGGRAPH '97*
- [Buxton92] William A. S. Buxton, « **Telepresence: integrating shared task and person spaces** », *Proceedings of Graphics Interface '92*, pp. 123-129, Morgan Kaufmann.
- [Cadoz94] Claude Cadoz, « **Le geste canal de communication homme/machine – la communication instrumentale** », *Technique et science informatiques*, Volume 13, n°1/1994, pp. 31-61, 1994
- [Cadoz96] Claude Cadoz, « **Réintroduire les sensations physiques** », spécial *La Recherche*, No. 285, pp. 80-84, mars 1996
- [Capin97] Tolga K. Capin, Hansrudi Noser, Daniel Thalmann, Igor Sunday Pandzic, Nadia Magnenat Thalmann, « **Virtual human representation and communication in VLNet** », *IEEE Computer Graphics and Applications*, Vol. 17, No. 2, mars-avril 1997
- [Cassel94] Justine Cassel, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost et Matthew Stone, « **Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation, for Multiple Conversational Agents** », *Proceedings Siggraph'94*, pp. 413-419, Orlando, FL, USA, 24-29 juillet 1994
- [Chaillou94] C. Chaillou, L. Wibaux, L. Barthe, « **Une expérience d'interface 3D exploitant la notion de pièce** », *Actes des Sixièmes Journées sur l'Ingénierie des Interfaces Hommes-Machines IHM'94*, Lille, 8-9 décembre 1994
- [Cohen93] M. Cohen. « **Throwing, Pitching and Catching Sound : Audio Windowing Models and Modes** » dans *Int. Journal of Man-Machines Studies*, Academic Press, No.39, 269-304, 1993.
- [Coninx97] K. Coninx, F. Van Reeth, E. Flerackers, « **A man-machine interface for immersive geometric modeling** », *XIV Imeko World Congress, ISMCR'97 Topical Workshop on Virtual Reality and Advanced Man-Machine Interfaces*, pp. 73-78, 01 - 06 June 1997, Tampere, Finland
- [Conversy97] Stéphane Conversy, « **Sons d'environnement naturel pour le suivi d'activités d'arrière-plan** », *IHM 97*, pp. 191-197, 9^{èmes} journées sur l'Ingénierie de l'Interaction Homme-Machine, Poitiers Futuroscope, France, 10-12 septembre 1997
- [Croisy95] Pascal Croisy, « **Collecticiel temps réel et apprentissage coopératif : des aspects sociaux et pédagogiques jusqu'au modèle multi-agent de l'interface de groupe** », *Thèse de Doctorat*, Université des Sciences et Technologies de Lille, 1995
- [Derycke93] A.C. Derycke, C. Chaillou, L. Dumesges, « **La troisième dimension dans les interfaces de collecticiels** », *Document de travail des Cinquièmes Journées sur l'Ingénierie des Interfaces Hommes-Machines IHM 93*, pp. 81-86, Lyon, 19-20 octobre 1993
- [Doenges97] Peter K. Doenges, Tolga K. Capin, Fabio Lavagetto, Joern Ostermann, Igor S. Pandzic, Eric D. Petajan, « **MPEG-4: Audio/video & synthetic graphics/audio for mixed media** », *Image Comm. J.*, Elsevier, Amsterdam, 1997 ; http://drogo.cselt.stet.it/ufv/leonardo/icjfiles/mpeg-4_si/paper7.htm
- [Emslie-Smith89] D. Emslie-Smith, C. R. Patterson, T. Scratcherd, N. W. Read, « **Textbook of physiology** », 11^{ème} édition, *Churchill Livingstone*, Londres, 1989
- [Fahlén93] L. E. Fahlén, C. G. Brown, O. Ståhl, C. Carlsson, « **A space based model for user interaction in shared synthetic environments** », pp. 43-48, *ACM InterCHI'93*, 24-29 April 1993
- [Feyereisen91] Pierre Feyereisen, « **Gestures and speech** », *Cambridge [England] University Press*, 1991

- [Fish92] Robert S. Fish, Robert E. Kraut, Robert W. Root, Ronald E. Rice, « **Evaluating video as a technology for informal communication** », *Proceedings of the Conference on Computer Human Interaction (CHI) '92*, Monterey, CA, mai 1992, pp. 37-48
- [Foley90] James D. Foley, Andries van Dam, Steven K. Feiner, John F. Hughes, « **Computer graphics, principles and practice** », Second edition, Addison-Wesley, 1990
- [Furnas86] George W. Furnas « **Generalized fisheye views** », *Human Factors in Computing Systems CHI '86 Conference Proceedings*, Boston, April 13-17, 1986, 16-23
- [Gale90] Stephen Gale, « **Human aspects of interactive multimedia communication** », *Interacting with Computers*, Vol. 2, No. 2, 1990, pp. 175-189
- [Gettys96] James Gettys, Robert W. Scheifler, « **Xlib – C language X interface** », X Consortium Standard, X version 11, Release 6.3
- [Gibson66] J. J. Gibson, « **The senses considered as perceptual systems** », *Boston: Houghton Mifflin*
- [Gibson79] J. J. Gibson, « **The ecological approach to visual perception** », Boston: Houghton Mifflin, 1979
- [Goral84] C. M. Goral, K. E. Torrance, D. P. Greenberg, B. Battaile, « **Modelling the interaction of light between diffuse surfaces** », *Computer Graphics (SIGGRAPH '84 Proceedings)* 18:3, juillet 1984, pp. 212-222
- [Greenhalgh94] Chris Greenhalgh, « **An experimental implementation of the spatial model** », *Proc. 6th ERCIM workshops*, Swedish Institute of Computer Science (SICS), Stockholm, 1-3 juin 1994, Suède
- [Greenhalgh95] Chris Greenhalgh, Steve Benford, « **MASSIVE: a collaborative virtual environment for tele-conferencing** », *ACM Transactions on Computer Human interfaces (TOCHI)*, Volume 2, Numéro 3, pp. 239-261, ISSN 1073-0516, ACM Press, septembre 1995
- [Harrison95] Beverly L. Harrison, Hiroshi Ishii, Kim J. Vicente, William A. S. Buxton, « **Transparent layered user interfaces: an evaluation of display design to enhance focused and divided attention** », *Proceeding of CHI'95*, ACM Press, 317-324, mai 1995.
- [Henderson86] D. Austin Henderson, Jr., S. K. Card, « **Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface** », *ACM Transactions on Graphics*, Vol. 5, No. 3, July 1986, pp. 211-243
- [Hoogstoel95] Frédéric Hoogstoel, « **Une approche organisationnelle du travail coopératif assisté par ordinateur—Application au projet Co-Learn** », *Thèse de doctorat*, Université des Sciences et Technologies de Lille, 1995
- [Hoogstoel97] Frédéric Hoogstoel, Patricia Plénacoste, Jean-Claude-Tarby, « **ODESCA/SCENIC/SPIN : Rôles et fonctions** », document de travail, laboratoire Trigone
- [Ishii91] Hiroshi Ishii, Kazuho Arita, « **ClearFace: translucent multiuser interface for TeamWorkStation** », *Proceedings of the Second European Conference on Computer-Supported Cooperative Work*, pp. 163-174, Bannon, L., Robinson, M. & Schmidt, K. (Editors), 25-27 septembre 1991, Amsterdam, Pays-Bas.
- [Ishii92] Hiroshi Ishii, Minoru Kobayashi, « **ClearBoard: a seamless medium for shared drawing and conversation with eye contact** », *ACM SIGCHI'92*, pp. 525-532, 3-7 mai 1992
- [Ishii94] H. Ishii, M. Kobayashi, K. Arita, « **Iterative Design of Seamless Collaboration Media** », *Communications of the ACM (CACM)*, Special Issue on Internet Technology, ACM, Vol. 37, No. 8, août 1994, pp. 83-97
- [Kabbash94] Paul Kabbash, William Buxton et Abigail Sellen, « **Two-handed input in a compound task** », *Proceedings of CHI'94*

- [Kramer94] Axel Kramer, « **Translucent patches -Dissolving windows-** », *UIST'94*, pp. 121-130, 2-4 novembre 1994
- [Kuzuoka92] Hideaki Kuzuoka, « **Spatial workspace collaboration: a sharedview video support system for remote collaboration capability** », *CHI'92*, pp. 533-540, 3-7 mai 1992
- [Mackinlay91] J.D. Mackinlay, G.G. Robertson et S.K. Card, « **The perspective wall: detail and context smoothly integrated** », *Proceedings of CHI'91*, pp. 173-180
- [Mackinlay94] Jock D. Mackinlay, George G. Robertson, and Robert DeLine, « **Developing Calendar Visualizers for the Information Visualizer** », *UIST'94*
- [Morris77] Desmond Morris, « **Manwatching** », *New York : H. N. Abrams*, 1977
- [Neisser75] Ulric Neisser, Robert Becklen « **Selective looking: attending to visually specific events** », *Cognitive Psychology*, Vol.7, 1975, pp. 480-494
- [Nye92] Adrian Nye, « **Xlib programming manual, Volume one** », *O'Reilly & Associates Inc.*, 1992
- [Okada94] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa et Y. Matsushita « **Multiparty videoconferencing at virtual social distance: MAJIC design** », *ACM 94 Conference on CSCW*, pp. 385-393
- [Ochsman74] Robert B. Ochsman, Alphonse Chapanis, « **The effects of 10 communication modes on the behavior of teams during co-operative problem solving** », *International Journal of Man-Machine studies*, Vol. 6, 1974, pp. 579-619
- [Pang94] A. Pang, « **Spray rendering** », *IEEE Computer Graphics and Applications*, Vol. 14, No. 5, septembre 1994, pp. 57-63
- [Pang97] Alex Pang, Craig Wittenbrink, « **Collaborative 3D visualization with CSpray** », *IEEE Computer Graphics and Applications*, vol.17, No. 2, mars-avril 1997, pp. 32-41
- [Peters94] Sophie Peters Van Deirse, « **Devenir ingénieur en communication grâce à la visioconférence** », *L'Usine Nouvelle*, No. 2448, p. 126, mars 1994
- [PictureTel94] « **Qu'est-ce qu'un système de visioconférences ?** », *Série 101 d'introduction aux systèmes de visioconférences PictureTel*, PictureTel Corporation, 1994
- [Poupyrev97] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, Tadao Ichikawa, « **A framework and testbed for studying manipulation techniques for immersive VR** », *ACM VRST '97*, pp. 21-28, Lausanne, Suisse, 15-17 septembre 1997
- [Proesmans97] Marc Proesmans, Luc Van Gool, « **Reading between the lines - a method for extracting dynamic 3D with texture** », *ACM VRST '97*, pp. 95-102, Lausanne, Suisse, 15-17 septembre 1997
- [Quéau93] Quéau, P., « **Televirtuality: the merging of telecommunications and virtual reality** », *Computers and Graphics*, 17(6), 1993, 691-693.
- [Robertson91] G.G. Robertson, S.K. Card et J.D. Mackinlay, « **Cone Trees: animated 3D visualizations of hierarchical information** », *Proceedings of CHI'91*
- [Robertson93] G. G. Robertson, S. K. Card et J. D. Mackinlay « **Information visualization using 3D interactive animation** », *Communications of the ACM*, avril 1993, Vol. 36, No. 4
- [Sarkar92] M. Sarkar, M. Brown, « **Graphical fisheye views of graphs** », *Proceedings of CHI'92*, pp. 83-91, Monterey, Californie, 3-7 mai 1992
- [Saugis95] G. Saugis, C. Chaillou, « **Utilisation de techniques de synthèse d'images 3D pour la construction d'interfaces** », *3èmes journées de l'Association Française d'Informatique Graphique*, Marseille, 22-23-24 novembre 1995
- [Saugis97a] G. Saugis, C. Dumas, C. Chaillou, « **A new model of interface for synchronous CSCW** », *XIV Imeko World Congress, ISMCR'97 Topical Workshop on Virtual*

Reality and Advanced Man-Machine Interfaces, pp. 221-226, 01 - 06 June 1997, Tampere, Finland

- [Saugis97b] G. Saugis, C. Dumas, C. Chaillou, P. Carlier, P. Plénacoste, J.-C. Tarby, Vidéo « **SPACE: A new generation of human-computer interface for CSCW** », *Proceedings of ECSCW'97*, Lancaster, UK, 7-11 septembre 1997
- [Saugis97c] Grégory Saugis, Cédric Dumas, Christophe Chaillou, Patrice Carlier, Jean-Claude Tarby, Alain Derycke, « **SPACE une nouvelle génération d'interface homme-machine pour le TCAO** », vidéo *IHM 97*, pp. 211-212, 9^{èmes} journées sur l'Ingénierie de l'Interaction Homme-Machine, Poitiers Futuroscope, France, 10-12 septembre 1997
- [Scheiffler92] R. Scheiffler, J. Gettys, J. Flowers, D. Rosenthal, « **X Window System: The Complete Reference to Xlib** », X protocol, ICCCM, XFLD, X Version 11, Release 5. Digital Press, 1992
- [Schmidt91] Kjeld Schmidt, « **Riding a tiger, or Computer Supported Cooperative Work** », *ECSCW'91 Proceedings*, pp. 1-16, Amsterdam, 1991
- [Sederberg 86] Thomas W. Sederberg, Scott R. Parry, « **Free-form deformation of solid geometric models** », pp. 151-160, *Computer Graphics (SIGGRAPH '86 Proceedings)*, Vol. 20, Dallas, Texas, 18-22 août, 1986
- [Seligmann95] Dorée Duncan Seligmann, Rebecca T. Mercuri, John T. Edmark, « **Providing assurances in a multimedia interactive environment** », *Proceedings of CHI'95*, pp. 250-256, 7-11 mai 1995, Denver, Colorado, USA
- [Sheelagh95] M. Sheelagh T. Carpendale, David J. Cowperthwaite, F. David Fracchia, « **3-Dimensional pliable surfaces: for the effective presentation of visual information** », *UIST'95*, 14-17 novembre 1995
- [Sheelagh97] M. Sheelagh T. Carpendale, David J. Cowperthwaite, F. David Fracchia, « **Extending distortion viewing from 2D to 3D** », *IEEE Computer Graphics and Applications*, juillet-août 1997, pp. 42-51
- [Shu94] Li Shu, Woodie Flowers, « **Teledesign: groupware user experiments in three-dimensional computer-aided design** », *Collaborative Computing*, volume 1, number 1, pp. 1-14, 1994
- [Simsarian97] Kristian T. Simsarian, « **Virtual Reality control of a semi-autonomous robot through a multi-reality and speech control interface** », *XIV IMEKO World Congress, ISMCR'97 Topical Workshop on Virtual Reality and Advanced Man-Machine Interfaces*, pp. 164-169, 1-6 juin 1997, Tampere, Finlande
- [Slater93] M. Slater et M. Usoh, « **Simulating peripheral vision in immersive virtual environments** », *Computer & Graphics*, vol. 17, No. 6, pp. 643-653, Novembre/Décembre 1993
- [Smith89] Randall B. Smith, Tim O'Shea, Claire O'Malley, Eileen Scallon, Josie Taylor, « **Preliminary experiments with a distributed, multi-media, problem solving environment** », *Proceedings of the First European Conference on Computer Supported Cooperative Work: EC-CSCW'89*, Londres, UK, Septembre 1989, pp. 19-34. Réédité dans *Computer Studies in Computer Supported Cooperative Work: Theory Practice and Design*, J. Bowers et S. Benford (Eds.), Amsterdam, Elsevier Science Publishers B. V., 1991
- [Smith97] Gareth Smith, John Mariani, « **Subjective views to enhance 3D applications** », *ACM VRST '97*, pp. 139-146, Lausanne, Suisse, 15-17 septembre 1997
- [Staples93] L. Staples, « **Representation in virtual space: visual convention in the graphical user interface** », *ACM INTERCHI'93*
- [Stoakley95] Richard Stoakley, Matthew J. Conway, Randy Pausch, « **Virtual Reality on a WIM: Interactive Worlds in Miniature** », *Proceedings of CHI'95, CHI95*, pp. 265-272, 7-11 mai 1995, Denver, Colorado, USA

- [Tanaka96] Shunsuke Tanaka, Ken-ichi Okada, Shukei Kurihara, Yutaka Matsushita, « **Desktopconferencing system using multiple still-pictures: Desktop-MAJIC** », *CSCW'96*, Short Paper, novembre 1996
- [Tang92] John C. Tang, Ellen Isaacs, « **Why do users like video ? Studies of multimedia-supported collaboration** », *SMLI Technical Report Series TR-92-5*, décembre 1992, Sun Microsystems Laboratories, Inc.
- [Venolia93] Dan Venolia, « **Facile 3D Direct Manipulation** », *Proceedings of InterCHI'93*
- [Vernier97] Frédéric Vernier, Laurence Nigay, « **Représentations multiples d'une grande quantité d'information** », *IHM 97*, pp. 183-190, 9^{èmes} journées sur l'Ingénierie de l'Interaction Homme-Machine, Poitiers Futuroscope, France, 10-12 septembre 1997
- [Viaud95] Marie-Luce Viaud, Agnès Saulnier, « **Real Time Analysis and Synthesis Chain** », *Proceedings of International Workshop on Automatic Face and Gesture Recognition*, 1995
- [Viaud96] Marie-Luce Viaud, Agnès Saulnier, Pierre-Emmanuel Chaut, « **Analysis/synthesis system for facial animation** », International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11 *Coding of moving pictures and audio MPEG96*, septembre 1996
- [Viéville95] Claude Viéville, « **Structuring conversation in asynchronous communication systems to support collaborative learning at a distance** », *ED-MEDIA'95*, Graz, Autriche, 21 juin, 1995.
- [Viitaniemi97] J. Viitaniemi T. Määttä, K. Pekonen, J. Helminen, « **Virtual cockpit simulator** », *XIV Imeko World Congress, ISMCR'97 Topical Workshop on Virtual Reality and Advanced Man-Machine Interfaces*, pp. 127-131, 1-6 June 1997, Tampere, Finland
- [Wanger92] L. R. Wanger, J. A. Ferwerda, D. P. Greenberg, « **Perceiving spatial relations in computer-generated images** », *IEEE Computer Graphics and Applications*, pp. 44-57, mai 1992
- [Ware93] Ware, C., Arthur, K., and others, « **Fish Tank Virtual Reality** », *InterCHI*, 1993, 37-42.
- [Whitted80] T. Whitted, « **An improved illumination model for shaded display** », *Communications of the ACM* **23**:6, 1980, pp. 343-349
- [Williams77] Ederyn Williams, « **Experimental comparisons of face-to-face and mediated communication** », *Psychological Bulletin*, Vol. 84, No. 5, 1977, pp. 963-976
- [Zhai94] Shumin Zhai, William Buxton, Paul Milgram, « **The "SILK CURSOR": investigating transparency for 3D target acquisition** », *CHI'94 Proceedings*, Boston, Massachusetts USA
- [Zhai96] Shumin Zhai, William Buxton, Paul Milgram, « **The partial-occlusion effect: utilizing semitransparency in 3D human-computer interaction** », *ACM Transactions on Computer-Human Interaction*, Vol. 3, No. 3, septembre 1996, pp. 254-284

Annexe 1 – Définitions

Les définitions présentées ici n'ont pas la prétention d'être œcuméniques. De nombreux termes sont souvent employés avec des sens différents, ce qui manque pas de générer de vives querelles de clochers... Nous ne faisons ici que préciser le sens que nous-mêmes attribuons à certains termes dans le but de clarifier la lecture de ce mémoire.

ACTEUR : représentation d'une personne dans notre interface. Cette personne peut être réelle (l'acteur est un clone ou un avatar) ou artificielle (un agent, un assistant).

ACTIVITE : ensemble des actions et interactions au sein d'un groupe.

AGENT : un processus logiciel ayant un comportement autonome. Il n'est pas nécessairement représenté par une entité graphique [Capin97].

AVATAR : toute représentation d'une personne (smiley, photographie, vidéo, icône, *EyeCon*, personnage 3D, etc.). Si l'avatar est réaliste, on parle alors de **clone**.

BANDEAU : dans SpIn, ensemble des objets entourant la table ; l'espace accueillant ces objets.

CLONE : représentation réaliste d'une personne. La notion de « réalisme » est discutable, elle peut s'appliquer à l'aspect physique (traits du visage...), à l'expressivité, à la gestuelle.

CONSCIENCE DU REGARD (*gaze awareness*) : capacité de déterminer le centre d'intérêt d'une personne en suivant son regard.

CONTACT VISUEL : situation où deux personnes se regardent dans les yeux. C'est un élément social important, qui aide par exemple à déterminer qui va prendre la parole.

CONTINU : c'est la traduction que nous employons pour « seamless ». Introduit par Ishii [Ishii91], ce terme, qui signifie littéralement « sans couture » ou « sans soudure », qualifie un environnement homogène dans lequel tous les éléments sont intégrés sans séparation visible. Le substantif correspondant est « continuité » (« *seamlessness* »).

CVE (*Collaborative Virtual Environment*) : monde virtuel « collaboratif », partagé par plusieurs utilisateurs et utilisant les techniques de Réalité Virtuelle.

DECOR : éléments [statiques] dans l'interface ne visant qu'à situer, mettre en place l'**activité**. Notre table fait partie du décor.

DESKTOP CONFERENCING : tout système de réunion distante sur un [ordinateur de] bureau. Ce terme est plus large que la **visioconférence**.

DOCUMENT : tout objet contenant de l'information. Nous l'employons parfois très abusivement pour désigner tout objet n'étant pas un **acteur** : document au sens commun du terme mais aussi application, outil...

EXPLORATION : découverte d'un nouvel espace [Vernier97]. Dans notre contexte restreint, elle est synonyme de **navigation**.

MANIPULATION : ensemble des techniques d'interaction avec les objets de la scène. En 3D, trois actions de base doivent être réalisables : sélectionner, positionner et orienter un objet [Poupyrev97].

NAVIGATION : tout déplacement et changement du **point de vue** de l'utilisateur. Contrairement à l'**exploration**, ce terme implique qu'il existe un objectif à atteindre [Vernier97].

OBJET : dans notre interface, un objet est soit un **acteur** soit un **document**. Ne sont donc pas concernés les éléments de décor, comme la table par exemple.

ODESCA : le serveur centralisé d'activité. Voir « ODESCA : gestion de l'activité » (p. 99).

PARTICIPANT : une personne réelle qui participe à l'environnement virtuel en réseau et qui est représentée par un **avatar**.

POINT DE VUE : repérage spatial de l'observateur dans la **scène**. Ce repérage peut être géométrique (position et orientation comme dans MASSIVE [Benford93c]) ou plus abstrait comme dans notre interface.

SCENE : ce que voit l'utilisateur fait face. La scène de Spln est constituée d'**objets**, de décor, mais aussi du pointeur, d'indices visuels sur l'interaction (télépointeurs)...

SCENIC : le lien entre Spln et ODESCA. Voir « ScenIC : représentation abstraite de la scène » (p. 101).

SEAMLESS/SEAMLESSNESS : voir **continu**.

SPACE : notre projet dont le but est de définir un nouveau modèle d'interface 3D pour le **travail coopératif synchrone**.

SPIN : la partie de SPACE chargée de la représentation visuelle de la **scène**.

TÂCHE : travail particulier à réaliser, avec un but à atteindre.

TCAO : Travail Coopératif Assisté par Ordinateur

TELECONFERENCE : toute communication de groupe. Ce terme n'implique pas la production ni même le partage de documents

TELEPOINTEUR : indice visuel de l'activité matérialisant le point d'action d'une personne distante.

TRAVAIL COOPERATIF SYNCHRONE : pour nous, il s'agit des activités de travail à plusieurs faisant intervenir la *communication* entre personnes dans le processus de production (votes, design participatif...)

VIDEOCONFERENCE : toute technique de **téléconférence** utilisant la vidéo.

VISIOCONFERENCE : **vidéoconférence** sur ordinateur. La visioconférence apporte généralement le partage d'applications ou de documents en plus des techniques audiovisuelles classiques.

WIDGET : élément graphique associé à un **objet** de la scène, agissant comme un « périphérique virtuel ». Une action simple sur un widget déclenche des mécanismes compliqués difficiles à exécuter par manipulation directe. Exemples dans les interfaces **WIMP** : bouton, ascenseur, menu déroulant...

WIMP (*Window Icon Mouse Pointer*) : terme qualifiant toutes les interfaces 2D à fenêtres commercialisées depuis Xerox Star.

WYSIWIS (*What You See Is What I See*) : se dit d'une application partagée dans laquelle tout le monde a la même vue (mêmes fenêtres, mêmes emplacements, mêmes contenus).

Annexe 2 — Table des figures

Figure 1 : une téléconférence audiovisuelle	15
Figure 2 : miroirs semi-transparents pour préserver le contact visuel	15
Figure 3 : la salle du professeur à l'ENIC	16
Figure 4 : le « visiocentre » de l'ENIC	16
Figure 5 : les portholes de Xerox	17
Figure 6 : un outil de visioconférence	18
Figure 7 : Bill Buxton converse par l'intermédiaire de trois Hydra Units	22
Figure 8 : la TeamWorkStation, avec ses deux caméras	23
Figure 9 : présentation en « mosaïque » sur la TeamWorkStation	24
Figure 10 : visages surimposés et mains en transparence	24
Figure 11 : le ClearBoard 1	25
Figure 12 : le ClearBoard 2	26
Figure 13 : MAJIC	26
Figure 14 : le MultiConference Manager de Genesys	31
Figure 15 : une interlocutrice	31
Figure 16 : le viewpoint avec la ligne matérialisant le point d'action	33
Figure 17 : EyeCons	34
Figure 18 : EyeCons, Spraycans et vues séparées	35
Figure 19 : deux personnes se rejoignent dans une pièce virtuelle	36
Figure 20 : l'utilisateur voit ses interlocutrices et un document	37
Figure 21 : l'activité vue dans des fenêtres	37
Figure 22 : un cockpit virtuel, vues intérieure et extérieure [Viitaniemi97]	38

Figure 23 : une capture d'écran d'un client BSX	39
Figure 24 : « Head Mounted Display »	40
Figure 25 : le pipe-line de rendu classique	41
Figure 26 : un dispositif de suivi du regard, remplaçant ici la souris	43
Figure 27 : périphériques isométrique, isotonique et élastique	43
Figure 28 : champs visuels de l'œil gauche et de l'œil droit [Emslie-Smith89, p. 451]	46
Figure 29 : rendu normal (à gauche) et avec zone périphérique (à droite)	47
Figure 30 : projection classique et projection modifiée « grand angle »	47
Figure 31 : « blockies » de DIVE avec leurs auras	49
Figure 32 : différents avatars et clones dans DIVE	50
Figure 33 : quelques avatars de MASSIVE	51
Figure 34 : des clones dans VLNet	51
Figure 35 : le squelette et ses articulations	52
Figure 36 : les squelettes recouverts de metaballs	53
Figure 37 : placage de texture sur un modèle 3D	54
Figure 38 : chaîne de traitement temps réel	55
Figure 39 : un clone « bande dessinée »	57
Figure 40 : vue de face et de profil	58
Figure 41 : visages bruts (en haut) et lissés par filtrage (en bas)	58
Figure 42 : textures à appliquer aux clones	59
Figure 43 : modèles 3D de clones	59
Figure 44 mosaïque de fenêtres	61
Figure 45 : recouvrements entre fenêtres	62
Figure 46 : le mécanisme client/serveur de X-window [Nye92]	63
Figure 47 : patches translucides	66
Figure 48 : destruction d'un patch	68
Figure 49 : différents schémas de distorsion 3D	70
Figure 50 : Perspective Wall	71
Figure 51 : arborescence vue sur le Perspective Wall	72
Figure 52 : une suite de listes sur le Perspective Wall	72
Figure 53 : une arborescence « Cone Tree »	73
Figure 54 : une arborescence de fichiers et de répertoires sous forme de Cone Tree	73
Figure 55 : réutilisation des widgets 2D en Réalité Virtuelle	77
Figure 56 : les Worlds In Miniature	78
Figure 57 : le matériel de MAJIC	82
Figure 58 : dispositif MAJIC complet avec plusieurs interlocuteurs	82
Figure 59 : les neuf positions dans Desktop-MAJIC	84
Figure 60 : Desktop-MAJIC avec photos et télépointeurs	84
Figure 61 : bouche fermée et ouverte	85

Figure 62 : réunion autour d'un tableau dans MASSIVE	88
Figure 63 : l'architecture du projet SPACE	99
Figure 64 : un exemple d'arbre décrivant la scène	102
Figure 65 : le jeu Doom II	108
Figure 66 : le premier « bandeau »	112
Figure 67 : 4 puis 5 personnes réparties autour de la table	113
Figure 69 : Observateur fixe et objets venant dans son champ de vision	114
Figure 70 : objets placés autour d'une demi-table, dans le champ de vision	115
Figure 71 : disparition, à gauche, et réapparition simultanée, à droite, de l'objet X	116
Figure 72 : la distorsion, schématisée	118
Figure 73 : le bandeau subissant une distorsion	118
Figure 74 : boîte englobante de chaque objet graphique	124
Figure 75 : arrivée d'un participant	125
Figure 76 : objets composés et boîtes englobantes	126
Figure 77 : bandeau à cinq objets, sans distorsion	128
Figure 78 : distorsion angulaire	129
Figure 79 : distorsion horizontale	129
Figure 80 : transformations à appliquer à l'objet pour la distorsion	130
Figure 81 : les objets 3D périphériques sortent de l'écran	130
Figure 82 : aplatissement des objets périphériques	131
Figure 83 : pas d'aplatissement latéral, le clone de gauche disparaît	131
Figure 84 : le clone reste visible grâce à l'aplatissement latéral	132
Figure 85 : rotation de la tête et des yeux	133
Figure 86 : un clone expérimental	133
Figure 87 : un télépointeur	134
Figure 88 : la table et le bandeau contenant des documents	135
Figure 89 : Grégory prend place autour de la table	135
Figure 90 : Cédric montre un logo à Grégory	136
Figure 91 : le logo désigné par le clone de Cédric	136
Figure 92 : pointage sur un document	137
Figure 93 : le télépointeur de Cédric a bougé	137
Figure 94 : Grégory porte son regard sur le logo en question	138
Figure 95 : Cédric voit Grégory regarder l'autre logo	139
Figure 96 : la distorsion agrandit le centre d'intérêt	140
Figure 97 : manipulation 3D d'une molécule	141
Figure 98 : le clone regarde dans son espace privé	142
Figure 99 : une réunion à cinq	146
Figure 100 : une armoire accueille les objets « inactifs »	146
Figure 101 : l'objet sur la table est transparent	147

THESES

1997

- 06 Janvier **Philippe MESEURE**
Modélisation de corps déformables pour la simulation d'actes chirurgicaux.
- 07 Janvier **Olivier ROUSSEL**
L'achèvement des bases de connaissances en calcul propositionnel et en calcul des prédicats.
- 08 Janvier **Nouredine MELAB**
Gestion de la granularité et régulation de charge dans le modèle P³ d'évaluation parallèle des langages fonctionnels
- 09 Janvier **Cédric DUMOULIN**
Dream : une mémoire partagée répartie à cohérence programmable.
- 13 Janvier **Raymond NAMYST**
PM² : Un environnement pour une conception portable et une exécution efficace des applications parallèles irrégulières.
- 14 Janvier **Philippe MERLE**
CorbaScript - Corba Web : propositions pour l'accès à des objets et services distribués.
- 27 Janvier **David GALINEC**
Exécution asynchrone de programmes synchrones par transformations automatiques : application au traitement d'images temps-réel.
- 14 Mars **Côme RACZY**
Commandes optimales en temps pour des systèmes différentiellement plats. Application aux commandes de satellites et de grues.
- 27 Mars **Jean Jacques VANDEWALLE**
Projet OSMOSE : Mosélisation et implémentation pour l'interopérabilité de services carte à microprocesseur par l'approche orientée objet.
- 10 Juin **David SIMPLOT**
Langages de mots de figures, monoïdes inversifs et langages de mots à deux dimensions
- 20 Juin **Dominique SUEUR**
Algorithmes de redistribution de données. Application aux systèmes de fichiers parallèles distribués.

25 Juin

Olivier DELGRANGE

Un algorithme rapide pour une compression modulaire optimale.
Application à l'analyse de séquences génétiques.

THESES

1998

- 09 Janvier **David CARLIER**
Représentation permanente, coordonnée par une carte à microprocesseur, d'un utilisateur mobile.
- 16 Janvier **Yves DENNEULIN**
Conception et ordonnancement des applications hautement irrégulières dans un contexte de parallélisme à grain fin.
- 23 Janvier **Jaafar GABER**
Plongements et manipulations d'arbres dans les architectures distribuées.
- 30 Janvier **Grégory SAUGIS**
Interfaces 3D pour le travail coopératif synchrone, une proposition.

