



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*

Discipline : *Informatique*

Présentée et soutenue par *Sébastien BOTTECCHIA*

Le *20 septembre 2010*

Titre :

Système T.A.C. : Télé-Assistance Collaborative.

Réalité augmentée et NTIC au service des opérateurs et des experts dans le cadre d'une tâche de maintenance industrielle supervisée.

JURY

<i>Indira MOUTTAPA THOUVENIN</i>	<i>Docteur, titulaire de l'HDR</i>	<i>Rapporteur</i>
<i>Simon RICHIR</i>	<i>Professeur</i>	<i>Rapporteur</i>
<i>Franck LUTHON</i>	<i>Professeur</i>	<i>Examineur</i>
<i>Olivier PATROUX</i>	<i>Docteur</i>	<i>Examineur</i>
<i>Eric SENTURIER</i>	<i>Société Turboméca</i>	<i>Membre invité</i>
<i>Jean-Jacques SZTOGRYN</i>	<i>Société LCI</i>	<i>Membre invité</i>

École doctorale : *Mathématique Informatique TéléCommunication Toulouse (MITT)*

Unité de recherche : *Institut de Recherche en Informatique de Toulouse (IRIT)*

Directeur de thèse : *Jean-Pierre JESSEL (Co-directeur : Jean-Marc CIEUTAT)*

« Une idée simple et fausse aura toujours plus de poids qu'une idée vraie et complexe »

– Montesquieu –

« Le travail s'étale de façon à occuper le temps disponible pour son achèvement »

– Loi de Parkinson –

« La différence entre la théorie et la pratique, c'est qu'en théorie, il n'y a pas de différence entre la théorie et la pratique, mais qu'en pratique, il y en a une ! »

– Jan Van de Snepscheut (attribué à) –

TABLE DES MATIÈRES

LISTE DES FIGURES	9
LISTE DES TABLEAUX	13
LISTE DES ABRÉVIATIONS	15
INTRODUCTION GÉNÉRALE	17
Contexte	17
Problématique	18
Structure du mémoire	18
I TOUR D’HORIZON EN RÉALITÉ AUGMENTÉE	19
1 TERMINOLOGIE	20
1.1 Définitions et principes... l’évolution	20
1.2 Taxonomies	24
1.2.1 Taxonomie fonctionnelle	24
1.2.2 Taxonomie technique	25
1.2.3 Autres taxonomies	27
2 PRINCIPES TECHNOLOGIQUES	29
2.1 Les différentes briques technologiques	29
2.2 Le principe du rendu	31
2.3 Les technologies d’affichage	33
2.3.1 Les « Head Worn Displays »	33
2.3.2 Les « Handheld Displays »	35
2.3.3 Les « Projective Displays »	36
2.4 Suivi et recalage	37
2.4.1 Le suivi (Tracking)	38
2.4.2 Le recalage (Registration)	40
3 CONCLUSION	42

II L'HOMME, LA MAINTENANCE ET LA RÉALITÉ AUGMENTÉE	43
4 LE FACTEUR HUMAIN	44
4.1 Les mécanismes de l'erreur	44
4.1.1 Le processus cognitif	44
4.1.2 Topologie de l'erreur	45
4.1.3 Les formes d'erreur	46
4.2 Les erreurs dans l'activité de maintenance	47
4.2.1 La défaillance, prémices de l'erreur	47
4.2.2 Les principales erreurs	47
4.3 Taxonomie des erreurs	49
5 L'AIDE AU TRAVAIL	50
5.1 Pourquoi recourir à une aide?	51
5.1.1 Objectifs et définition	51
5.1.2 Les écarts	51
5.2 Les types d'aides	52
5.3 Procédures, documentations... une énigme?	53
5.3.1 Considérations générales	53
5.3.2 Dans l'industrie	55
5.4 L'assistance téléphonique	57
5.5 Conclusion sur l'aide « classique »	58
5.6 Potentiel de la RA dans l'industrie	58
5.6.1 Intérêts pour la tâche	58
5.6.2 Intérêts pour l'opérateur	59
6 LES SYSTÈMES ACTUELS DE R.A.	60
6.1 Les systèmes « Automatisés »	60
6.2 Les systèmes « Supervisés »	66
7 NAISSANCE D'UNE PROBLÉMATIQUE	69

III	L'ASSISTANCE HUMAINE DISTANTE : PROPOSITIONS	71
8	LE POURQUOI DE L'INTERVENTION HUMAINE	72
8.1	La connaissance	72
8.1.1	Les différents types de connaissances	72
8.1.2	Théories d'apprentissage	74
8.2	De la compétence à l'expertise	77
8.2.1	La compétence	77
8.2.2	Une représentation de la compétence	78
8.2.3	Vers l'expertise	79
8.3	Travail collaboratif	80
8.4	Conclusion sur l'intervention humaine	81
9	COMMUNICATION DISTANTE : LES PARAMÈTRES	81
9.1	Collaborer à distance	82
9.1.1	Les types de collaboration	82
9.1.2	Les asymétries	83
9.2	Les facteurs importants de communication	84
9.2.1	Partage de l'espace visuel	84
9.2.2	Les références ostensives	85
10	PROPOSITIONS POUR SOUTENIR LA COLLABORATION DISTANTE	87
10.1	La façon de percevoir	88
10.2	Paradigme d'interaction : P. O. A.	90
11	CONCLUSION	94
IV	SYSTÈME T.A.C : TÉLÉ-ASSISTANCE COLLABORATIVE	
	(Conception et Expérimentation)	95
12	LE SYSTEME T.A.C.	96
12.1	Rappel de fonctionnement	96
12.2	Diagramme de déploiement	97
12.3	Choix d'implémentation	98

12.4	Vue d'ensemble de l'architecture interne	99
12.4.1	Diagramme de relation	99
12.4.2	Description du rôle des composants de TAC-Opérateur	100
12.4.3	Description du rôle des composants de TAC-Expert	101
13	PERCEPTION : LES LUNETTES MOVST	103
13.1	La perception visuelle humaine	103
13.1.1	La luminance	104
13.1.2	Le champ visuel	105
13.1.3	La vision stéréoscopique	107
13.1.4	Interprétation de l'information	108
13.2	Caractérisation du prototype	109
13.2.1	Objectifs	109
13.2.2	Choix	110
13.2.3	Prototype MOVST	113
14	INTERACTION : IMPLÉMENTATION DE P.O.A.	114
14.1	L'opérateur : un être capable et en mouvement	114
14.2	Picking : Suivi et Latence	114
14.2.1	Le suivi de points d'intérêt	114
14.2.2	Implémentation du « Picking »	115
14.2.3	La latence entre l'expert et l'opérateur	117
14.3	Outlining-Adding : Reconnaissance et Localisation	118
14.4	L'interface « Expert »	119
15	ÉVALUATION EXPÉRIMENTALE	121
15.1	Ce que l'on cherche à évaluer	121
15.2	Description de l'évaluation	121
15.2.1	Descriptions des tâches et des moyens de téléassistance	121
15.2.2	Sujets et plan de recherche	122
16	ANALYSE ET INTERPRÉTATIONS	123
16.1	Méthode d'analyse des résultats	123
16.1.1	Pour le temps de réalisation	123
16.1.2	Pour les données du questionnaire	124

16.2 Résultats et interprétations	124
16.2.1 Pour le temps de réalisation	124
16.2.2 Pour les données du questionnaire	125
16.2.3 Discussion des résultats obtenus	127
17 SYNTHÈSE	129
CONCLUSION GÉNÉRALE ET PERSPECTIVES	131
Rappel du contexte	131
Synthèse du document	131
Résumé de nos contributions et résultats	132
Limitations et Perspectives	134
Références	135

LISTE DES FIGURES

1	Casque de réalité augmentée semi-transparent à vision stéréoscopique imaginé par Sutherland en 1966.	21
2	HUD installé dans un cockpit d'avion militaire. Les informations virtuelles coïncident avec le point de vue du pilote.	21
3	Présentation d'incrustations virtuelles.	23
4	Continuum Réalité-Virtualité.	25
5	Degré de connaissance du monde.	26
6	Degré de fidélité du monde en fonction de la technologie d'affichage et du rendu.	26
7	Degré d'immersion de l'utilisateur en fonction de la technologie d'affichage.	27
8	Taxonomie des différents mondes vue par Bérard.	27
9	Taxonomie de Dubois.	28
10	Briques constituant un système de réalité augmentée.	29
11	Utilisation d'un collier de connecteur d'antenne comme bouton rotatif en vue de modifier une valeur dans une interface virtuelle.	30
12	Proposition de dispositif visant à modifier la façon de présenter l'information en fonction des conditions de la scène.	30
13	Alignement réel-virtuel (extrait de [SD06]).	32
14	Cohérence spatio-temporelle (extrait de [SD06]).	32
15	Cohérence photométrique (extrait de [SD06]).	32
16	Réalité augmentée en vision directe (Optical See-Through) [MR08].	34
17	Réalité augmentée en vision indirecte (Video See-Through) [MR08].	34
18	Dispositif « Handheld Display » [MR08].	36
19	Réalité augmentée projetée.	37
20	Head Mounted Projection Display [MR08].	37
21	Différentes cibles codées.	40
22	Modèle de l'échelle double.	44
23	Modèle de l'enchaînement d'erreurs de Reason, dit « modèle de l'Emmental ».	45
24	« Décomposition cartographique » de la maintenance par rapport à la défaillance [Rey02].	47
25	Catégorisation des erreurs ayant induit des accidents.	48
26	Opérateur effectuant une tâche de maintenance en suivant les instructions virtuelles.	60
27	Câblage électrique assisté.	61
28	Aide apportée à l'opérateur de façon contextuelle [WKS03].	61

29	Principe du projet STARMATE.	62
30	Exemples de visualisation de données [AMR02].	63
31	Principe de la méthode de Tracking.	64
32	Le système de BMW.	64
33	Le système ARMAR.	65
34	Tâche d'entraînement. L'opérateur manipule les objets virtuels via un marqueur.	66
35	Le système WACL.	67
36	Dispositif porté par l'opérateur.	67
37	Interface de l'expert. À droite, le flux vidéo temps réel. À gauche, une image capturée en train d'être augmentée avant son envoi à l'opérateur.	68
38	CAMKA Gun.	68
39	Les différents types de connaissance en science cognitive.	73
40	Continuum des approches théoriques de l'apprentissage [BRW98].	76
41	Représentation de la compétence sous forme « atomique » selon [Jed94].	78
42	Matrice CSCW ou matrice Espace/Temps.	82
43	Notre proposition du principe de fonctionnement d'un système d'assistance à distance.	87
44	Représentation du champ visuel de l'opérateur portant un dispositif See Through sur l'oeil droit (intérieur du rectangle rouge).	88
45	Proposition de visualisation de l'espace de l'opérateur pour l'expert. La vue orthoscopique (intérieur du rectangle rouge) est placée dans la vue panoramique. Nous voyons ici l'intérêt de pouvoir restituer la vue orthoscopique par rapport à la vue panoramique.	90
46	Vue augmentée de l'opérateur après une interaction de type « Picking ». Ici, l'élément désigné est clairement discriminé parmi tous les autres. Inutile donc de le décrire, ce qui ici peut se révéler ardu !	92
47	Vue augmentée de l'opérateur après une interaction de type « Outlining ». L'expert a souhaité mettre en avant l'objet de la discussion (pièces détournées en rouge) tout en laissant apparaître une indication pouvant se révéler importante pour la suite des opérations.	93
48	Vue augmentée de l'opérateur après une interaction de type « Adding ». L'expert montre, par une animation virtuelle 3D recalée sur l'élément réel, l'enchaînement des opérations.	93
49	Principe de fonctionnement de T.A.C.	96
50	Design général de T.A.C.	97
51	Vue d'ensemble de l'architecture basée composants du système T.A.C.	99
52	Détails de l'oeil et du cortex visuel cérébral.	103

53	Spectre électromagnétique (haut) et spectre visible par l'oeil humain (bas).	104
54	Luminance relative en fonction de la longueur d'onde visible par l'oeil humain (à intensité lumineuse égale).	105
55	Champ de vue horizontal et champ de vue vertical.	106
56	Zones d'action du champ visuel.	106
57	Les illusions d'optique sont la manifestation de notre cerveau à recréer de l'information en étant influencé par le contexte et l'expérience.	109
58	Nos lunettes MOVST.	113
59	Méthode proposée pour l'implémentation du « Picking ».	116
60	Procédé de suivi des points d'intérêt dans la mémoire tampon en fonction de la différence inertielle Δ_{inertiel}	118
61	Interface de l'expert.	119

LISTE DES TABLEAUX

2	Taxonomie de la réalité augmentée selon Mackay.	28
3	Différents HWD utilisés en réalité augmentée.	35
4	Critères de performance minima pour un HMD suivant l'application.	35
5	Principaux facteurs affectant les performances humaines.	48
6	Motif du non-respect des procédures par les opérateurs de maintenance dans l'industrie [Emb00].	55
7	Problèmes liés aux manuels de maintenance.	56
8	Ce que les opérateurs en maintenance pensent des manuels.	56
9	Transformation de la connaissance [NT95].	74
10	Comparaison des notions de coopération et de collaboration selon [Pue03] (résumé par [Ros04]).	80
11	Quelques gammes d'éclairage.	105
12	Effet inter-sujet sur le temps de complétion.	125
13	Notes moyennes pour chaque mode en fonction des questions.	126
14	Degré d'accord entre les différents sujets sur le classement des systèmes.	126

LISTE DES ABRÉVIATIONS

API : Advanced Programming Interface
BDD : Base De Données
CAO : Conception Assistée par Ordinateur
CMOS : Complementary Metal Oxide Semiconductor
CSCW : Computer Supported Cooperative Work
HAD : Hole Accumulation Diode
HMD : Head Mounted Display
HWD : Head Worn Display
IHM : Interface Homme Machine
KLT : Kanade Lucas Tomasi (algorithme)
LCD : Liquid Crystal Display
LCOS : Liquid Cristal On Silicon
MOVST : Monocular Orthoscopic Video See Through
NTIC : Nouvelles Technologies de l'Information et de la Communication
OST : Optical See Through
PAL : Phase Alternating Line
PiP : Picture in Picture
P.O.A. : Picking Adding Outlining
POSIT : Pose from Orthography and Scaling with Iterations
RA : Réalité Augmentée
RAM : Random Access Memory
SDK : Software Development Kit
SURF : Speeded Up Robust Feature (algorithme)
T.A.C. : Télé-Assistance Collaborative
TUI : Tangible User Interface
UKCAA : United Kingdom Civil Aviation Authority
VA : Virtualité Augmentée
VGA : Video Graphics Array
VST : Video See Through
WIMP : Windows, Icons, Menus and Pointing device

INTRODUCTION GÉNÉRALE

Contexte

Nous assistons ces dernières années à une attente forte du monde industriel pour intégrer de nouveaux outils d'assistance technologique. Ce besoin exprime les difficultés rencontrées par les techniciens en maintenance actuellement confrontés à la grande variété de systèmes mécatroniques toujours plus complexes, et à un renouvellement des gammes de plus en plus rapide.

La compression des durées de formations et la multiplicité des procédures de maintenance favorisent l'apparition de nouvelles contraintes liées à l'activité des opérateurs. On peut citer par exemple le manque de « visibilité » du système à maintenir et l'incertitude des opérations à réaliser. Ces constatations tendent à obliger les mécaniciens à se former « sur le tas », ce qui peut à terme impliquer un accroissement des erreurs de procédures et par ce biais augmenter les coûts de maintenance, pour au final engendrer des pertes de temps considérables.

Dans ce contexte mondial fortement concurrentiel, la demande des industriels à augmenter l'efficacité des tâches du support technique et de maintenance passe alors par l'intégration de nouvelles technologies de l'information et de la communication (NTIC).

Avec l'avènement de la réalité augmentée, plusieurs projets ont donc vu le jour dans l'espoir de créer des systèmes d'aides efficaces. Tous ont pour vocation d'accompagner un opérateur dans l'accomplissement de sa tâche en fournissant de l'information contextualisée (visuelle ou sonore) et en temps réel. L'utilisateur doit pouvoir porter son attention sur la tâche à réaliser et non sur l'utilisation de l'outil lui-même, d'où les différentes stratégies de chaque projet dans l'élaboration des prototypes. Aussi, le choix du dispositif d'affichage est important, car l'objectif peut être de pouvoir minimiser le recours aux supports classiques (papier) et laisser ainsi les mains de l'opérateur libres.

Enfin, tous ces systèmes sont particulièrement pertinents lorsque les tâches sont régies par des règles attribuant une action à une situation précise, c'est-à-dire dans le cadre d'une procédure opérationnelle standard. On parle alors de « connaissance explicite ». Mais l'accès à cette connaissance n'est pas forcément suffisant pour savoir comment l'utiliser. Ce savoir, dit tacite (ou implicite), appartient au domaine de l'expérience, des aptitudes ou du savoir-faire. Ce type de connaissance est personnel et est difficile à traduire en termes de représentation. Ainsi, les systèmes actuels de réalité augmentée pour la maintenance sont impuissants lorsque survient une situation imprévue. Il est alors parfois nécessaire d'avoir recours à une personne distante ayant le niveau de qualification requis.

Problématique

Lorsqu'un opérateur travaillant seul a besoin d'aide, il n'est pas forcément évident de trouver à travers le monde la personne avec le niveau de compétence et de connaissance qui puisse l'assister. Grâce à l'explosion des débits de communication et du World Wide Web, la téléassistance temps réel devient accessible. Cette collaboration entre un expert et un opérateur engendre de multiples bénéfices comme le contrôle qualité ou le retour d'information. Encore faut-il avoir un système permettant de supporter des interactions à distance !

Avec la réalité augmentée, nous pouvons désormais envisager un système de collaboration à distance permettant à l'expert d'être virtuellement coprésent avec l'opérateur. Comme le processus de collaboration est imprévisible et indéterministe, cela implique que l'expert puisse avoir à sa disposition des possibilités d'interactions apparentées à celles de l'opérateur et à celles qui lui sont disponibles dans le réel. On pense notamment à la désignation ou à la capacité de pouvoir mimer des actions. En ce sens, en permettant à l'expert de voir ce que voit l'opérateur (en se mettant virtuellement à sa place), il peut alors en temps réel interagir avec lui via un paradigme d'interaction adéquat.

Notre recherche s'axe donc sur la possibilité d'avoir un système de collaboration à distance permettant d'interagir de façon aussi efficace et naturelle qu'en situation de coprésence. Nous nous sommes alors appuyés sur la réalité augmentée pour proposer de nouvelles modalités de communication, modalités expérimentées ici dans le cadre d'activités de maintenance industrielle.

Structure du mémoire

Dans le chapitre 1, nous allons commencer par retracer l'évolution de la définition de la réalité augmentée. À l'issue de cet historique, nous proposerons au lecteur notre propre définition afin d'établir une compréhension commune de ce terme au fil du mémoire. Nous nous consacrerons ensuite à présenter les fondements technologiques qui constituent la réalité augmentée, l'intention étant de nous éclairer sur ses principes nous ayant permis de proposer notre solution.

Comment les erreurs naissent-elles ? Pourquoi certaines aides sont-elles inefficaces ? Quel pourrait être l'intérêt d'introduire les nouvelles technologies comme la réalité augmentée dans le processus d'assistance ? Est-ce que les propositions actuelles permettent de soutenir efficacement un opérateur dans sa tâche ? C'est à ces questions que le chapitre 2 va répondre, et les manquements constatés nous amèneront à formuler notre problématique.

Le chapitre 3, lui, va au préalable montrer pourquoi à l'heure actuelle l'humain est encore un chaînon indispensable en termes d'assistance. Nous verrons par la suite les mécanismes d'une communication interpersonnelle réussie et en quoi la collaboration, même distante, est bénéfique. C'est à la lumière de ces observations que nous énoncerons nos propositions, en termes de perception et d'interaction, visant à « simuler » la coprésence d'un expert auprès d'un opérateur dans le cadre d'une tâche de maintenance supervisée, et cela, grâce à la réalité augmentée.

Enfin, dans le chapitre 4, nous allons présenter le système T.A.C., système se révélant être l'implémentation de nos propositions de visualisation et d'interaction. Nous avons été également en mesure d'expérimenter une partie de nos principes, nous en présenterons les premiers résultats.

Première partie

TOUR D'HORIZON EN RÉALITÉ AUGMENTÉE

L'objectif de ce chapitre n'est pas de faire un énième état de l'art sur la réalité augmentée, mais plutôt de nous éclairer sur ses principes et ses technologies nous ayant permis de proposer notre solution.

Ainsi, ce chapitre pourrait se découper en deux parties. La première retrace la réalité augmentée à travers l'évolution de ses définitions et de ses classifications depuis sa création. Nous verrons que les principes mêmes sont dépendants tant de la communauté qui l'énonce que de la technologie employée. À l'issue de cet historique, nous proposerons alors au lecteur notre propre définition du terme « réalité augmentée ». Nous souhaitons ainsi établir, pour la lecture de ce document, une base commune du concept même de ce que nous pensons être la réalité augmentée. Mais au-delà de fournir une nouvelle définition, nous avons voulu ouvrir des pistes de réflexion sur l'expression « réalité augmentée » qui pourraient, à terme, fournir une caractérisation de l'essence même de ce concept.

La deuxième partie quant à elle, vise à présenter les principes technologiques indispensables à ce domaine. Du type de rendu aux problèmes de recalage des informations, nous nous apercevons des difficultés qui subsistent encore à ce jour. Mais cela n'a en rien empêché que la réalité augmentée soit aux portes de la dernière décennie, la nouvelle ayant même déjà commencé à la faire découvrir au grand public.

1 TERMINOLOGIE

1.1 Définitions et principes... l'évolution

Réalité¹ : n.f (latin médiéval *realitas*, du latin classique *realis*, de *res*, chose)

1. Caractère de ce qui est réel, de ce qui existe effectivement.
2. Ce qui est réel, ce qui existe en fait, par opposition à ce qui est imaginé, rêvé, fictif.

Augmenter¹ : v.tr (bas latin *augmentare*, du latin classique *augere*, accroître)

1. Rendre quelque chose plus grand, plus considérable.
2. Rendre quelque chose plus important, plus intensif.

L'association de ces deux mots en une seule expression « réalité augmentée » a été introduite au début des années 90 par Caudell [CM92]. La réalité augmentée désigne alors tout système interactif présentant à un individu une combinaison en temps réel d'entités virtuelles (générées par ordinateur) avec des éléments physiques de l'environnement réel. L'idée sous-jacente non exprimée est cependant que cette combinaison implique une relation sémantique forte entre le réel et le virtuel. Ainsi, les propriétés et relations intrinsèques des éléments de l'environnement réel seraient révélées à nos sens².

La réalité augmentée compléterait donc la réalité sans avoir pour vocation de la remplacer [Azu97] comme c'est le cas en réalité virtuelle. Toutefois, le concept même de la réalité augmentée ne date pas des années 90. Ce sont en effet les travaux d'Ivan Sutherland qui ont jeté les bases de la réalité augmentée au milieu des années 60 [Sut68]. Il réalise alors le premier système basé sur un casque semi-transparent stéréoscopique permettant de combiner la vue de la réalité avec une image de synthèse. L'affichage des entités virtuelles s'effectuait alors par des tubes cathodiques miniatures et le suivi des mouvements de la tête par un bras mécanique lié au casque (Fig. 1).

Il est à noter que l'application présentée par Sutherland était destinée à visualiser des molécules en trois dimensions et que l'auteur précise que « les objets affichés peuvent être détachés du monde physique ou coïncider avec des cartes, des bureaux, des murs ou des touches d'une machine à écrire ». Bien que l'on trouve dans le système proposé l'essentiel des éléments constituant les systèmes de réalité augmentée modernes, nous pouvons nous interroger quant à le classer comme tel sachant qu'il n'existe pas à proprement parler de lien sémantique entre le réel et le virtuel. Toutefois, avant les années 2000, nous n'avons pas trouvé dans la littérature l'obligation d'existence d'un tel lien³.

Malgré tout, et ce de manière implicite, les années 80 ont accordé une grande place aux informations sémantiquement liées. En effet, la généralisation des dispositifs d'affichage tête haute (Head Up Display ou HUD) a trouvé nombre d'applications dans le domaine militaire, notamment dans les

1. Définition étymologique du dictionnaire Larousse.

2. L'idée même de révéler à nos sens ces informations implique que le domaine de la réalité augmentée n'est pas cloisonné au seul organe visuel comme le rappelle Azuma [ABB⁺01]. Toutefois, nous nous intéresserons dans ce manuscrit uniquement à la modalité visuelle.

3. Remarquons cependant que dans la communauté IHM, l'objectif est autre. En effet, suite aux travaux de Weller [Wel93] faisant référence dans le domaine, l'idée prédominante est de rendre accessible le virtuel via le réel. On parlera alors ici plus d'affordance [Gib86][Nor88] que de lien sémantique.

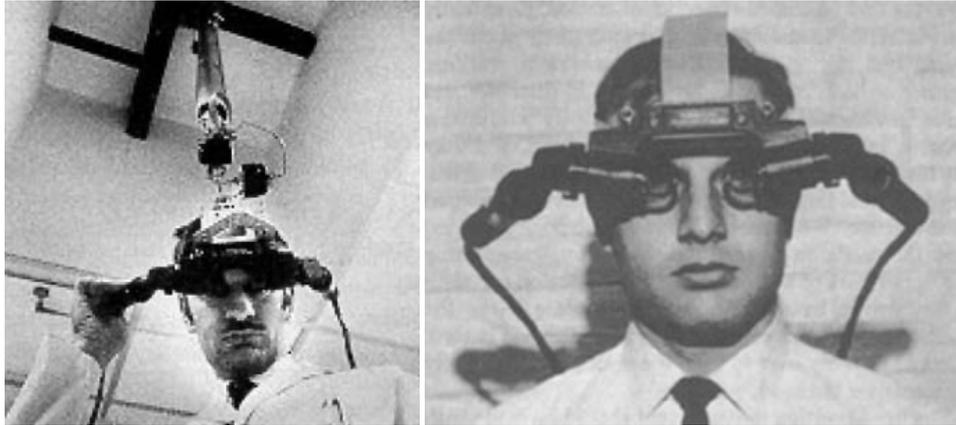


FIGURE 1: Casque de réalité augmentée semi-transparent à vision stéréoscopique imaginé par Sutherland en 1966.

cockpits des avions de combat (Fig.2). Via un miroir semi-transparent sur lequel on projette les informations utiles à la navigation (cap, inclinaison...), le pilote perçoit de son point de vue réalité et virtualité.



FIGURE 2: HUD installé dans un cockpit d'avion militaire. Les informations virtuelles coïncident avec le point de vue du pilote.

Il faudra attendre l'article de Milgram datant de 1994 [MMT⁺94] pour voir réellement apparaître une discussion sur l'éventuelle définition de la réalité augmentée⁴. Et l'on se rend compte que la définition a tendance à varier en fonction de la provenance (journal, communauté...). Nous retiendrons notamment deux définitions à ce stade choisies par Milgram. La première (à laquelle Milgram adhère) est une définition plutôt large :

Réalité augmentée : « augmenter la rétroaction naturelle de l'opérateur avec des indices virtuels ».

La seconde est une définition plus restrictive dans le sens où elle fait intervenir la technologie pour justifier l'approche :

Réalité augmentée : « une forme de réalité virtuelle où l'afficheur tête-porté⁵ du participant est transparent, ce qui permet une vision claire du monde réel ».

4. Dans le cas présent, l'auteur cherche plus à effectuer une classification qu'établir une véritable définition.

5. En anglais Head Mounted Display (HMD). Cette technologie d'affichage sera présentée plus loin dans le chapitre.

Néanmoins, dans l'une ou l'autre des définitions, rien ne laisse entrevoir la relation qui doit exister entre le réel et le virtuel. Cependant, la deuxième définition est clairement orientée vers la modalité visuelle et exclut donc toute autre modalité.

Trois ans plus tard, Azuma propose une définition beaucoup plus restrictive [Azu97] en imposant une relation d'existence concomitante entre le réel et le virtuel :

Réalité augmentée : « permet à l'utilisateur de voir le monde réel avec une superposition, composée ou par-dessus, d'objets virtuels. Toutefois, la réalité augmentée complète la réalité, plutôt que de la remplacer. Idéalement, l'utilisateur devrait percevoir dans le même espace des objets réels et virtuels coexister ».

Définition qu'il affina dès 2001 sous la forme [ABB⁺01] :

Réalité augmentée : « système qui complète le monde réel avec des objets virtuels (générés par ordinateur) de telle sorte qu'ils semblent coexister dans le même espace que le monde réel ». Ce qui dans les deux cas l'amène à définir les caractéristiques d'un système de réalité augmentée par les trois propriétés suivantes :

1. « Combiner le réel et le virtuel ». Au monde réel en trois dimensions doivent être intégrées des entités également en trois dimensions.
2. « Interactivité en temps réel ». Cela exclut notamment les films bien que la condition précédente soit respectée.
3. « Recalage en trois dimensions ». Cela permet de faire coïncider visuellement les entités virtuelles avec la réalité.

Là encore, cette définition caractérisée de la réalité augmentée est définitivement tournée vers l'aspect visuel, ne laissant aucune place à la multimodalité. Même si cette définition fait référence encore à l'heure actuelle, elle est discutable sur au moins deux points (surtout d'un point de vue ergonomique [Ana06]). Premièrement, l'exigence d'entités virtuelles en trois dimensions n'est pas forcément justifiée. Si nous considérons que les augmentations peuvent porter de l'information fortement liée au réel, alors dans ce cas des entités en deux dimensions peuvent suffire (flèche, texte...). Deuxièmement, l'exigence de recalage parfait entre le réel et le virtuel n'est pas justifiée si l'on s'attache par exemple, à présenter des informations utiles à la tâche (consignes, instructions...). Bien qu'imparfaitement recalés, le système cognitif de l'utilisateur associera alors aisément la zone réelle ciblée par l'action aux indices virtuels. Les figures ci-dessous (Fig. 3a et 3b) illustrent la discutabilité des précédents points en proposant deux systèmes présentant des instructions virtuelles, l'une respectant totalement les principes d'Azuma, l'autre non.

Finalement, il faudra attendre 2001 pour que Fuchs et Moreau proposent dans [FMP01] de revenir sur une définition axée sur un aspect technologique non restrictif :

Réalité augmentée : « regroupe l'ensemble des techniques permettant d'associer un monde réel avec un monde virtuel, spécialement en utilisant l'intégration d'images réelles avec des entités virtuelles : images de synthèse, objets virtuels, textes, symboles, schémas, graphiques... ».

Même si cette définition nous paraît plus appropriée, voire généraliste, sans exclure la définition d'Azuma, elle exclut totalement la multimodalité. Toutefois, cette définition semble tenir compte de l'importance du lien qui unit le réel et virtuel. Nous verrons par la suite que l'auteur s'est effectivement attaché à classer en catégories fonctionnelles l'importance de la relation entre le réel et le virtuel.



(a) Coexistence parfaite du réel et du virtuel selon les principes d'Azuma.



(b) Augmentations via des entités 2D non parfaitement recalées.

FIGURE 3: Présentation d'incrustations virtuelles.

Définition des termes retenus :

Le terme « réalité augmentée » est critiquable d'un point de vue lexical ou sémantique. En effet, philosophiquement parlant, la réalité est TOUT. Augmenter la réalité pose donc un problème de sens. Mais nous ne discuterons pas plus ici cette dénomination qui a l'avantage d'être communément admise dans la littérature et commence à avoir un pouvoir évocateur fort pour le grand public. Dans notre cas, notre définition sera donc :

Réalité augmentée : « Combinaison de l'espace physique avec l'espace numérique dans un contexte sémantiquement lié ».

Cette définition, bien que globale, a l'avantage de ne pas exclure les précédentes, profitant ainsi des différentes classifications établies au cours des ans, classifications que nous allons maintenant détailler.

1.2 Taxonomies

La modalité sensorielle privilégiée par la réalité augmentée étant la vision, la littérature s'est progressivement enrichie de classification liée à l'évolution des possibilités offertes par les différents systèmes de visualisation. À ce jour, nous pouvons citer essentiellement deux systèmes de classification :

1. Fonctionnel, classifiant les relations entre le réel et le virtuel.
2. Technique, classifiant un système selon la technologie utilisée.

1.2.1 Taxonomie fonctionnelle

Proposé par Fuchs et Moreau [FMP01], ce classement a pour objectif de dresser une liste des différents types d'associations que l'on peut réaliser entre le réel et le virtuel, tout en s'affranchissant des dispositifs employés pour faire ladite association. Numérotées d'un à cinq, les auteurs précisent toutefois qu'il n'existe pas intrinsèquement de hiérarchie entre les différentes fonctionnalités, si ce n'est qu'un ordre approximativement croissant des difficultés scientifiques et techniques les caractérisant. Rappelons que cette taxonomie tient compte de l'importance de la relation entre le réel et le virtuel sur une même vue (écran, visio-casque...), ce qui exclut les combinaisons réels/virtuels sans aucun rapport sémantique. Même si les auteurs l'évoquent sans vraiment l'intégrer à leur classement, il nous semble juste d'ajouter une fonctionnalité zéro aux cinq déjà existantes.

fonctionnalité 0 : absence de relation entre le contenu des éléments virtuels et des éléments réels présentés sur une même vue. Les auteurs emploient le terme de « cadre d'affichage » pour regrouper ces différents cas. Le système créé par Sutherland en 1960 pourrait être classé en fonctionnalité zéro.

fonctionnalité 1 : « réalité documentée » ou « virtualité documentée ». Couvre toute existence d'un lien sémantique basique entre les images réelles et l'élément virtuel présenté dans deux cadres d'affichage différents. Généralement sous forme textuelle, les éléments virtuels permettent de mieux comprendre les images réelles. C'est la réalité documentée. En ce qui concerne la virtualité documentée, c'est cette fois l'incorporation de parties réelles dans une représentation virtuelle, visant à faciliter la compréhension de la scène en cours.

Fonctionnalité 2 : « réalité à compréhension ou à visibilité augmentée ». Ici, il est important de noter que les images réelles et les éléments virtuels sont dans un même cadre d'affichage, ce qui implique la nécessité de recalage entre les deux.

- La réalité à compréhension augmentée est l'incrustation d'éléments virtuels véhiculant de l'information sémantique passive. Ces éléments virtuels apportent de l'information complémentaire sur des objets réels via des titres, symboles, légendes... À noter que dans cette fonctionnalité, l'inversion entre les images réelles et les indices virtuels n'est pas envisageable.
- La réalité à visibilité augmentée permet la mise en valeur des objets réels, facilitant ainsi leur perception. Cette mise en valeur consiste généralement en l'utilisation de modèles « fil de fer » ou en transparence des objets réels. Pour assurer la correspondance géométrique, il est bien entendu nécessaire au préalable de posséder les modèles de ces objets réels.

Fonctionnalité 3 : association visuelle du réel et du virtuel en fonction du niveau d'intégration des objets virtuels dans la scène réelle, cette fonctionnalité présente deux cas :

- L'association par superposition, c'est-à-dire en incrustation (overlay) des objets virtuels par devant les objets réels. La superposition n'implique aucune occultation du virtuel par le réel.
- L'association tridimensionnelle, c'est-à-dire l'intégration d'objets virtuels dans le réel, les deux mondes paraissant cohabiter spatialement et temporellement. Toutefois, cela nécessite une gestion peu triviale des occultations entre les deux mondes. Il est à noter qu'une composition des deux cas précédemment cités est possible, voire même devenue très courante à l'heure actuelle.

Fonctionnalité 4 : association comportementale du réel et du virtuel. Après l'intégration tridimensionnelle, de plus en plus nous retrouvons l'intégration des propriétés physiques (masse, gravité, élasticité...) des objets virtuels, « interagissant » ainsi avec le réel. Bien entendu, cette interaction géométrique et physique nécessite au préalable une connaissance à priori de la scène réelle⁶.

Fonctionnalité 5 : « substitution du réel par le virtuel » ou « réalité virtualisée ». Ici, il est question d'offrir la possibilité de substituer la scène réelle (visuellement) par la scène virtuelle équivalente. Cependant, l'ensemble des propriétés de la scène réelle doit être connu et modélisé. Il devient alors possible de présenter à l'utilisateur l'un ou l'autre des mondes, soit à partir du même point de vue, soit à partir d'un point de vue alternatif dans le monde virtualisé, ceci afin de faciliter la compréhension de l'environnement.

1.2.2 Taxonomie technique

En 1994, Milgram cherche à classifier la relation qu'il peut exister entre un système de réalité augmentée et la technologie utilisée. Il introduit alors le terme de réalité mixte, concept désignant l'hybridation dans un continuum reliant le monde physique au monde numérique (Fig. 4).

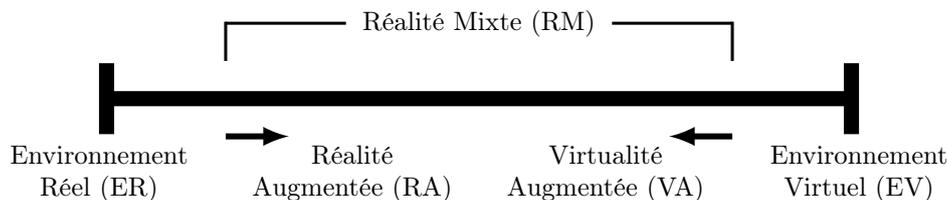


FIGURE 4: Continuum Réalité-Virtualité.

Ce continuum traduit en pratique le degré de mélange entre réel et virtuel, surtout en termes de modalités visuelles. C'est la dominance de l'environnement sur un autre qui permet de distinguer si le système est dit de réalité augmentée ou de virtualité augmentée. Dans cette représentation, la réalité augmentée est donc une composante de la réalité mixte. Partant de ce principe, Milgram

⁶. Il est intéressant de noter ici que la superposition, occultant partiellement la réalité, « diminue » la perception du réel.

propose une taxonomie de facteurs qui importe pour catégoriser les différentes technologies d’affichage. Elle est basée sur trois principes (ou dimensions), eux-mêmes représentés par une échelle de graduation :

- Le degré de modélisation (connaissance) du monde.
- Le degré de fidélité de reproduction du monde.
- Le degré d’immersion de l’utilisateur (ou métaphore de présence).

Degré de connaissance du monde :



FIGURE 5: Degré de connaissance du monde.

Ici, cette dimension s’intéresse à caractériser le degré de modélisation que l’on peut avoir. À l’extrémité gauche se trouve un monde non modélisé, à l’extrémité droite le monde totalement modélisé (cas de la réalité virtuelle par exemple). En ce qui concerne le monde partiellement modélisé, nous distinguons une granularité le caractérisant par les termes « où », « quoi » et « où et quoi ». « Où » se rapportant au fait de connaître la localisation d’un objet, sans pour autant connaître sa nature. « Quoi » répondant à l’inverse au fait d’avoir une connaissance sur l’objet regardé sans pour autant être forcément capable de le localiser⁷. Enfin, plus le monde tend à être complètement modélisé, plus il devient possible de répondre simultanément aux critères « où et quoi ».

Degré de fidélité de représentation du monde :



FIGURE 6: Degré de fidélité du monde en fonction de la technologie d’affichage et du rendu.

Cette dimension s’intéresse au degré de réalisme induit par deux facteurs fortement liés que sont la fidélité du rendu et le mode d’affichage associé. En fonction de l’adéquation du mode d’affichage avec la vision humaine, les possibilités sur ce dernier tendent à permettre un rendu visuel des entités virtuelles plus ou moins réelles.

7. Ce cas ne respecterait donc pas la troisième caractéristique d’Azuma exigeant un recalage réel/virtuel.

Degré d’immersion de l’opérateur :

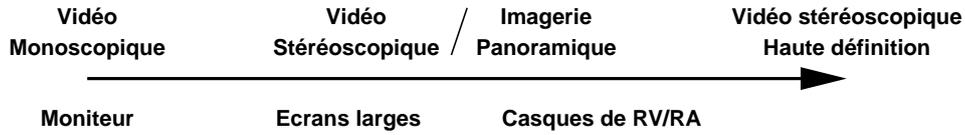


FIGURE 7: Degré d’immersion de l’utilisateur en fonction de la technologie d’affichage.

Cette dernière dimension présente la corrélation qui existe entre la technologie d’affichage et le degré d’immersion qu’un utilisateur pourra ressentir. À l’extrémité gauche, le degré d’immersion supposé est moindre compte tenu du mode d’affichage WoW (Window on World), contrairement à l’extrémité droite, censée renforcer le sentiment de présence de l’utilisateur (cas des systèmes de réalité virtuelle).

Cependant ici, le degré de fidélité de représentation du monde peut être élevé. Nous pouvons donc nous poser la question de savoir si forcément l’immersion à cette extrémité est fortement dégradée, ou tout du moins si le degré d’immersion est obligatoirement corrélé au degré du rendu.

1.2.3 Autres taxonomies

Il est important de noter que les taxonomies fonctionnelles et techniques que nous avons présentées ne sont pas les seules que l’on peut trouver dans la littérature. Pour le lecteur, nous citons celles proposées par Bérard [Bér94]. Celui-ci définit un espace organisé suivant deux dimensions, « action » et « objet », admettant deux valeurs, « réel » et « virtuel », l’intérêt étant de situer dans un même espace la réalité augmentée et la réalité virtuelle par rapport au monde réel (Fig. 8).

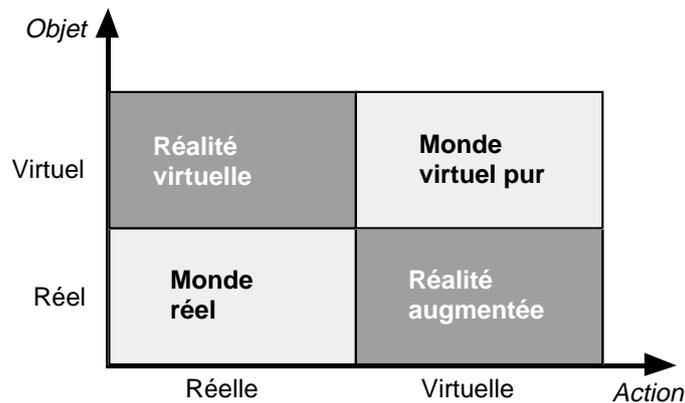


FIGURE 8: Taxonomie des différents mondes vue par Bérard.

Nous pouvons également citer la taxonomie proposée par Mackay [Mac96] où les applications de réalité augmentée sont classées suivant la stratégie adoptée : augmenter l’utilisateur, augmenter un objet, augmenter l’environnement (Tab. 2).

Type d'augmentation	Approche	Technologie
L'utilisateur	Porter des dispositifs sur le corps	Casque de RV, gants de données...
Les objets physiques	Instrumenter les objets	GPS, centrale inertielle...
L'environnement proche des objets et de l'utilisateur	Projeter des images et les enregistrer	Caméra vidéo, projecteur vidéo...

TABLE 2: Taxonomie de la réalité augmentée selon Mackay.

Enfin, Dubois nous propose une classification [DN00] basée en deux points. Tout d'abord, une approche déterminant l'interaction (Fig. 9a) :

- En réalité augmentée, l'interaction avec le monde réel est augmentée par l'ordinateur.
- En virtualité augmentée, l'interaction avec l'ordinateur est augmentée par des objets et actions dans le monde réel.

Puis, une approche fonction de deux types d'augmentation (Fig. 9b) : exécution et évaluation, appliquées aux précédentes interactions.

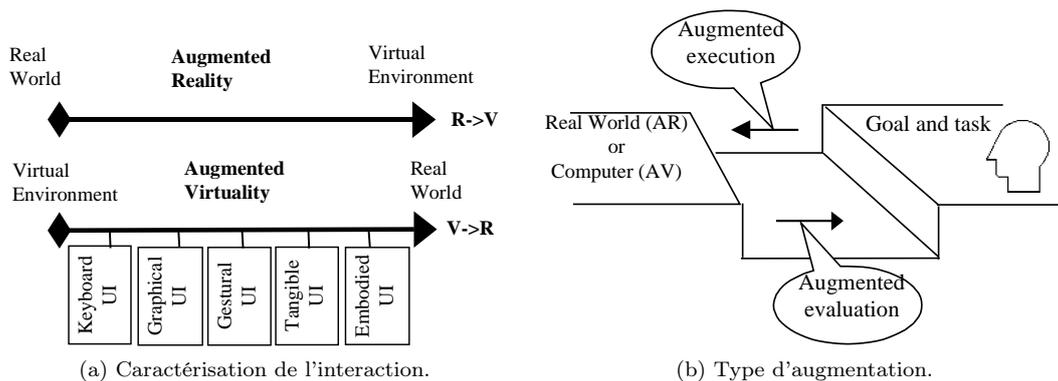


FIGURE 9: Taxonomie de Dubois.

Nous ne développerons pas plus ces différentes taxonomies, nous invitons cependant le lecteur à consulter les références s'y rapportant.

2 PRINCIPES TECHNOLOGIQUES

Les systèmes de réalité augmentée sont généralement construits suivant une architecture matérielle commune, et ce qu'ils soient fixes ou mobiles. D'ordinaire, une caméra filme la scène, un ordinateur génère des entités virtuelles en fonction de ce qui est perçu⁸ et des informations relatives à différents capteurs (position, inclinaison...), puis le résultat est visualisé par l'utilisateur via un dispositif d'affichage. Nous allons dans cette partie nous attacher à présenter le « design » général d'une application de réalité augmentée avant d'en décrire de manière plus précise les éléments constituant la base même.

2.1 Les différentes briques technologiques

Dans [BR05], Bimber illustre sous forme de briques technologiques la constitution générale d'un système de réalité augmentée (Fig. 10).

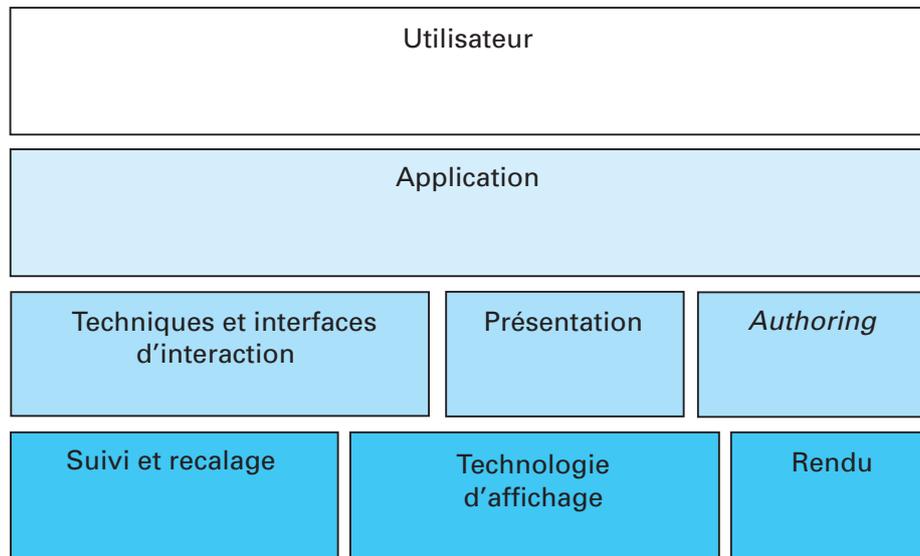


FIGURE 10: Briques constituant un système de réalité augmentée.

Le design général est composé de quatre couches de composants :

Niveau de base : c'est le plus bas niveau du système, incluant les composants fondamentaux, tant logiciel que matériel, à savoir le suivi et le recalage, les technologies d'affichage jusqu'au rendu des images générées. La plupart des efforts de recherche en réalité augmentée à l'heure actuelle se concentrent à ce niveau. Nous allons détailler de manière plus précise ces trois briques dans les paragraphes suivants (2.2, 2.3 et 2.4).

Par-dessus ce niveau de base se trouvent des éléments qui sont dérivés pour la plupart des techniques de la réalité virtuelle, adaptées à la réalité augmentée.

8. Si on considère uniquement des informations relatives à la caméra, alors la réalité augmentée devient un problème d'analyse d'image.

Niveau intermédiaire : ce niveau, principalement implémenté sous forme logicielle, s’occupe des méthodes d’interactions avec l’utilisateur, de la façon de présenter les informations virtuelles et des moyens pour les mettre en scène :

- Les Interactions avec l’utilisateur : les moyens modernes de sélectionner ou de manipuler des éléments virtuels utilisent le paradigme WIMP (Window, Icon, Menu, Pointing), paradigme parfaitement adapté aux interactions 2D des systèmes d’exploitation actuels. L’un des défis de la recherche dans le domaine de l’IHM est donc de trouver des paradigmes mieux disposés à une utilisation dans l’espace physique. C’est le cas notamment des TUI (Tangible User Interface) dont l’objectif est d’utiliser les objets réels en vue d’interagir avec les éléments du monde numérique [SH10]. On pourra citer les travaux d’Henderson et Feiner [HF10] visant à utiliser l’affordance de n’importe quel objet réel comme interface avec le virtuel (Fig. 11), procédé que les auteurs nomment « Opportunistic Controls ».

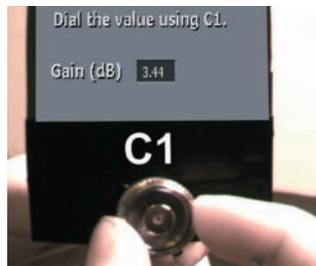


FIGURE 11: Utilisation d’un collier de connecteur d’antenne comme bouton rotatif en vue de modifier une valeur dans une interface virtuelle.

- La Présentation : cette brique s’occupe de la façon dont sera présenté l’ensemble des informations, qu’elles soient textuelles ou graphiques, augmentant le monde réel. Donnons l’exemple de l’étude faite par Gabbarg [GSH⁺07] examinant l’influence du style et de la couleur d’un texte affiché en surimpression de la réalité. Il propose ainsi un dispositif de présentation de l’information textuelle fonction des conditions d’illuminations et des textures de la scène réelle (Fig. 12).

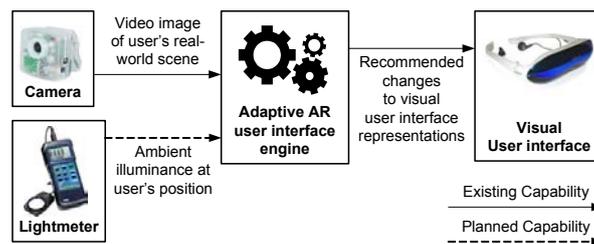


FIGURE 12: Proposition de dispositif visant à modifier la façon de présenter l’information en fonction des conditions de la scène.

- L’Authoring : c’est la scénarisation du monde virtuel associé au monde réel. L’intérêt premier est de déterminer une trame préétablie des différents éléments à « jouer » pour une localisation donnée. Il peut s’agir alors d’une animation qui guidera l’opérateur dans l’accomplissement de sa tâche. Cette mise en scène, non triviale, nécessite d’utiliser des méthodes et outils adaptés comme le logiciel composAR [SLB08] ou encore BuildAR, développés au HitLab [Hit08], remplaçant avantageusement les techniques plus souples mais plus fastidieuses utilisant des bibliothèques spécialisées comme OpenSceneGraph.

Niveau Application : ce niveau, entièrement réalisé de façon logicielle, est l'application de réalité augmentée en elle-même, application complètement tournée vers l'usage dévolu à la spécificité de la tâche. À l'heure actuelle, il n'est pas rare de voir des applications de réalité augmentée développées grâce à des API⁹ ou SDK¹⁰ dédiées aux jeux vidéos comme Irrlicht, OGRE ou SourceSDK. L'avantage est de pouvoir profiter au maximum des puissantes fonctionnalités offertes comme les interactions physiques 3D, la détection de collision ou le rendu 3D complexe.

Niveau Utilisateur : en définitive, une application de réalité augmentée doit faire face à l'utilisateur final. Impossible donc aujourd'hui de ne pas intégrer les spécificités de celui-ci dans la boucle de conception de l'application. En effet, les facteurs sociaux, la résistance au changement ou encore le niveau d'exigence du domaine ciblé sont des éléments primordiaux à prendre en compte pour l'acceptance de l'utilisateur vis-à-vis de l'application. Généralement, cette étude de la compréhension des mécanismes humains est à imputer aux disciplines scientifiques formant les sciences cognitives.

2.2 Le principe du rendu

Le rendu dépend de la modalité ciblée. À la réalité augmentée, il est souvent implicitement associé le rendu visuel. En informatique, le rendu est un processus permettant de passer d'une scène géométrique 3D à une image 2D en respectant un point de vue, les textures, l'illumination et les ombres associées. Le but ultime étant de ne pas pouvoir distinguer les objets virtuels des objets réels lors du mélange des deux mondes. Notons tout de même qu'une application de réalité augmentée peut ne pas viser cette perfection suivant le degré de réalisme souhaité (cf. Taxonomie technique de Milgram).

Cependant, pour que le rendu visuel soit correct, il doit respecter plusieurs règles. Au nombre de quatre, les trois premières sont fondamentales pour toute application interactive :

- La contrainte de temps réel, impliquant l'utilisation de méthodes rapides de génération d'images. À l'heure actuelle, l'utilisation de techniques d'illumination globale comme le Ray-Tracing est donc à exclure.
- L'alignement réel-virtuel, plus communément appelé « recalage » dans la littérature, est la mise en correspondance entre les objets réels et virtuels, donnant ainsi l'illusion d'une coexistence entre les deux mondes (Fig. 13). Techniquement, cela revient à déterminer les propriétés du point de vue de l'observateur (caméra ou oeil) dans la réalité. Ces paramètres sont ensuite appliqués à une caméra virtuelle dans le monde numérique. Les deux points de vue étant maintenant identiques, l'ordinateur peut finalement générer des entités virtuelles semblant coexister avec le monde physique (voir paragraphe 2.4.2 sur les différentes techniques de recalage).

9. « Application Programming Interface »

10. « Software Development Kit »



FIGURE 13: Alignement réel-virtuel (extrait de [SD06]).

- La cohérence spatio-temporelle, résultante de l’alignement réel-virtuel au cours du temps, ainsi que la gestion des occultations et des interpénétrations du réel avec le virtuel (Fig. 14). À l’heure actuelle, la gestion des occultations est un domaine très actif de la recherche en réalité augmentée. On citera pour l’exemple les références comme [LB00] ou [MS09].



FIGURE 14: Cohérence spatio-temporelle (extrait de [SD06]).

- La cohérence photométrique, c’est-à-dire la prise en compte dans la scène virtuelle de l’illumination et des ombres résultant de la scène réelle (Fig. 15). Le résultat final sera en accord avec le degré de réalisme choisi tout en respectant la vitesse de traitement associé (contraintes temps réel).



FIGURE 15: Cohérence photométrique (extrait de [SD06]).

En réalité augmentée, une fois le rendu visuel établi, le résultat peut être visualisé suivant deux modes distincts, à savoir en vision directe ou bien en vision indirecte :

- La vision indirecte regroupe les techniques où la vue du réel est restituée par une caméra sur un afficheur généralement vidéo. Ainsi, l'utilisateur ne perçoit pas directement réalité et virtualité, le mélange étant effectué par un ordinateur puis présenté sur un écran (Fig. 17).
- Au contraire, la vision directe est la technique qui permet à l'utilisateur de percevoir directement le réel sans processus de traitement. Ceci est rendu possible en utilisant des systèmes optiques placés devant les yeux (Fig. 16). Ici, c'est donc le système optique qui effectue le mélange entre les rayons lumineux de la scène réelle avec les rayons lumineux provenant de l'afficheur vidéo miniature présentant les entités virtuelles.

2.3 Les technologies d'affichage

Nous avons vu précédemment que le rendu visuel pouvait se faire soit en mode indirect, soit direct, le mode étant fonction de la technologie d'affichage choisie. Contrairement à la réalité virtuelle, qui nécessite des dispositifs permettant une immersion correcte avec de grands champs de vision (casque immersif ou HMD) ou bien des dispositifs de grande taille comme les CAVE (Cave Automatic Virtual Environment), la réalité augmentée doit permettre à l'utilisateur de garder pieds dans son environnement réel. Ce sont donc les aspects mobile et non intrusif qui sont privilégiés et ont amené l'utilisation de dispositifs types lunettes « See-Through » ou « Window On World » (Tablet-PC, assistants personnels...).

Ainsi, Azuma [ABB⁺01] classe les dispositifs de visualisation en trois types distincts :

1. Head Worn Display (HWD) ou Head Mounted Display (HMD). Ce sont tous des dispositifs de type casque ou lunettes qui sont portés par l'utilisateur. Ces casques sont étiquetés comme « See-Through » pour permettre à l'utilisateur de percevoir le monde réel, soit en vision directe, soit indirecte.
2. Handheld Display, c'est-à-dire tout dispositif offrant une fenêtre augmentée sur le monde. Ce sont des appareils types Smartphone (iPhone, Blackberry...) qui sont de plus en plus utilisés à l'heure actuelle.
3. Projective Display, dispositifs utilisant des projecteurs augmentant les objets réels par projection directe d'informations sur les objets eux-mêmes.

2.3.1 Les « Head Worn Displays »

Les HWD sont des dispositifs d'affichage portés sur la tête de l'utilisateur, les augmentations étant présentées soit sur un oeil (système monoculaire) soit sur les deux (système bi-oculaire si les images présentées aux deux yeux sont identiques, système binoculaire si elles sont différentes et forment une paire stéréoscopique).

Les HWD sont classés suivant la façon dont ils combinent la vue du réel avec le virtuel. Nous distinguons principalement les dispositifs de type Optical-See-Through (OST) qui apportent une vision directe (Fig. 16) du monde réel sur laquelle est surimprimé le contenu virtuel. Ici, virtuel

et réel sont donc combinés de façon optique par l'intermédiaire de prisme(s) et/ou de miroir(s) semi-transparent(s).

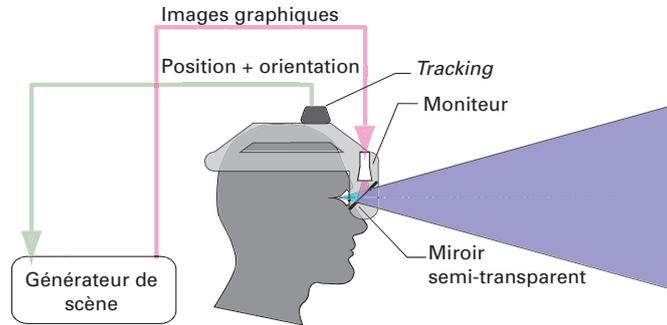


FIGURE 16: Réalité augmentée en vision directe (Optical See-Through) [MR08].

D'autre part, nous distinguons les dispositifs de type Video-See-Through (VST) qui apportent une vision indirecte (Fig. 17) du monde réel. Le monde réel est perçu grâce à une ou deux caméras disposées sur le casque, le virtuel étant superposé numériquement sur le signal vidéo avant d'être présenté à l'utilisateur via les afficheurs vidéo du VST.

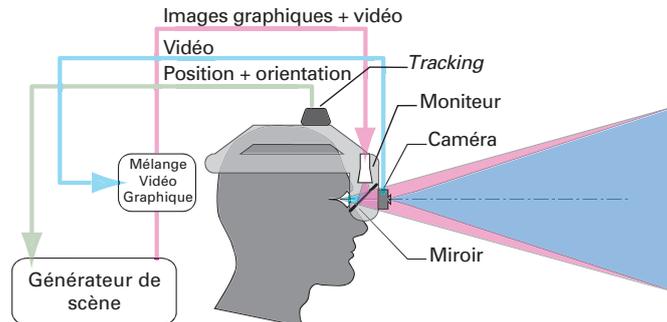


FIGURE 17: Réalité augmentée en vision indirecte (Video See-Through) [MR08].

Le tableau 3 présente une sélection de plusieurs dispositifs suivant certains critères technologiques. Parmi ces critères, on retrouve le champ de vision horizontal offert, fortement inégal en fonction des constructeurs, mais aussi le type d'affichage (stéréo ou mono). Mais d'autres critères sont aussi primordiaux. La résolution par exemple, surtout si elle est faible, sera à prendre en compte en fonction de l'importance du niveau de détail recherché dans les images à afficher. Le poids du dispositif peut aussi influencer sur la durée d'utilisation, certaines études comme [ZAK05] mettant en avant que le port de HWD peut entraîner des problèmes de charge physique (musculo-squelettiques). Enfin, la technologie d'affichage revêt une importance qui peut aussi influencer sur le choix d'un HWD, notamment en termes de luminosité ou de contraste. La plupart des dispositifs utilisés sont des afficheurs LCD (Liquid Cristal Display) ou LCOS (Liquid Cristal On Silicon), mais on trouve aussi des « Retinal Displays », c'est-à-dire des diodes laser de faible puissance formant par balayages une image directement sur la rétine de l'utilisateur (une bonne luminosité étant ainsi garantie, au détriment cependant d'un affichage monochrome).

Tableau 1 – Différents <i>head worn displays</i> utilisables en réalité augmentée						
						
Marque	Saabtech	Microvision	Myvu	Kaiser	Nvis	Trivisio
Modèle	Addvisor 150	Nomad	<i>Personal media player</i>	Proview XL 40/50 STm	nVisor ST	ARvision-3D HMD ou Goggles
Type	Casque	Monocle	Lunettes	Casque	Casque	Casque ou lunettes
Type <i>see-through</i>	optique	optique	optique	optique	optique	vidéo
Stéréo	oui	non	non	oui	oui	oui
Résolution/œil	1 280 × 1 024	800 × 600	320 × 240	1 024 × 768	1 280 × 1 024	800 × 600
Champ de vue (<i>overlap</i>)	46°	29°	10°	43°	50°	40°
Type couleur	<i>True color</i>	Monochrome rouge	<i>True color</i>	Monochrome vert	<i>True color</i>	<i>True color</i>
Depth couleurs	24 bits	5 bits	24 bits		24 bits	18 bits
Contraste	30:1			25:1	100:1	
Distance image	0,7 à 25 m	30 cm à l'infini	2 m		infini	
Type affichage	LCOS				LCOS	
Standard vidéo	SXGA	VGA				
Masse	< 1 kg	230 g		800 g	1,3 kg	230 g

TABLE 3: Différents HWD utilisés en réalité augmentée.

Le choix d'un HWD et donc une étape critique, devant tenir compte des limitations des dispositifs suivant l'application envisagée. Afin d'avoir une meilleure idée du choix technologique à effectuer suivant le domaine d'application visée, [KC98] propose un tableau récapitulatif des critères de performance minima acceptables (Tab. 4).

Application	FOV	Resolution	Color Depth	Head Mobility	Tracking	Stereo / Accommodation	Design Form
Mobil Data	30°	high	monochrome OK, color better	high	none	none	Monocle - very small / light, Flip-up, Binocular non-HMD OK
Surgery	30°	very high	very high	medium	highly registered, medium working volume, moderate lag	S: very important A: helpful	Look around, Video see-through, Flip up, Comfort for long wear times
Remote Manipulation	60°	medium	medium	low-medium	accurate, varied working volume, moderate lag	S: very important A: unimportant	Short wear times, Mechanical tracking
Pilot Optical See-Through	120°	high	high	very high	highly registered, small working volume, minimal lag	none	Integrate with helmet, Mechanical tracking, Robust
Walk Through / Immersion	≥ 120° (h) ≥ 60° (v)	medium	high	medium	medium accuracy, large working volume, moderate lag	S: important A: helpful	Very light, Untethered, Large vertical FOV
Entertainment	60°	medium	medium	high	low accuracy, varied working volume, moderate lag	S: useful A: unimportant	robust

TABLE 4: Critères de performance minima pour un HMD suivant l'application.

2.3.2 Les « Handheld Displays »

L'idée des Handheld Displays est d'offrir un dispositif portable à l'utilisateur présentant une « fenêtre augmentée » sur le monde (Fig. 18). Bien qu'à l'heure actuelle, le grand public soit en train de découvrir la réalité augmentée grâce aux téléphones portables, on distingue en tout trois classes

de Handheld Displays ; les Smartphones, les assistants personnels (PDA) et les Tablet-PC (ou UMPC¹¹). Tous constituent un compromis entre taille, poids, puissance et coût. Les Smartphones sont relativement bien adaptés à la réalité augmentée en extérieur, car on les trouve de plus en plus équipés de GPS, de boussole numérique, caméra vidéo et dispositifs de communication (WiFi, Bluetooth, 3G...). Mais leur puissance de calcul étant limitée, ils ne sont pas de bons candidats pour des applications de réalité augmentée 3D riches. Parmi les Smartphones les plus connus, on peut citer l'iPhone d'Apple, avec des applications de réalité augmentée comme Wikitude, AroundMe ou Layar.

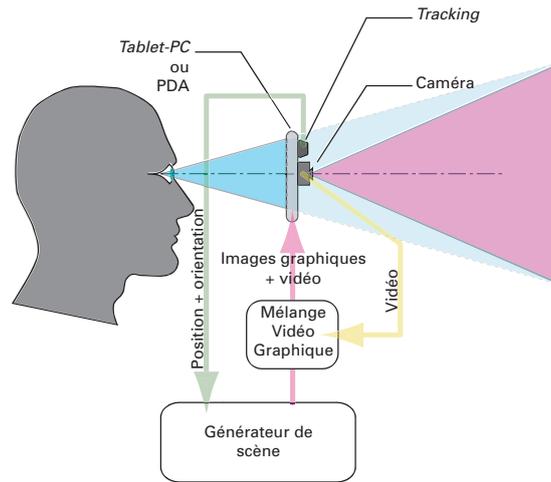


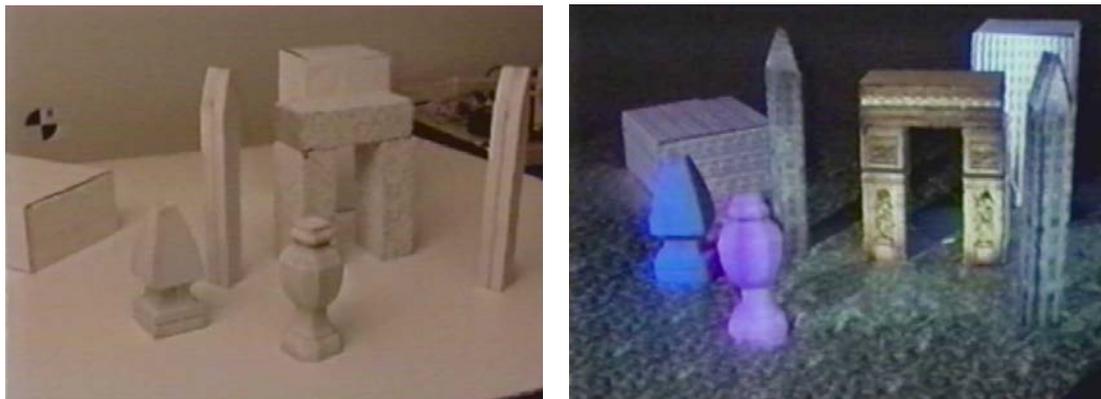
FIGURE 18: Dispositif « Handheld Display » [MR08].

Les Tablet-PC, eux, à l'inverse des Smartphones, ont une puissance de calcul nécessaire et suffisante pour tout type d'application en réalité augmentée, à la condition de les équiper avec le matériel adéquat [RD06]. Ils ont toutefois l'inconvénient de rarement pouvoir être tenus d'une seule main, leur encombrement et leur poids ne permettant pas une utilisation prolongée. Enfin, les assistants personnels sont un compromis entre les Smartphones et les Tablet-PC. Mais la montée en puissance de calcul des Smartphones et leur succès depuis l'intégration des écrans tactiles font des assistants personnels des plates-formes de plus en plus délaissées.

2.3.3 Les « Projective Displays »

Les « Projective Displays » sont des dispositifs utilisant des projecteurs vidéo pour augmenter directement un objet réel. Les premiers travaux utilisant cette technique permettaient d'augmenter une surface plane [RBLM97]. Par la suite, divers travaux ont permis d'augmenter n'importe quelle partie d'une pièce (tables, sols, murs...), soit en utilisant un miroir mobile pour orienter la projection [Pin01], soit en utilisant un vidéoprojecteur monté sur un système motorisé [BSC04]. Mais très vite des travaux sont apparus en ne se limitant plus seulement à des projections sur des surfaces planes, mais plus généralement à des maquettes physiques [RWC99] (Fig. 19). Le lecteur remarquera que pour recourir à ce type d'augmentation, il est nécessaire d'avoir une connaissance géométrique des éléments à augmenter.

11. Ultra Mobile PC.



(a) Une maquette physique sans augmentation. (b) La même maquette augmentée par vidéo-projection.

FIGURE 19: Réalité augmentée projetée.

Plus récemment, courant 2008, la miniaturisation a permis l'émergence d'une nouvelle catégorie de vidéoprojecteur, les picoprojecteurs. Ces dispositifs ultras compacts et légers ont relancé l'intérêt de la projection dans le domaine de la réalité augmentée, notamment pour les dispositifs mobiles. On citera en ce sens les travaux de Pranav Mistry et de son dispositif baptisé « Sixthsens » [MMC09].

Enfin, il convient d'évoquer un dispositif de projection développé par Jannick Rolland et son équipe, dispositif appelé « Head Mounted Projection Display » [RBHL⁺05]. Ce dispositif est par conséquent un HWD utilisant un mini-projecteur afin d'augmenter ce qui se trouve en face de l'utilisateur (Fig. 20).

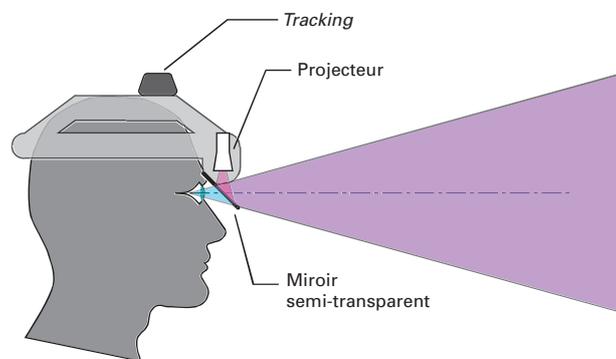


FIGURE 20: Head Mounted Projection Display [MR08].

2.4 Suivi et recalage

La réalité augmentée a pour objectif de faire coïncider visuellement le virtuel sur le physique, et ce en temps réel. Pour ce faire, il faut être capable de connaître de manière précise et récurrente la localisation (position et orientation) de la scène réelle (ou d'un objet) par rapport à un point de vue (celui de l'utilisateur par exemple). Ce problème de suivi (Tracking) est fondamental, mais pose encore problème à l'heure actuelle, même si de nombreuses approches existent pour le réaliser. Notons que le manque de précision qu'implique ces méthodes influe sur la précision du recalage (Registration) entre réel et virtuel.

2.4.1 Le suivi (Tracking)

Le Tracking est le processus qui détermine la position-orientation d'un objet dans l'espace physique. Idéalement, le Tracking doit être effectué de façon incrémentale, être temporellement stable et précis. Dans la réalité, différents types d'informations (visuelles ou physiques via des capteurs) sont utilisés selon les contraintes liées à l'application (réalisme, rapidité, précision...) Les techniques de Tracking sont communément classifiées en deux taxonomies distinctes ; suivant le type de relations entre objets-capteurs et suivant la technologie employée.

La première est elle-même constituée de deux classes, à savoir :

- Le Tracking Outside-In : le sujet à suivre évolue dans un environnement instrumenté par des capteurs. Le sujet est alors généralement équipé de dispositifs (actif ou passif), comme des récepteurs infrarouges par exemple, afin de permettre aux différents capteurs de déduire sa position-orientation. Ce type de Tracking est donc particulièrement adapté pour des applications en intérieur (indoor), car le volume de suivi est limité, volume allant de quelques mètres carrés à quelques centaines de mètres carrés au maximum. Les principales technologies de capteurs utilisés pour le Tracking Outside-In sont les technologies à ultrasons, magnétiques et optiques.
- Le Tracking Inside-Out : contrairement au Tracking Outside-In, ici c'est le sujet (objet ou utilisateur) qui est muni de capteurs. La position-orientation du sujet est donc déterminée de façon relative en fonction d'éléments (ancres visuelles...) situés dans l'environnement. Ce type de Tracking est donc mieux adapté pour des applications mobiles à l'extérieur (Outdoor). Généralement, les technologies employées sont optiques, c'est-à-dire utilisant une ou plusieurs caméras chargées de suivre soit des cibles, soit des primitives géométriques de l'environnement. On utilise également de plus en plus des dispositifs de localisation extérieure comme les GPS ou les D-GPS (Differential GPS).

La seconde classification, d'ordre technologique, ordonne les différents capteurs utilisés pour le suivi en cinq catégories ; mécanique, magnétique, ultrasons, inertiel et optique. Leur objectif est de mesurer le mouvement du sujet en une série de six grandeurs (trois positions et trois angles) à une fréquence d'échantillonnage compatible avec la précision recherchée :

- Les technologies mécaniques : historiquement utilisé par Sutherland, cela se présente sous forme de bras articulé utilisant des dispositifs de mesures analogiques ou numériques. Mais ces technologies sont aujourd'hui très peu utilisées malgré leurs performances (grande résolution, précision et temps de réponse), car elles limitent fortement le volume d'utilisation et les mouvements de l'utilisateur.
- Les technologies magnétiques : trois champs électromagnétiques orthogonaux sont émis par une base fixe, leurs distorsions étant mesurées par trois capteurs magnétiques, déterminant ainsi la position orientation du porteur. Bien que d'une grande précision, la présence d'objets ferromagnétiques dans l'espace de suivi perturbe fortement la mesure. De plus, on trouve une restriction sur le volume de travail (quelques mètres cubes) liée au fait que la puissance du champ requise n'évolue pas linéairement en fonction du volume à couvrir. On pourra citer le système développé et vendu par Polhemus utilisant des champs magnétiques alternatifs sinusoïdaux (AC), puis le système développé par Ascension Technologie utilisant des courants continus pulsés (DC) supposés réduire la sensibilité à la présence d'objets ferromagnétiques.

- Les technologies à ultrasons : il s’agit de capteurs télémétriques utilisant la variation du « temps de vol¹² » pour déterminer la distance séparant la source ultrasonore d’un élément de l’environnement. En utilisant ce principe de mesure du temps aller-retour par trois capteurs et respectivement capté par trois microphones, il est possible de déterminer les 6 degrés de liberté composant la position-orientation de l’objet à suivre. Toutefois, même si ces technologies sont peu onéreuses et que leur précision est de l’ordre du centimètre, elles peuvent se heurter à des erreurs de mesure principalement dues au phénomène d’absorption ou de réflexions multiples (échos) des matériaux environnants. On notera également une fréquence d’échantillonnage faible ne permettant pas une utilisation dans des applications nécessitant une haute réactivité.
- Les technologies inertielles : une centrale inertielle est un dispositif de mesures composé généralement d’un accéléromètre trois axes, d’un gyroscope trois axes (mesurant les taux de rotation) et d’un magnétomètre trois axes (mesurant le champ magnétique terrestre). Ce type de technologie permet de déterminer avec une grande précision l’orientation, moyennant l’utilisation de filtres numériques en charge d’éliminer les dérives de mesures dues aux perturbations extérieures. Bien que théoriquement il soit possible à partir de l’accélération de calculer le déplacement relatif du sujet lié à la centrale, les mesures de [Mai07] ont montré qu’à l’heure actuelle il était difficile d’obtenir sans dérive une position pendant plus de 480 ms.
- Les technologies optiques : il existe différentes façons de faire du suivi en employant ces méthodes. On distingue notamment le Tracking optique basé multi-caméra pour trianguler la position dans l’espace de points mesurés. Ces systèmes utilisent généralement des caméras infrarouges surmontées de projecteurs infrarouges. Cette lumière est projetée sur des matériaux réfléchissants fixés sur l’utilisateur. Les caméras sont alors utilisées pour enregistrer les mouvements des marqueurs et ainsi reconstruire la position-orientation du sujet « tracké ». À noter que certains systèmes suppriment les projecteurs infrarouges et les matériaux réfléchissants en équipant directement le sujet avec des diodes infrarouges. Ce type de système optique est très utilisé en réalité virtuelle ou encore dans l’industrie cinématographique et du jeu vidéo pour animer les mouvements de personnages 3D. On pourra citer pour l’exemple la société VICON qui utilise ce procédé pour son produit de « motion capture ». La seconde façon de faire du Tracking optique est d’utiliser l’analyse d’images. Ce Tracking basé vision exploite une ou plusieurs caméras pour suivre soit des cibles codées, soit des éléments de l’environnement. Les cibles codées sont apparues au milieu des années 90 grâce aux travaux de Jun Rekimoto [Rek98]. Nommées Matrix puis Cybercode en 2000 [RA00], ces cibles (Fig. 21) à disposer sur l’objet à suivre représentent des marqueurs plans sur lesquels sont imprimés des motifs uniques. L’avantage est double. Cela permet d’une part de déterminer la position-orientation de la caméra qui les observe, et d’autre part de pouvoir associer un code unique qui permet de choisir l’augmentation en fonction de la cible. Il faut néanmoins voir en permanence les cibles pour ce faire. Toutefois, la taille des cibles est un problème qui peut influencer sur le recalage. En effet, plus la dimension de la cible est réduite, plus il est difficile dans l’image fournie par la caméra d’extraire les points la composant. L’une des solutions palliant cet inconvénient consiste à utiliser plusieurs cibles réparties sur l’objet augmenté. Par la suite, d’autres systèmes de cibles codées ont été proposés, notamment le devenu célèbre ARTToolKit [KB99](Fig. 21b), ou encore VisTracker [NF02] dont la spécificité réside dans la forme circulaire de ses cibles (Fig. 21c). Mais l’utilisation de cibles peut s’avérer très contraignante dans des environnements publics ou extérieurs. En ce sens, nombre de travaux sont menés

12. TOF : Time Of Flight.

à l'heure actuelle pour s'en dispenser. L'objectif n'est plus de suivre des cibles, mais de suivre des primitives (connues ou non) dans l'environnement comme des courbes, des points ou des contours..., on citera pour l'exemple [HS88], [Low04], [Can86] ou [BETVG08]. L'inconvénient majeur de ces techniques est que le temps de traitement n'est pas garanti, contrairement aux méthodes basées cibles¹³, ce qui peut être problématique pour une application temps réel.

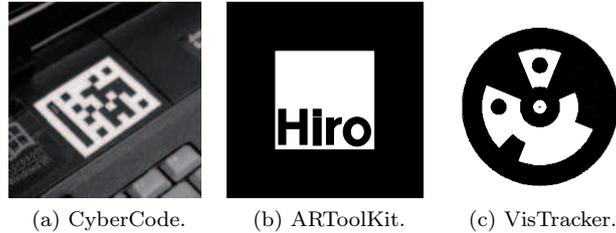


FIGURE 21: Différentes cibles codées.

2.4.2 Le recalage (Registration)

Le recalage entre le réel et le virtuel est un problème complexe qui aujourd'hui encore n'est pas totalement résolu. L'alignement entre les objets du monde virtuel et ceux du monde physique doit se faire de telle manière à ce que l'utilisateur ait l'impression d'une coexistence parfaite entre ces deux mondes. Dans la réalité, ce besoin de précision dans la superposition dépend surtout du type de l'application visée, comme cela doit être le cas en médecine ou en mécanique. À défaut, nombre d'études comme [LA08] mettent en avant les problèmes qui en découlent, tant pour le Tracking que pour l'utilisabilité de l'application. Les erreurs d'alignement sont définies en deux types, statique et dynamique [Azu97] :

1. L'erreur statique (alignement spatial) qui apparaît lorsque le point de vue de l'utilisateur est fixe (ainsi que les objets de l'environnement). Holloway, dans [Hol95, Hol97], définit quatre types de cause amenant cette erreur :
 - (a) La distorsion optique, problème lié aux propriétés des lentilles des optiques de caméras. On pourra citer le problème de distorsion radiale (image « concavisée » ou « sphérisée »).
 - (b) L'erreur du système de suivi, problème lié aux valeurs de sortie générées par lesdits systèmes. Ces erreurs sont généralement difficiles à caractériser.
 - (c) L'erreur d'alignement mécanique, c'est-à-dire la divergence qui existe entre les spécifications de matériels ou de modèles vis-à-vis des propriétés réelles du système. Par exemple, citons l'erreur d'alignement qu'il peut y avoir entre une caméra et un afficheur dans un casque VST (problème de parallaxe).
 - (d) La mauvaise estimation des paramètres de visualisation (champ de vue, distance du capteur, distance interpupillaire...), paramètres essentiels qui seront reportés dans la matrice de visualisation utilisée pour générer les images d'augmentation.

¹³. La détection de cible requiert des traitements identiques, quelle que soit la cible (binarisation, identification du marqueur, calcul de sa position-orientation). Ceci assure un traitement déterministe et en temps fini. Les méthodes basées primitives dépendant fortement du nombre de primitives présentes dans l'image, il est difficile de pouvoir garantir un recalage en un temps donné.

2. L'erreur dynamique (alignement temporel) est l'erreur de recalage apparaissant durant le mouvement, faisant ainsi perdre l'illusion de coexistence entre le monde physique et numérique. Cela est dû aux retards temporels inhérents à la chaîne de traitement du système de réalité augmentée, car chaque élément le composant nécessite un temps minimal (latence relative) pour fournir un résultat.

Il existe plusieurs solutions pour réduire cette latence :

- Réduire la latence du système. Bien que cela passe par le choix de la performance des différents composants, l'évolution technologique amènera une réduction dans les temps de traitement. Dans l'attente, diverses possibilités existent pour réduire ce temps, par exemple en utilisant des systèmes fonctionnant en parallèle, travaux déjà initiés dans les années 95 par [OCMB95] et [Wlo95]. Encore à l'heure actuelle, les recherches sont effectuées pour profiter au maximum des systèmes graphiques dotés de hautes performances calculatoires [AHAS08, BB09, SFPG06], tâches jadis dévolues aux microprocesseurs¹⁴.
- Réduire la latence apparente, technique appelée « Image Deflection » qui consiste à introduire la mesure d'orientation la plus récente dans le processus de rendu [SG92, SG97]. Le rendu ayant auparavant été calculé sur une image plus grande que nécessaire, le système choisit la portion d'image correspondant à la mesure d'orientation venant d'être fournie. Malheureusement, cette technique ne fonctionne pas pour des mouvements de translation.
- Faire correspondre les flux temporels, problèmes rencontrés surtout sur les HMD non optiques. Une des techniques consiste à ajouter un retard supplémentaire aux flux vidéo pour compenser le retard dû à la génération des images virtuelles. De ce fait, les images réelles et virtuelles sont maintenant synchronisées, mais décalées par rapport à la réalité. Qui plus est, le retard entre le flux vidéo et le rendu peut ne pas être constant, ce qui nécessite une gestion de celui-ci. [JLS97] propose notamment une méthodologie pour ordonner et gérer les différents flux.
- Faire une prédiction du futur point de vue en l'extrapolant par des valeurs courantes et anciennes de la position-orientation. Si la future localisation est connue, il devient alors possible de commencer la génération de l'image virtuelle. Pour ce faire, on utilise des algorithmes tels que le filtre de Kalman [Kal60].

Nous venons de voir que le suivi et le recalage ont une importance cruciale dans une application de réalité augmentée, avec des contraintes fortes en termes de précision, de robustesse et de temps réel. Mais encore aujourd'hui, de nombreux problèmes persistent et constituent le principal verrou de la réalité augmentée, faisant ainsi l'objet de recherches actives dans ce domaine.

14. Ces techniques sont connues sous le nom de GPGPU, c'est-à-dire utiliser la carte graphique (GPU) pour en faire un usage général (General Purpose ou GP). Cela a été rendu possible grâce à l'API CUDA de la société NVIDIA pour ses propres cartes graphiques.

3 CONCLUSION

Au fil de la première partie de ce chapitre, nous avons pu nous rendre compte que les concepts de la réalité augmentée ne sont pas nouveaux. C'est finalement l'évolution technologique qui a permis un regain d'intérêt dans ce domaine. Nous avons pu discuter de la richesse conceptuelle de ces idées, et il ressort de ce fait une grande diversité d'interprétations sur la définition même de la réalité augmentée. En ce sens, pour avoir une vision commune au fil du document avec le lecteur, nous avons proposé notre propre définition de la réalité augmentée :

Réalité augmentée : « Combinaison de l'espace physique avec l'espace numérique dans un contexte sémantiquement lié ».

Comme nous l'avons déjà précisé, cette définition, bien que globale, n'exclut pas les précédentes et profite ainsi des diverses taxonomies établies au cours des décennies.

Nous avons tenu également à faire remarquer que l'expression « réalité augmentée » en elle-même souffre toujours à l'heure actuelle d'un problème sémantique, bien qu'acceptée par les communautés informatiques et maintenant par le grand public. Les pistes que nous avons explorées parallèlement à ce travail de thèse nous mènent à dire que ce n'est pas la réalité¹⁵ qui est augmentée, mais plutôt la perception que nous avons de celle-ci. Ainsi, les différentes technologies de la réalité augmentée nous permettraient d'étendre la représentation de notre propre réalité perçue.

Les moyens technologiques d'y parvenir ont justement été exposés dans la deuxième partie. Nous avons voulu présenter les fondements de ce qui caractérise l'implémentation technologique de la réalité augmentée. Nous nous sommes principalement intéressés à la modalité visuelle de ce vaste domaine, la vue étant généralement privilégiée par l'humain aux autres sens. Les augmentations visuelles et leurs types sont largement définis en fonction de la technologie d'affichage choisie. Il n'existe pas à l'heure actuelle de lunettes de réalité augmentée ultimes. Comme nous le verrons, c'est le plus souvent l'exigence de la tâche qui détermine les caractéristiques et le mode d'affichage des augmentations. Il en va de même pour l'architecture du système, et des contraintes de suivi et de recalage inhérentes à l'application.

15. Au sens du philosophe Kant, la réalité absolue (ou noumène) ne nous est pas accessible. La seule réalité à laquelle nous avons accès (le phénomène) l'est à travers notre perception.

Deuxième partie

L'HOMME, LA MAINTENANCE ET LA RÉALITÉ AUGMENTÉE

Au coeur de toutes activités industrielles se trouve l'Homme. Il est l'élément essentiel de par sa capacité à s'adapter à son contexte. Mais il est également confronté à un environnement de travail toujours plus riche dont la complexité ne cesse de croître. Ces nouvelles contraintes liées à son activité favorisent l'apparition de comportements pouvant conduire de façon involontaire à divers incidents ou accidents. Ces « erreurs humaines » sont le fruit du fonctionnement cognitif de tout un chacun, et nous allons voir les mécanismes de notre pensée qui nous amènent à accorder confiance à notre raison, souvent à tort !

Le fait de se tromper trahit bien souvent un savoir incomplet, et l'un des meilleurs moyens de pallier ce problème est la mise en place de systèmes d'aide. Encore faut-il qu'ils soient utilisés à bon escient, pour peu qu'ils demeurent corrects dans leur conception. Alors, quel pourrait être l'intérêt d'introduire de nouvelles technologies comme la réalité augmentée dans le processus d'assistance ? Est ce que les propositions actuelles permettent de soutenir efficacement un opérateur dans sa tâche ?

4 LE FACTEUR HUMAIN

« Errare humanum est, perseverare diabolicum »¹⁶.

La première partie de cet adage est bien connue et généralement sert d'excuse à l'inévitable faillibilité humaine. La seconde partie de cet adage, elle, nous met en garde contre la récurrence qui peut être dommageable. Nous sommes censés apprendre de nos erreurs !

Cependant, il est illusoire de penser supprimer les erreurs, résultat découlant des mécanismes de traitement de l'information acquis par l'humain. Comprendre les processus amenant l'erreur est donc essentiel pour ne pas rendre vaine toute tentative de les éviter.

4.1 Les mécanismes de l'erreur

4.1.1 Le processus cognitif

L'erreur¹⁷ prend naissance dans la façon dont le système cognitif humain traite l'information lors de la résolution de problèmes. Rasmussen dans [Ras86] propose à ce titre un modèle, dit de l'échelle double, illustrant la démarche générale de résolution de problèmes par un opérateur humain (Fig.22).

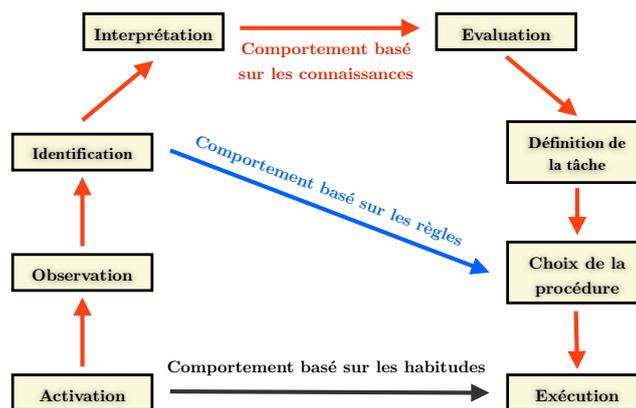


FIGURE 22: Modèle de l'échelle double.

La taxonomie SRK issue de ce modèle définit trois niveaux de processus cognitifs :

- Skill Based (niveau de compétence, habileté) : ce niveau correspond aux actions effectuées de manière quasi automatique après un apprentissage pouvant être relativement long. Une fois intégré, l'appel à ses fonctions nécessite un minimum de ressources (mentales) pour être mis en oeuvre.
- Rule Based (Niveau des règles) : ce niveau correspond aux règles (procédures, algorithmes...) apprises par l'homme pour définir son action. Il nécessite un effort cognitif plus important dans la mesure où à partir de ses perceptions, l'Homme doit au préalable reconnaître une situation, sélectionner une règle appropriée puis entreprendre son action.

16. « L'erreur est humaine, persévérer (dans l'erreur) est diabolique ». Locution latine attribuée au philosophe Lucius Annaeus Seneca (-4 av. J.-C., 65 ap. J.-C.) qui fut le tuteur de l'empereur romain Nero.

17. L'étude de l'erreur appartient à un domaine à la croisée de nombreuses disciplines nous obligeant à un développement succinct de ces principales idées faisant référence.

- Knowledge Based (Niveau des connaissances déclaratives) : ce dernier niveau fait appel aux cas où aucune règle ou procédure connue ne permet de résoudre le problème donné. Ce niveau intervient lorsqu'une situation est inconnue. L'Homme doit alors trouver la solution par lui-même à partir de ses connaissances. Ce processus nécessite beaucoup plus de ressources cognitives que les précédents niveaux et peut prendre beaucoup de temps avant de permettre l'élaboration d'un plan d'action.

4.1.2 Topologie de l'erreur

C'est en se basant sur les études préalables de Rasmussen que James Reason, l'un des théoriciens de l'erreur, développa une topologie d'erreur généralement commise lors de l'utilisation d'un système par un opérateur [Rea93]. Tout d'abord, Reason introduit une distinction fondamentale entre deux catégories d'erreur, les erreurs actives et les erreurs latentes. Les erreurs actives ont un effet immédiat lorsqu'elles sont commises par leurs exécutants. Quant aux erreurs latentes, elles se développent « à partir des activités de ceux qui sont éloignés de l'interface de contrôle direct, à la fois dans le temps et dans l'espace : les concepteurs, les décideurs, les ouvriers qui construisent le système, les directeurs et personnels de maintenance ». Les erreurs latentes « peuvent rester longtemps en sommeil dans le système » avant de provoquer des conséquences néfastes. C'est souvent la combinaison de ces deux catégories d'erreurs qui est à l'origine des principaux accidents dans l'industrie. Cet enchaînement d'erreurs a d'ailleurs abouti au célèbre modèle de l'Emmental (Fig.23).

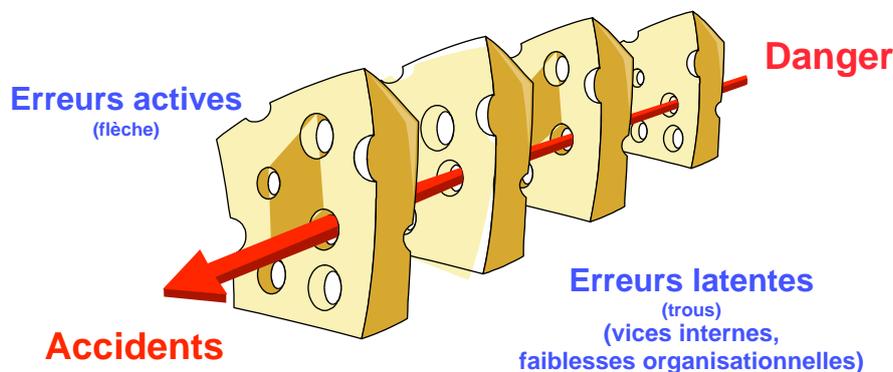


FIGURE 23: Modèle de l'enchaînement d'erreurs de Reason, dit « modèle de l'Emmental ».

Comme l'écrit Reason en guise de conclusion dans son ouvrage, « plutôt que de considérer les opérateurs comme les principaux instigateurs de l'accident, il faut comprendre qu'ils sont les héritiers des défauts du système, créé par une mauvaise conception, une mauvaise mise en place, un entretien défectueux et de mauvaises décisions de la direction ». Nous pouvons donc raisonnablement penser que l'on peut « casser » cet enchaînement en minimisant les erreurs actives, et ce, en agissant sur au moins l'une des strates du modèle de Reason.

Pour revenir à la topologie d'erreur de Reason (fondée sur les trois niveaux de fonctionnement cognitif de Rasmussen), elle est définie en trois types. On entend par type d'erreur les erreurs renvoyant à leur origine présumée.

Ces trois types sont :

- Les ratés/ les lapsus : basés sur les automatismes du sujet. L'action est mise en échec, soit par l'inattention, en omettant les vérifications d'usage, soit par l'attention excessive, en effectuant les vérifications à un moment inapproprié.
- Les fautes basées sur les règles : erreurs découlant d'une mauvaise application de mauvaises règles et erreurs dues à l'application de règles fausses¹⁸.
- Les fautes basées sur les connaissances déclaratives : apparaissant lors de situations nouvelles, elles sont le résultat d'une limitation des connaissances tournées vers un objectif dont la conscience du contexte est toujours incomplète.

4.1.3 Les formes d'erreur

Indépendamment du type de l'erreur, les formes d'erreur (c'est-à-dire les formes récurrentes de défaillance apparaissant quel que soit le type d'activité) sont conditionnées par les mécanismes de recherche cognitive regroupés en deux grands facteurs. Le premier est l'appariement par similarité, mécanismes prédominant lorsque les indices du contexte sont adéquats pour déterminer une connaissance unique, et que le niveau de connaissance de l'humain est important. On pourrait résumer cela par l'adage : « à situation similaire, traitement similaire ! ». C'est a contrario lorsque les indices du contexte sont insuffisants, voire ambigus, et que le domaine de connaissance est limité qu'intervient le mécanisme d'appariement par fréquence. Cette illusion des séries, menant à penser que la situation va se répéter, induit d'agir comme si cela était le cas, ce qui conduit relativement souvent à l'erreur.

Cette explication des mécanismes de l'erreur est basée sur l'hypothèse que l'Homme possède deux modes de contrôle de l'activité et du traitement de l'erreur :

- Le mode de contrôle attentionnel : ce mode est caractérisé par un fonctionnement contrôlé, c'est-à-dire en tout état de conscience. Il est lié à la mémoire de travail qui a entre autres pour rôle d'établir des objectifs en sélectionnant les moyens nécessaires à leur accomplissement... Ce mode de contrôle demande lui-même un effort important.
- Le mode de contrôle schématique : caractérisé par des traitements inconscients automatiques. Il s'appuie sur la reconnaissance de régularité, de situations familières, et est donc peu efficace face à la nouveauté. Pour autant, il a l'avantage de ne demander quasiment aucun effort cognitif.

Selon Reason, c'est donc les modes de contrôle qui entraînent les formes d'erreur. L'erreur humaine découlerait alors des mêmes sources mentales que celles de la perspicacité et en établirait « le revers de la médaille ». Les mécanismes précédemment décrits sont à l'origine de nombreuses erreurs commises dans le monde du travail, et nous allons plus particulièrement voir celles que l'on rencontre dans le domaine de la maintenance industrielle.

18. Les bonnes règles sont celles qui n'ont pas été mises en défaut dans une situation particulière. Elles ont donc de grandes chances d'aboutir à un succès dans le futur. Si elles sont mises en défaut, alors elles engendreront des règles spécifiques s'adaptant à l'éventail de situations particulières.

4.2 Les erreurs dans l'activité de maintenance

4.2.1 La défaillance, prémices de l'erreur

Dans l'industrie, les erreurs sont amenées à être commises lors d'une tâche de maintenance sur une pièce (ou système) ayant préalablement subi une défaillance. Avant de présenter dans les paragraphes suivants les erreurs résultantes, il convient de définir les termes « défaillance » et « maintenance » :

Défaillance¹⁹ : « Cessation de l'aptitude d'un bien à accomplir une fonction requise ». La défaillance peut être vue comme l'évènement faisant basculer un bien dans un état de panne. Les raisons en sont nombreuses, pouvant être dues à la fabrication, à la conception ou encore à l'usure.

Maintenance²⁰ : « Ensemble de toutes actions techniques, administratives et de management durant le cycle de vie d'un bien, destiné à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise ». La maintenance est catégorisée en deux principales parties, la maintenance préventive et corrective (Fig.24). La maintenance préventive a pour objectif la réduction de la probabilité de défaillance ou de dommage d'un bien. La maintenance corrective elle, vise à rétablir un bien dans l'état d'accomplissement de la fonction requise.

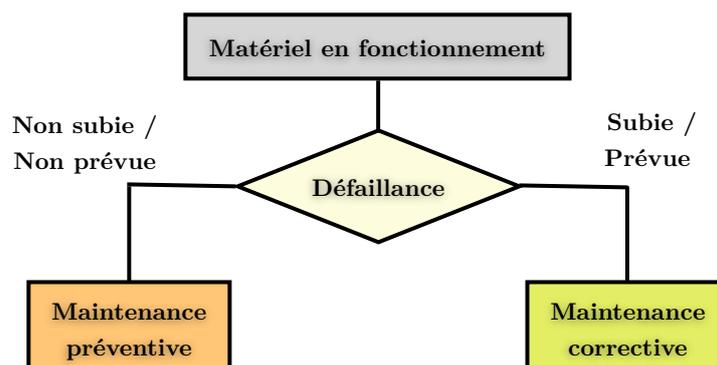


FIGURE 24: « Décomposition cartographique » de la maintenance par rapport à la défaillance [Rey02].

L'objectif ici n'étant pas d'établir de façon exhaustive les différents types de maintenance, le lecteur pourra se reporter à [dB09] traitant de façon complète l'activité de maintenance.

4.2.2 Les principales erreurs

Dans le domaine de la maintenance, surtout aéronautique, les erreurs prennent de plus en plus de place, entraînant ainsi incidents et accidents. Plusieurs études en font l'écho comme celles de la United Kingdom Civil Aviation Authority [UKC92], Boeing [MG93] ou encore dans l'étude rapportée par Reason [Rea97]. Plus récemment, la UKCAA a rendu les résultats d'une étude effectuée sur une période de 10 ans (1996 à 2006) dans les activités de maintenance [UKC09]. Dans cette étude, les erreurs ayant amené des incidents ont été catégorisées en trois parties (Fig.25) :

19. Définition AFNOR du document NF EN 13 306 sur la « terminologie en maintenance ».

20. Définition AFNOR du document NF EN 13 306 X 60-319 sur la « terminologie en maintenance ».

- « Contrôle de maintenance », c'est-à-dire ayant subi un contrôle de maintenance inefficace (dysfonctionnel ou inadéquat), soit environ 1/4 des erreurs (26,22 %).
- « Maintenance incomplète » ou inachevée, soit environ 1/4 des erreurs (20,68 %).
- « Action incorrecte de maintenance », c'est-à-dire que l'activité de maintenance est arrivée à son terme tout en étant porteuse d'omissions ou de négligences. Cela représente la moitié des erreurs (53,11 %).

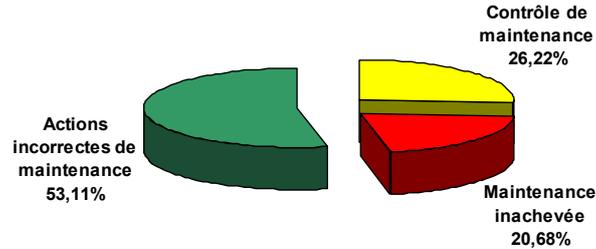


FIGURE 25: Catégorisation des erreurs ayant induit des accidents.

Il est important de remarquer que plus de la moitié des erreurs proviennent de la même catégorie, à savoir celle ayant impliqué des actions de maintenance incorrectes. Cela nous renvoie aux modes de contrôle définis par Reason, tant attentionnel que schématique.

Au-delà des modèles cognitifs expliquant le mécanisme intrinsèque de résolution de problème, les erreurs de maintenance trouvent également racine dans les nombreux facteurs liés à l'immersion de l'Homme au sein de ce domaine qu'est la maintenance industrielle. Les principaux facteurs affectant la performance de l'opérateur ont été identifiés [Pat02] en facteurs organisationnels et individuels (Tab.5).

Facteurs Organisationnels	Facteurs Individuels
Hardware, les équipements, les outils, le manque de ressources, pas assez de personnels	La santé physique
La conception, la configuration, les pièces	La fatigue
Le management de la maintenance, la direction, la supervision, le politique d'entreprise	Les contraintes de temps
Les processus de travail, les procédures, les informations	La pression du management
Error-enforcing conditions, (les règles) les manières de faire, influence de l'entourage	La suffisance / la complaisance
Le ménage	La taille et la force de l'individu
Les objectifs incompatibles	Un évènement ou stress personnel
Les processus de communication	La distraction sur le lieu de travail
Les structures organisationnelles, le changement d'entreprise, l'action syndicale	Le manque de conscience
La formation, les connaissances techniques, compétences	Le manque de connaissance
Les défenses	Manque de compétences de communication
L'environnement, installation	Le manque d'affirmation de soi
Le manque de travail en équipe	

TABLE 5: Principaux facteurs affectant les performances humaines.

La diversité des facteurs d'erreurs, qu'elle soit liée au contexte social, à l'environnement physique ou à la psychologie de l'opérateur, nous renseigne quant à la difficulté de pouvoir « éviter » les erreurs actives, et encore plus les erreurs latentes. Mais la connaissance de ces informations est déjà un moyen d'agir de façon locale, à défaut d'une action globale!

4.3 Taxonomie des erreurs

Reconnu et servant de référence dans la littérature, Dupont a proposé dans [Dup97] une classification sur les principales causes aboutissant à une erreur de maintenance. Au nombre de 12 (dirty dozen²¹), ces erreurs sont directement à relier à la topologie de l'erreur proposée par Reason :

- Le manque de communication : cela relate simplement l'échec dans le processus de partage des informations, orales ou écrites. Dupont préconise une communication verbale directe entre les interlocuteurs, avec quelques précautions toutefois, car lors d'un échange oral, un pourcentage faible d'information est retenu, généralement le début et la fin de celui-ci. En conséquence, il est important de terminer un discours par l'énonciation des points-clés. En termes de communication, il ne faut donc jamais rien présumer ou supposer.
- La complaisance : définie comme l'autosatisfaction. Celle-ci s'accompagne généralement d'une perte de conscience du danger. C'est la répétition dans la tâche qui amène l'opérateur à voir ce qu'il s'attend à voir, pouvant ainsi manquer d'importants signaux. Ce biais de fréquence est résumé par l'adage : « j'ai regardé 1000 fois et je n'ai jamais trouvé quelque chose qui n'allait pas! ».
- Le manque de connaissances : le changement technologique grandissant contribue largement à l'erreur. Les formations toujours plus courtes et la capacité de l'Homme à ne retenir qu'un faible pourcentage de ce qu'il apprend (au profit des connaissances souvent utilisées) sont autant de facteurs qui entraînent une erreur de jugement. Dupont préconise d'employer des opérateurs qualifiés ou faire en sorte qu'ils le deviennent. Si le manque de connaissances se fait sentir, il ne faut surtout rien supposer et demander de l'aide à une personne en ayant les capacités.
- Les distractions : liées à tout ce qui peut distraire l'esprit d'une tâche mentale ou/et manuelle, même un instant. D'après les psychologues, la distraction serait la cause numéro un entraînant l'oubli. Nous pensons plus vite que ce que nos mains exécutent, laissant croire que la tâche est plus avancée que ce qu'elle est réellement. La moindre distraction ferait donc reprendre la tâche à une mauvaise étape. Cela serait la cause d'environ 15 % des erreurs de maintenance. Il est donc préconisé de finir sa tâche avant tout, ou de noter l'étape à laquelle elle doit être reprise.
- Le manque de travail en équipe : plus une équipe s'agrandit, plus ce facteur prédomine. À relier directement au manque de communication, les objectifs doivent être clairs et tous les membres doivent être impliqués dans le processus de décision.
- La fatigue : l'humain n'a généralement pas conscience de son état de fatigue jusqu'à ce qu'il devienne extrême. Le jugement est alors altéré, contribuant grandement à l'erreur. Il est préconisé de sensibiliser les opérateurs aux symptômes caractérisant cet état, afin de pouvoir le déceler soi-même ou chez les différents membres de l'équipe. Si tel est le cas, il est alors nécessaire de demander à faire contrôler son travail par un tiers.

21. Dupont ayant certainement voulu faire un clin d'oeil au film américain du même nom réalisé par Robert Aldrich en 1967.

- Le manque de ressources : ce manque peut interférer sur la capacité d’accomplir une tâche. Il peut être de l’ordre du défaut d’approvisionnement, du manque d’assistance technique ou encore de biens de qualité médiocre. Dans tous les cas, il est préconisé de commander et de stocker les biens attendus avant que le besoin ne s’en fasse sentir, ou de prévoir des solutions de mise en commun ou de prêts.
- La pression : la pression est définie comme l’attitude mentale consistant à fixer son mental sur un but à atteindre. Cette attitude, contrairement aux apparences, n’est pas le résultat de circonstances extérieures, mais la capacité qu’a l’être humain de se projeter vers un futur inexistant et de le vivre comme s’il était réel. Il est donc important que l’opérateur puisse désamorcer ce processus en communiquant ses inquiétudes, doutes ou incertitudes, et surtout de savoir dire non.
- Le manque d’assurance : juste milieu entre la passivité et l’agressivité²², l’opérateur doit être autoritaire et intransigeant vis-à-vis du respect des normes, des réglementations ou des principes face à des situations où il serait contraint à la précipitation.
- Le stress : « la réponse de stress est le résultat d’un déséquilibre entre la perception individuelle des exigences de la situation et celle des ressources de l’individu pour faire face à ces demandes » [LF84]. Le stress affecte l’activité humaine et s’avère important en situation dégradée. Les erreurs commises ont alors souvent pour origine différents biais comme celui de confirmation (justifier une analyse immédiate en sélectionnant les informations qui la confirment).
- Le manque de conscience : défini comme « un échec à reconnaître toutes les conséquences d’une action, ou un manque de prévoyance ». Cela traduit en général un défaut d’observation qui amène un conflit dans une situation donnée. D’après l’étude de Dupont, ce problème se produirait le plus souvent chez les opérateurs expérimentés ne se posant plus les questions ayant trait aux conséquences (« que se passerait-il si... ? »).
- Les normes : Les normes sont les règles non écrites ou tolérées qui existent au sein d’un groupe. Le besoin d’appartenance [Mas54] est un besoin primaire de l’individu qui le pousse de façon générale à respecter ces comportements dictés par la majorité. Bien que ces normes puissent être bonnes ou mauvaises, Dupont préconise de toujours travailler selon les instructions officielles.

5 L’AIDE AU TRAVAIL

Nous avons vu précédemment que le mode de réflexion humaine est enclin à commettre des erreurs. Certaines études ont été consacrées à établir les facteurs et les causes les plus probables compromettant une tâche de maintenance. Pour pallier cela, l’humain a recours à tout procédé lui permettant de s’acquitter de son objectif, autrement dit à une aide. L’aide vise donc à améliorer l’efficacité d’un opérateur pour effectuer son travail de manière qualitative.

22. En psychologie, l’agressivité est « la tendance ou un ensemble de tendances qui s’actualisent dans des conditions réelles ou fantasmatiques, celle-ci visant à nuire à autrui, à le détruire, le contraindre ou à l’humilier ». Extrait du dictionnaire de la langue française du Centre National de Ressources Textuelles et Lexicales (www.cnrtl.fr).

5.1 Pourquoi recourir à une aide ?

5.1.1 Objectifs et définition

Dans l'industrie moderne, l'opérateur en maintenance est confronté à une complexification croissante des systèmes technologiques. Ces systèmes sont très souvent à la croisée de divers domaines tels que l'électronique, la mécanique et l'informatique. Il est alors improbable et difficile d'utiliser ou d'intervenir sur de tels systèmes sans aide. À ce titre, pour illustrer ce propos, nous citerons le concepteur du langage C++ Bjarne Stroustrup : « J'ai toujours rêvé d'un ordinateur qui soit aussi facile à utiliser qu'un téléphone. Mon rêve s'est réalisé, je ne sais plus comment utiliser mon téléphone ! ».

Au-delà de cette complexité, s'ajoute un renouvellement des gammes de produits proposés par les industriels, s'accompagnant le plus souvent d'une évolution technique/technologique. Pour un opérateur, la formation est alors le passage obligé pour acquérir de nouvelles connaissances de façon régulière. Cependant, les formations ont un coût non négligeable en termes de conception et de mise en oeuvre.

Nous verrons par la suite que même si elles permettent l'acquisition de nouvelles connaissances, cela se révèle insuffisant, l'ensemble des connaissances n'étant pas intégralement retenu. Ici, l'aide doit offrir la possibilité de fournir des moyens d'atteindre un objectif sans pour autant nécessiter une pratique professionnelle très longue. Elle ne se substitue pas pour autant à l'expertise.

Selon [Gro98], aider un opérateur « c'est estimer qu'il risque de rencontrer des difficultés et que l'aide améliorera la réalisation de sa tâche, donc sa performance au travail. L'aide apparaît lorsque trois éléments sont réunis : une tâche, un opérateur chargé de réaliser la tâche et une aide (humaine ou technique) chargée d'assister l'opérateur dans sa tâche ». L'objectif des aides serait alors de décrire l'activité attendue en codifiant l'exécution du travail.

5.1.2 Les écarts

Il arrive malgré tout que l'effet escompté ne soit pas celui obtenu dans la réalité de la tâche. Ainsi, au lieu de promouvoir efficacité et performance pour l'opérateur, l'aide peut devenir une contrainte supplémentaire, augmentant par la même occasion la charge de travail initiale.

[Bra07] estime qu'une aide adéquate dépend principalement de trois facteurs : de la pertinence des fonctionnalités (adaptation de l'aide aux objectifs de la tâche), du niveau d'utilisabilité (compatibilité de l'aide avec les caractéristiques physiques, cognitives et sociales de l'opérateur vis-à-vis de son travail) et des modes de régulation psychosociale (compromis entre les interactions humaines — aides techniques — organisation).

C'est la non-prise en compte de ces facteurs qui peut faire devenir l'aide « inutile, détournée, sabotée, inefficace et même complexifier le travail alors que l'objectif est de le simplifier », et plus particulièrement les aides peuvent [Bra07] :

- Fournir des procédures détaillées, mais qui ne permettent pas d'aider la planification de l'activité en cas d'erreur ou d'incident.

- Favoriser la constitution de représentations d'ensemble des objets traités ; mais ces schémas risquent également de masquer des informations qu'une activité cognitive doit reconstruire.
- Améliorer les heuristiques des opérateurs novices et ainsi présenter des qualités pédagogiques, mais s'avérer inefficaces pour des individus expérimentés.
- Réduire les temps d'apprentissage en faisant accomplir une partie du travail par le dispositif d'aide, tout en entraînant ainsi une déqualification de l'opérateur : il devient incapable de faire seul son travail.
- Diminuer la capacité des opérateurs à se rappeler rapidement les informations essentielles.
- Détériorer les compétences liées au contrôle des actions des opérateurs, tout en permettant le développement de nouvelles formes de compétences collectives.
- Restreindre les possibilités d'anticipation et de gestion des situations qui ne sont pas prévues dans les aides.
- Être d'une très mauvaise ergonomie, qui complexifie l'accès et l'usage de l'aide.
- Déplacer les responsabilités et créer un climat de surconfiance : l'opérateur risque de ne plus contrôler la complétude et la pertinence des informations et d'accepter en l'état les propositions ou les décisions de l'aide. Du coup, l'attribution des responsabilités devient un problème, surtout en cas d'incidents.
- Réduire les gratifications, car la réussite d'une opération est vue comme étant due à l'aide et non plus aux individus.
- Rendre difficile, voire impossible, de se constituer une vision globale du fonctionnement des installations, surtout si l'aide ne porte que sur une portion réduite du travail.
- Déqualifier l'opérateur, car il n'est plus utilisé que sur une petite partie du domaine de l'expertise, ce qui ne favorise pas un apprentissage global du métier.
- Ne pas améliorer la qualité et la durée de la prise de décision des utilisateurs ayant des attitudes de résistance à l'égard de l'aide ; ceci étant ces attitudes peuvent évoluer dans le temps.
- S'accompagner de modifications de l'organisation du travail en remettant notamment en cause la division du travail. Ces systèmes, comme beaucoup de nouvelles technologies, ne se contentent pas de se diffuser dans les entreprises, mais ils contribuent à leurs transformations.

Impossible alors de réduire l'aide à un simple glissement de connaissance de l'individu. Donc, fournir de l'aide, oui, mais pas n'importe comment !

5.2 Les types d'aides

Dans [Bra01], l'auteur propose de placer les dispositifs d'aide en fonction du but qui leur est assigné. Ces types d'aides peuvent se combiner entre eux et peuvent prendre de nombreuses formes (manuels, logiciels, assistance humaine...).

Les aides sont au nombre de trois : l'aide informative, collaborative et supplétive.

- L'aide informative : « Elle vise à donner des informations, souvent de nature procédurale, sur des problèmes posés à un utilisateur pour réaliser une tâche donnée. L'aide informative organise le champ informationnel de l'opérateur. Les supports écrits, muraux et informatiques structurent l'activité de l'opérateur dans la mesure où ils induisent une division des tâches et une organisation des procédures ».
- L'aide collaborative : « Elle permet la collaboration entre un agent d'assistance, humain ou technique, et un opérateur disposant de ressources limitées. L'opérateur entre en interaction avec l'assistance, qui ne se limite pas seulement à une présentation des informations, mais lui pose des questions, explique les décisions, aide au diagnostic, etc. L'assistant effectue un tri des informations, traite les informations estimées pertinentes et conseille l'opérateur. L'assisté va construire son raisonnement au fur et à mesure des interactions avec l'assistance et, pour finir, opter pour une décision ».
- L'aide supplétive : « Elle prend la place de l'homme dans un certain nombre de tâches qu'il ne peut réaliser. Cette assistance est conçue pour pallier les déficiences humaines (handicaps moteurs ou sensoriels) ou l'inaccessibilité d'une situation (télésurveillance, télémaintenance, robotique...) ».

Le découpage proposé par Brangier a l'avantage de caractériser l'aide suivant un niveau graduel. Elle s'échelonne d'un degré basique où l'information est présentée de manières plus ou moins adaptées et pertinentes, en passant par une aide interactive ciblée jusqu'à un dispositif d'assistance remédiant aux incapacités humaines.

5.3 Procédures, documentations... une énigme ?

Dans l'histoire du travail, l'écriture a progressivement remplacé la communication orale, modifiant ainsi la distribution des connaissances de l'individu, déqualifié dans certains domaines, expert dans d'autres. Mais dans notre société moderne, l'accumulation de savoirs rend difficile l'acquisition de multiples connaissances, l'humain étant alors obligé d'avoir recours à un système d'aide. Les procédures et autres documentations sont autant de référentiels pour un opérateur sur la description codifiée de sa tâche. Il y trouve ce qui doit être fait et comment cela doit être fait. Pour autant, les documents censés normaliser l'action ne sont pas exempts de défauts, ce qui amène souvent leur concepteur à se poser la question : « pourquoi les utilisateurs ne lisent-ils pas cette #@%* documentation ? ».

L'étude de Mitchell dès 1993 montra les attentes des utilisateurs envers une documentation [Mit93], que ce soit sous forme physique ou immatérielle. Il doit être facile d'y « naviguer », les explications devant être d'un niveau de détails techniques approprié, et permettant de trouver aussi bien que de résoudre des problèmes à travers des exemples et des scénarii.

5.3.1 Considérations générales

Alors, exaucés les vœux des utilisateurs ? Pas si sûr ! Une décennie plus tard, Brangier et Barcenilla se sont attachés à évaluer l'utilisabilité des aides textuelles fournies aux opérateurs faiblement qualifiés dans une dizaine d'entreprises couvrant divers domaines. Voici le résultat de l'évaluation de ces documents [BB03] :

- 56,9 % des documents n'ont pas une bonne visibilité.
- 43,4 % présentent des abréviations et sigles inexplicables.
- 34,9 % ont des formes syntaxiques complexes.
- 76,7 % des documents ont des phrases courtes.
- 58,9 % ont des phrases grammaticalement mal construites.
- 14,1 % des documents présentent les prérequis des actions à réaliser.
- 17,6 % seulement expliquent les actions à réaliser.
- 19 % seulement intègrent aux modes opératoires les procédures de récupération des erreurs.
- 60,7 % ne présentent pas correctement l'objectif des procédures.

Ces chiffres expriment clairement qu'une aide mal pensée est obligatoirement soumise à interprétation de la part des opérateurs, l'objectif descriptif et codifié ayant ici échoué.

David Novick s'est également intéressé à savoir pourquoi les utilisateurs étaient réticents à recourir aux documents, et ce qu'ils en attendaient [NW06a][NW06b]. La première remarque intéressante provient du fait que l'aide papier en tant que telle est peu appréciée. Les manuels sont considérés trop lourds, encombrants et donc difficilement utilisables. Les utilisateurs se tournent plus volontiers vers les systèmes informatiques pour chercher ce dont ils ont besoin... avant de l'imprimer sur papier ! Nous pensons que ce résultat est cohérent en termes d'efficacité, l'opérateur exploitant le meilleur des deux mondes ; l'informatique pour sa capacité de traitement de l'information, le physique (papier) pour sa manipulation aisée dans l'environnement de la tâche.

Concernant les attentes, les opérateurs ne connaissent pas toujours la bonne terminologie et pensent qu'il serait plus efficace de pouvoir chercher une information par les synonymes qui lui sont associés. De ce fait, ils ne sont pas toujours certains que l'information recherchée est effectivement dans la documentation.

Autre point, la documentation est utilisée uniquement pour résoudre un problème qui se présente à eux durant la tâche. En les interrogeant, il ressort de cela que les informations devraient être présentées plutôt en termes de diagnostics et de résolution des problèmes. Si cela s'avère insuffisant, les documents devraient inclure des pointeurs sur des informations additionnelles ou sur les personnes à contacter pour porter assistance.

Pour ce qui est du niveau des explications, les opérateurs trouvent qu'elles ne répondent pas souvent aux besoins de leur public cible, la majorité pensant que le niveau est soit trop complexe, soit trop général pour répondre avec succès à leurs questions. Quant aux outils multimédias, il est déploré le manque d'illustrations visuelles et d'animations. Il est clair qu'ici, l'informatique serait un moyen efficace pour permettre d'adapter le niveau de détail et d'explication visuelle en fonction des capacités de l'opérateur.

Enfin, Novick fait ressortir le niveau de frustration que provoque l'utilisation d'une aide, tant informatique que physique, la difficulté venant de lire et d'exécuter les instructions en même temps. Il est à noter que ceci est à relier au fait que les opérateurs préfèrent imprimer des pages qui par la suite pourraient être facilement positionnées, attachées, scotchées sur le lieu de la tâche.

5.3.2 Dans l'industrie

Dans l'industrie, où l'activité de maintenance est fortement corrélée au niveau de sécurité, la mise en oeuvre des systèmes d'aide est davantage codifiée. En effet, les manuels de maintenance²³ doivent respecter une organisation, une forme et des informations rédigées par des détenteurs de certificats (les constructeurs en majorité) en suivant des règles imposées (ou recommandées) par des organismes de réglementation. Néanmoins, les opérateurs de maintenance ne sont pas forcément plus respectueux des procédures, ce qui bien entendu ici est l'un des facteurs contribuant à l'apparition d'incidents. [Emb00] en a d'ailleurs étudié les raisons, dans des industries dites à haut risque comme l'Aérospatiale, la chimie... (Tab.6).

Les procédures ne sont pas utilisées parce que...		%
Exactitude / Précision	Elles sont imprécises	21
	Elles sont dépassées	45
Possibilité de réalisation	Elles sont impraticables	40
	Elles rendent plus difficile le travail à faire	42
	Elles sont trop restrictives	48
	Elles sont trop longues	44
	Si elles sont suivies à la lettre, le travail ne peut pas être effectué dans les temps	62
Optimisation	Les individus trouvent une meilleure manière de faire le travail	42
	Elles ne décrivent pas la meilleure manière de faire le travail	48
Présentation	Il est difficile de savoir quelle est la bonne procédure	32
	Elles sont trop complexes et difficiles à utiliser	42
	Il est difficile de trouver l'information dont on a besoin dans une procédure	48
Accessibilité	Il est difficile de localiser la bonne procédure	50
	Les individus ne sont pas au courant qu'il existe une procédure pour le travail qu'ils sont en train de réaliser	57
Politique	Les individus ne comprennent pas pourquoi elles sont nécessaires	40
	Aucune politique claire sur le moment où elles doivent être utilisées	37
Usage	Les individus expérimentés n'en ont pas besoin	19
	Les individus acceptent mal qu'on leur dicte comment faire leur travail	34
	Les individus préfèrent compter sur leurs propres compétences et expériences	72
	Les individus supposent qu'ils savent ce qui se trouve dans la procédure	70

TABLE 6: Motif du non-respect des procédures par les opérateurs de maintenance dans l'industrie [Emb00].

Ces facteurs influençant l'utilisation des procédures sont, à bien y regarder, assez similaires à ceux précédemment exposés. On y retrouve par exemple les difficultés dues à la présentation, c'est-à-dire des informations dont le niveau n'est pas adapté à l'utilisateur, ou bien la difficulté de trouver la bonne procédure. À ce niveau, il faut savoir qu'il peut effectivement ne pas être évident de localiser une procédure puis d'y naviguer. En effet, la documentation est généralement organisée en tâches faisant référence à des sous-tâches, elles-mêmes faisant référence à des tâches... Cela implique de nombreux « aller-retour » dans les manuels, ce qui peut faire perdre le fil de la tâche. Nous sommes donc ici en présence d'une des « Dirty dozen » de Dupont : la distraction. Pourtant, il est à noter l'avantage d'une telle organisation de la documentation qui facilite grandement la mise à jour des

23. Ils sont le plus souvent rédigés en anglais technique simplifié, ce qui lors des traductions (notamment françaises) peut expliquer que la longueur des phrases est relativement courte, problèmes pointés par l'étude de Brangier (op. Cit.)

procédures (unicité de la procédure), et permet de réduire efficacement la taille et le volume de celle-ci.

Enfin, il est intéressant de remarquer que 70 % des individus préfèrent compter sur leurs propres compétences, autrement dit faire appel au niveau S et R du modèle cognitif de Rasmussen, niveaux propices aux lapsus et fautes basés sur les règles. Ceci peut expliquer en grande partie l'origine de la majorité des erreurs commises. À cela, si l'on y ajoute les 70 % d'individus qui supposent connaître ce qui se trouve dans les manuels, nous pouvons comprendre le « désamour » des opérateurs pour la documentation.

Comme Brangier, [CGCS02] a lui aussi étudié la pertinence des manuels de maintenance, et ses conclusions n'en sont que peu différentes. D'après les résultats, les manuels ne seraient pas forcément appropriés aux activités de maintenance et souffriraient d'imprécisions, d'ambiguïtés et d'un certain manque d'information. Voici les principaux résultats (Tab.7 & Tab.8) :

	Jamais	Rarement	Occasion- nellement	Souvent	Très souvent
Poucentage de fréquence dans l'année où un opérateur a réalisé un travail d'une meilleure manière que celle proposée par le manuel	13%	25%	40%	16%	6%
Procédure ambiguë imprécise ou incomplète ayant entraîné une pièce endommagée	38%	44%	14%	-	-
Consultation d'autre(s) opérateur(s) face à une procédure ambiguë, imprécise ou incomplète	-	12%	23%	41%	18%
Propre manière de réaliser la tâche face à une procédure ambiguë ou incomplète	12%	23%	41%	18%	-
Fréquence des problèmes dans les manuels rapportés	3%	21%	30%	31%	15%

TABLE 7: Problèmes liés aux manuels de maintenance.

	D'accord	Pas d'accord	Neutre
Le manuel décrit la manière la plus facile pour conduire chaque procédure	18%	47%	35%
Les rédacteurs comprennent la façon dont les opérateurs réalisent les tâches	13%	54%	33%

TABLE 8: Ce que les opérateurs en maintenance pensent des manuels.

Ici, seulement 18 % des opérateurs pensent que les manuels décrivent la bonne manière de faire, ce qui a pour conséquence que ces mêmes opérateurs préfèrent d'abord consulter leurs homologues (41 % souvent et 18 % très souvent), ou accomplir la tâche de leur propre manière (41 % occasionnellement et 18 % souvent). [MIMS05] confirme également cet aspect de « tenter autre chose »

avant de recourir à la documentation officielle. L'origine de l'inadaptation des manuels serait que les rédacteurs ne comprennent pas la façon dont les opérateurs réalisent leur tâche, mais dans le même temps, seulement 46 % des utilisateurs rapportent les problèmes qui y apparaissent.

D'une manière générale, la documentation mise à disposition des opérateurs souffre de nombreux défauts : accessibilité, imprécision, décomposition des procédures non appropriées, mauvaise organisation... Tout cela est autant de sources d'incitation pour l'opérateur à ne pas suivre les procédures officielles, préférant ainsi utiliser des moyens parallèles, parfois valides et utiles, mais pouvant largement contribuer à l'échec.

5.4 L'assistance téléphonique

Quand l'utilisation des aides textuelles échoue, les opérateurs peuvent recourir à une expertise via le média le plus couramment utilisé : le téléphone. Bien qu'il existe à l'heure actuelle nombre de systèmes de communication à distance (courriel, forum, chat...), le téléphone permet aux interlocuteurs de véhiculer de l'information suffisamment riche, et ce, de manière synchrone.

Toutefois, l'assistance téléphonique entre un expert et un novice n'est pas sans difficulté, certaines informations ne pouvant être retranscrites qu'à travers le canal visuel. [DSAO⁺97] a néanmoins observé qu'il était possible d'atteindre au téléphone un niveau de performance équivalent à une communication face à face moyennant un grand effort verbal de la part des interlocuteurs. Ces difficultés liées à la communication vont dépendre de la capacité des intervenants à faire coïncider leur savoir, adapter leur vocabulaire et à se faire une représentation mentale de la situation. [Mol88] parle alors de mettre en commun les « sphères personnelles », les éléments de connaissance (expérience, compétence...) étant propres et uniques à chaque individu.

Qui plus est, cette aide ne serait pas neutre dans le sens où elle souligne la dépendance de l'appelant vis-à-vis de l'appelé [RVDWB00]. L'usage et l'efficacité des aides téléphoniques chez les techniciens dépendraient alors de plusieurs facteurs :

- Du niveau d'expérience. Les débutants ont tendance à faire appel à l'assistance téléphonique, les expérimentés privilégiant l'aide écrite.
- De la pression agissant sur la situation de maintenance. Plus le client est important et moins il y a de temps pour réparer, plus les techniciens ont recours à l'aide téléphonique.
- Du niveau de risque encouru. Si le risque est interprété comme important, l'assistance téléphonique est une aide de choix.
- De l'autonomie offerte. Plus l'aide se présente comme étant discrète vis-à-vis de la hiérarchie ou des collègues, plus son usage sera privilégié.

Même si l'assistance téléphonique se révèle être le principal moyen de communication à distance dans le domaine de la maintenance industrielle, nous venons de voir qu'elle n'est pas exempte de défauts, le plus souvent liés à la co-construction à distance d'une représentation du problème de l'appelant ou à l'impression d'une perte d'autonomie.

5.5 Conclusion sur l'aide « classique »

L'opérateur en maintenance industrielle évolue dans un monde où il est confronté à la multiplicité et à la complexité des produits à maintenir. Sans assistance, il peut être délicat d'intervenir sur les divers systèmes. Mais l'aide apportée aujourd'hui est source de certaines complications. En effet, l'aide textuelle se heurte davantage à des problèmes structurels qu'à des problèmes dans l'utilisation qui en est faite. Quant à l'assistance téléphonique, elle est acceptée moyennant des concessions psychologiques, mais son caractère mono-modal conduit à des difficultés liées à l'interaction.

Le support d'information ne remplaçant pas une expertise (ou un diagnostic), et l'expertise distante étant directement soumise aux biais de la technologie classique, les NTIC ne seraient-elles pas une nouvelle voie permettant de tirer avantage des différents systèmes d'aide ?

5.6 Potentiel de la RA dans l'industrie

Bien que les années 60 aient vu naître la réalité augmentée, la concrétisation de son implémentation dans le domaine de la maintenance date du début des années 90. Sa vocation est alors d'accompagner un opérateur dans l'accomplissement de sa tâche en fournissant de l'information contextualisée, et ceci, en temps réel. Cela serait les deux conditions qui réduiraient alors les risques d'erreurs d'exécution [NM98].

Mais l'accent doit être mis sur la transparence dans l'interaction avec la machine. Il s'agit effectivement d'un point-clé de la réalité augmentée dans ce domaine. L'utilisateur doit pouvoir porter son attention sur la tâche à réaliser et non sur l'utilisation de l'outil lui-même, d'où les différentes stratégies que chaque projet peut suivre dans l'élaboration d'un prototype. Aussi, le choix du dispositif d'affichage est important, car l'objectif peut être de pouvoir minimiser le recours aux supports classiques (papier) et laisser ainsi les mains de l'opérateur libres [WN03].

Beaucoup d'études ont été menées début 2000 pour évaluer le réel potentiel de la réalité augmentée dans le domaine de la maintenance. Certaines sont contradictoires comme celles de [WOSL03] ou [HB03], ne permettant pas de conclure quant à l'efficacité de la réalité augmentée face à l'aide « classique » (textuelle). Une première explication possible est peut-être à chercher du côté des limitations fonctionnelles et du caractère encore émergent de la technologie à cette époque.

Mais déjà les principales conclusions sur l'utilisabilité de la réalité augmentée dans l'industrie apparaissent, tendant même actuellement vers un épilogue favorable comme nous allons le voir.

Les apports les plus couramment cités peuvent être réunis dans les deux groupes suivants : intérêts pour la tâche et intérêts pour l'opérateur.

5.6.1 Intérêts pour la tâche

- Réduction des temps de transition entre les tâches et sous-tâches, et donc potentiellement réduction du temps global pour l'accomplissement de l'activité [HF09]. L'application peut fournir de précieuses indications en guidant l'opérateur à localiser précisément l'emplacement d'intervention.

- Si le système est utilisé dans un cadre définissant une application stricte des procédures, la réalité augmentée permettrait alors de réaliser plus rapidement une tâche de montage/démontage, comparativement à un guidage papier [WOSL03].
- Fournir des informations contextualisées et en temps réel [NM98], ou accéder à des informations sur de précédentes séquences de réparation [HF07]. L'information est ciblée, l'opérateur n'a plus besoin de gérer et d'exploiter lui-même la quantité d'information [WOSL03].
- Réduction du nombre d'erreurs [WOSL03]. Selon [MG08], ce n'est pas tant le fait de passer d'un support papier à une application informatique qui améliore le taux de réussite, mais c'est l'apport des augmentations qui jouent un rôle essentiel. Toutefois, même si [TOBM03] constate lui aussi une diminution du nombre d'erreurs commises, les opérateurs ont tendance à moins revenir sur celles-ci pour les corriger (l'utilisateur focalise sur les instructions données par les systèmes de réalité augmentée, faisant entièrement confiance à la machine : c'est l'effet tunnel).

5.6.2 Intérêts pour l'opérateur

- Réduction des mouvements oculaires et de la tête [HF07], ce qui confirme les précédentes expérimentations de [TOBM03]. La réalité augmentée permettrait alors, grâce à l'utilisation de casques Video-See-Through, d'avoir les informations directement sous les yeux et d'être guidé directement sur la zone à agir. Moins de mouvement, donc potentiellement une réduction des troubles musculo-squelettiques [HF09].
- Réduction du « Context Switching », c'est-à-dire des changements liés au passage d'une activité à une autre (déconcentration en se détournant de la tâche pour chercher une information par exemple). L'opérateur peut se focaliser sur ce qu'il fait sans « distraction » [HF07] et ainsi limiter la charge mentale [TOBM03] (effort de transposition, de concentration...).
- Minimiser l'utilisation du support papier permet à l'opérateur de retrouver l'usage de ses deux mains pour les activités qui l'exigent [SH02][WN03].
- Apprendre par l'action (« Learning by doing ») en construisant des connaissances d'une manière active [FV02], ce qui pourrait être équivalent à une formation « sur le tas ». Durant cet apprentissage, mélanger réel et virtuel assurerait une meilleure compréhension des concepts théoriques [SS01], facilitant l'élaboration de représentation des relations spatiales, et de leur évolution temporelle [SH02].

6 LES SYSTÈMES ACTUELS DE R.A.

L'industrie a bien compris les avantages qu'elle pourrait retirer de la réalité augmentée. Mais l'introduction d'une technologie émergente dans le milieu professionnel se fait généralement en plusieurs étapes tributaires des améliorations progressives de la recherche dans le domaine. Ainsi, étalés sur un peu plus d'une décennie, divers projets préindustriels sont apparus, l'objectif étant de tester, adapter et valider les concepts et prototypes qui serviront de base dans le futur.

L'exhaustivité étant vaine, nous allons présenter les systèmes faisant référence, systèmes dévolus à l'aide des opérateurs dans leurs tâches de maintenance. Nous avons choisi de les classer en deux catégories : les systèmes « automatisés » et les systèmes « supervisés ».

6.1 Les systèmes « Automatisés »

Nous entendons par système automatisé tout système de réalité augmentée fonctionnant de manière autonome et largement déterministe. En d'autres termes, le système est un outil « personnel » pour l'opérateur (guide virtuel), l'aide administrée dépendant de scénarii préalablement caractérisés par les concepteurs.

Le projet KARMA

Parmi les systèmes de réalité augmentée dévolus à l'assistance des tâches de maintenance, le projet KARMA [FMS93] est certainement le plus connu, car à l'origine d'un tel concept dès 1993. KARMA, pour « Knowledge-based Augmented Reality for Maintenance Assistance » avait pour but le guidage d'un opérateur dans la réalisation de tâches de maintenance sur des imprimantes laser (Fig.26).



FIGURE 26: Opérateur effectuant une tâche de maintenance en suivant les instructions virtuelles.

Le système était constitué de capteurs ultrasons (ici sur le couvercle de l'imprimante, sur le bac à papier et sur le casque porté par l'utilisateur) et d'un système d'affichage type Optical See Through. La localisation de l'opérateur étant connue grâce aux capteurs ultrasons, une représentation 3D put être calculée²⁴ et affichée sur l'HWD. Le système intégrait déjà des règles définissant la manière d'effectuer une tâche, avec un suivi vérifiant les actions exécutées.

²⁴. Par le moteur graphique maison nommé IBIS (Intent Based Illustration System).

Le projet chez Boeing

Boeing est l'une des premières sociétés qui commença dès 1992 [CM92] à développer son propre système de réalité augmentée pour aider les techniciens dans le câblage électrique de leurs avions (Fig.27).

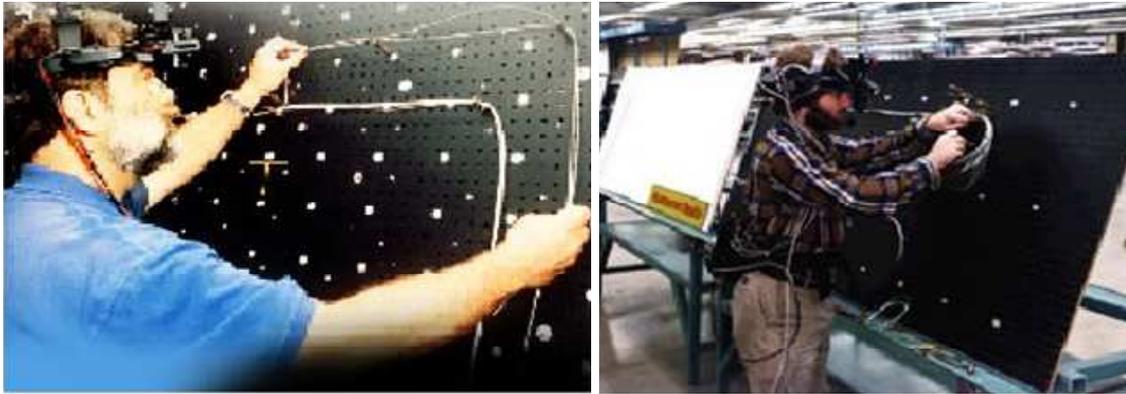


FIGURE 27: Câblage électrique assisté.

Le dispositif était composé d'un casque de réalité virtuelle surmonté d'une caméra. Des cibles étaient disposées dans l'environnement de l'opérateur, facilitant ainsi la localisation de celui-ci. ARTToolkit n'existant pas encore, Boeing développa son propre moteur d'analyse d'images et d'affichage basé sur OpenGL. Le système se contentait alors d'annoter textuellement les images du flux vidéo, et ce en temps réel. Parallèlement à ce projet, des études cognitives ont été réalisées [NM98] pour évaluer l'interaction humaine avec ce nouvel outil et déterminer ses bénéfices. Il fut mis en avant le potentiel de tels dispositifs dans l'avenir au sein de l'industrie.

Le projet ARVIKA (puis ARTESAS)

ARVIKA (Augmented reality for development, production and services) démarra en 1999 pour se terminer en 2003 [ARV99]. Projet national allemand dirigé par Siemens, l'objectif fut d'ordre exploratoire afin de déterminer comment utiliser les technologies de la réalité augmentée dans le secteur industriel, en l'intégrant au plus tôt dans le cycle de vie des produits.

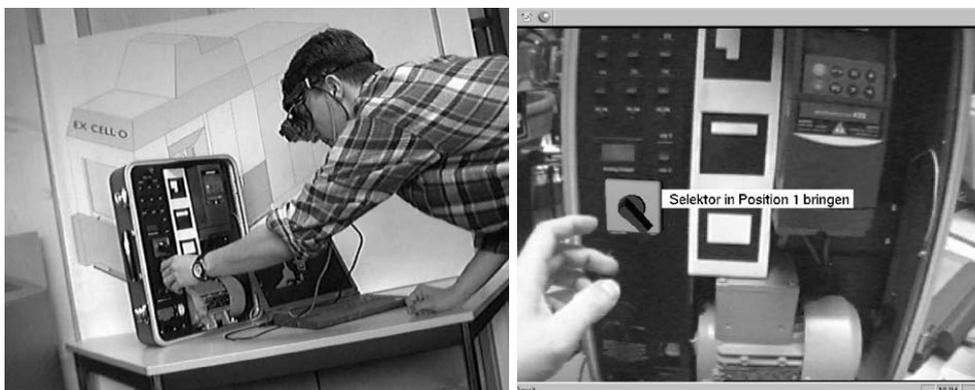


FIGURE 28: Aide apportée à l'opérateur de façon contextuelle [WKS03].

Partant d'une base commune, plusieurs systèmes ont été développés par les différents partenaires. Le système résultant fut en fait le développement d'un serveur informatique sur lequel les différents systèmes étaient capables de sélectionner et présenter la documentation, s'adaptant ainsi au contexte de la tâche de l'opérateur. Comme Boeing, la scène est instrumentée à l'aide de cibles pour localiser l'opérateur (Fig.28), les augmentations étant cependant de nature textuelle et/ou 3D [WKS03].

Le successeur du projet ARVIKA, nommé ARTESAS (Advanced augmented Reality TEchnologies for industrial Service Application), regroupe les mêmes acteurs durant la période 2004-2006. En se basant sur l'expérience acquise précédemment, l'accent était mis sur le suivi de position sans marqueurs pour les environnements industriels difficiles [WVS05], ainsi que sur les aspects ergonomiques des implémentations.

Le projet STARMATE

Le projet STARMATE (SysTem using Augmented Reality for Maintenance, Assembly, Training and Education) fut un programme à l'échelle européenne composé de l'institut Franhofer, EADS, Thales optronic (anciennement Thomson CSF), Dune, Tectnatom et CS-SI. Il s'est déroulé sur la période 2000-2002, son objectif étant de développer et concevoir un prototype permettant d'assister un opérateur dans l'accomplissement de tâches de maintenance sur des systèmes mécaniques complexes. L'autre objectif était de pouvoir fournir une assistance de formation aux opérateurs novices sur ces mêmes tâches [SFC⁺01][SdL03]. L'opérateur était équipé d'un dispositif plus complexe que les projets précédemment décrits, à savoir de lunettes de type See Through, d'écouteurs et d'un microphone couplé à un logiciel de reconnaissance vocale, d'un système de pointage pour désigner les objets à l'application, d'un système de suivi de mouvement (caméra avec filtre infrarouge) et enfin, d'une unité de transmission des données (Fig.29).



FIGURE 29: Principe du projet STARMATE.

Ce projet fut intéressant à plus d'un point en ce qui concerne l'architecture modulaire proposée pour la première fois dans un tel système (superviseur de régulation et conduction des flux de données temps réel, module de reconnaissance et suivi d'objets ou encore module de gestion des scénarii).

Le projet AMRA

Le projet AMRA pour « Augmented Reality Assistance for Maintenance » était un projet RNTL²⁵ placé sous la tutelle du ministère de la Recherche française durant la période 2002-2004. La préoccupation principale fut d'implémenter un système de réalité augmentée mobile pour une utilisation dans le domaine de la maintenance industrielle. Les objectifs étaient multiples : fournir une aide contextuelle à des opérateurs novices, apporter une assistance en donnant l'accès à des informations pertinentes aux agents expérimentés, et rendre disponible l'information via les techniques d'affichage de la réalité augmentée (Fig.30).

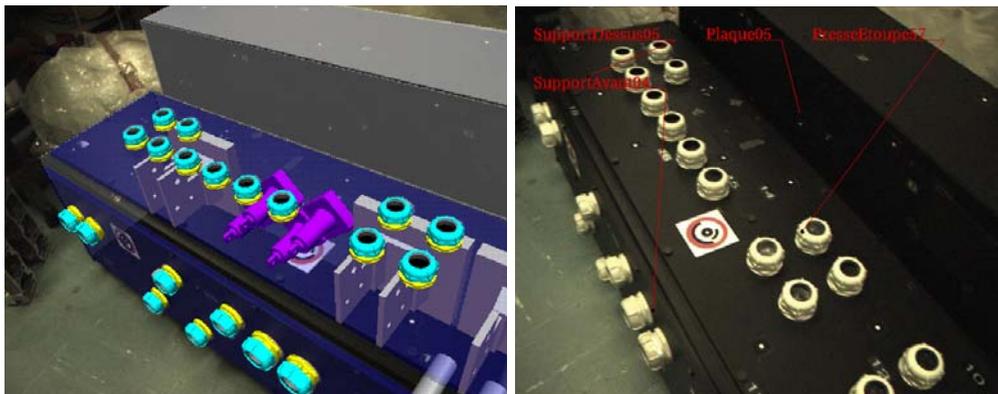


FIGURE 30: Exemples de visualisation de données [AMR02].

Le projet AMRA utilise une visualisation de données type « Window on the World », c'est-à-dire en vision indirecte via un tablet PC [DRM⁺05][Did05]. Notons que ce choix diffère des précédents, et nous pouvons nous interroger quant à l'utilisabilité lorsque l'opérateur a besoin de ses deux mains durant l'exécution d'une action.

Le projet de BMW

La société munichoise Metaio, en collaboration avec BMW, a développé pour celle-ci en 2006 un prototype de « service du futur » pour aider les opérateurs en maintenance sur les automobiles de la marque. L'originalité de ce prototype réside dans la méthode de *Tracking* proposée, particulièrement efficace et robuste, et ce, en temps réel [PHMG06]. Dans le principe, il faut au préalable se constituer, sous différentes orientations, une banque d'images clés représentant le système à maintenir (Fig.31a). Celles-ci obtenues, dans chacune d'entre elles vont être détectés des points caractéristiques. Ces points sont ensuite projetés sur un modèle CAO afin d'établir une correspondance 2D/3D (Fig.31b). Cette phase d'initialisation terminée, le système informatique, pour chaque image provenant du flux vidéo de l'opérateur, va comparer les points caractéristiques obtenus avec ceux précédemment enregistrés. Par comparaison, il sera alors capable de déterminer l'orientation de l'opérateur, et donc d'afficher des informations virtuelles recalées en fonction de l'avancement de la tâche. En ce qui concerne l'équipement, l'opérateur est relié à un ordinateur portable et est équipé d'un casque constitué d'une caméra et de lunettes types See Through pour visualiser les augmentations produites par la firme BMW (Fig.32).

25. Réseau National des Technologies Logicielles.

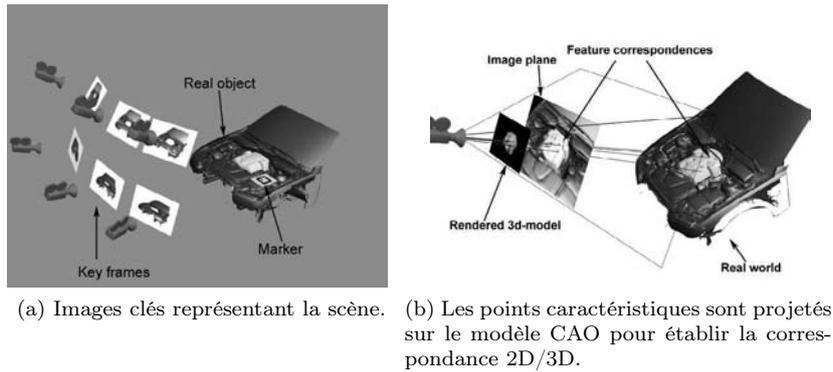


FIGURE 31: Principe de la méthode de Tracking.

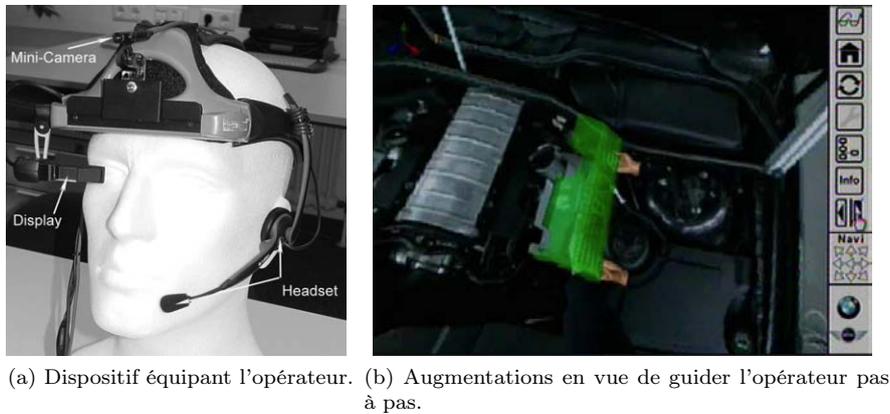


FIGURE 32: Le système de BMW.

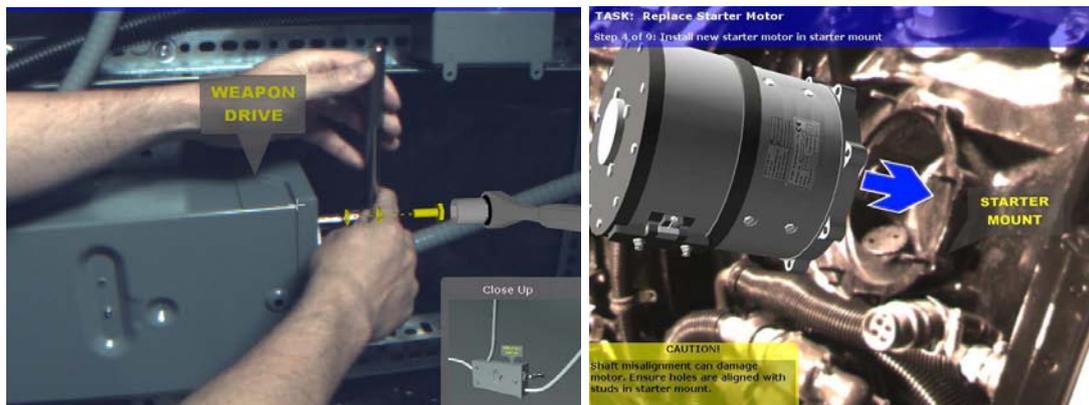
Le projet ARMAR

Débuté en 2007, « Augmented Reality for MAintenance and Repair » explore l'utilisation de la réalité augmentée pour aider à l'exécution de tâches procédurales dans le domaine de la réparation et de la maintenance. Plus particulièrement, [ARM07] s'intéresse à savoir comment les graphismes numériques superposés et recalés en temps réel peuvent augmenter la productivité, la précision et la sécurité du personnel de maintenance [HF09]. Pour cela, Feiner et Henderson ont développé un système de réalité augmentée [HF07] composé d'un afficheur Video See Through binoculaire surmonté d'un dispositif de suivi de mouvements et d'un système de calcul mobile relié à un réseau sans fil (Fig.33). Les concepteurs se sont également intéressés à développer une nouvelle façon d'interagir avec le système en lieu et place des classiques commandes physiques [HF08].

Avec le système proposé par BMW, ARMAR fait partie des prototypes les plus aboutis et fonctionnels à ce jour en termes d'utilisabilité.



(a) Dispositif de visualisation développée pour ARMAR.



(b) Exemples d'augmentations durant une tâche de maintenance et interaction avec le système.

FIGURE 33: Le système ARMAR.

Conclusion sur les systèmes « Automatisés »

Tous les projets présentés proposent des dispositifs plus ou moins similaires dans leur fonctionnement. Ces derniers ont majoritairement utilisé des dispositifs d'affichage tête portée, ont travaillé sur le recalage d'information avec ou sans marqueurs et ont cherché à apporter la connaissance dans la zone de travail de l'opérateur. Tous ces systèmes sont particulièrement pertinents lorsque les tâches sont régies par des règles attribuant une action à une situation précise, c'est-à-dire dans le cadre d'une procédure opérationnelle standard.

Or, nous l'avons vu précédemment, les procédures définies par les concepteurs des divers produits à maintenir sont souvent imprécises, ou fréquemment inadaptées à l'environnement de l'opérateur. Qui plus est, avoir accès à une connaissance dite explicite (voir 8.1.1 page 72) n'est pas forcément suffisant pour savoir comment l'utiliser. Il existe un autre type de connaissance, une connaissance dite tacite, appartenant au domaine de l'expérience, des attitudes et du savoir-faire. Aujourd'hui, ces informations sont difficilement transposables dans l'univers informatique.

Ainsi, les systèmes actuels de réalité augmentée pour la maintenance sont impuissants lorsque survient une situation imprévue. Il est alors parfois nécessaire d'avoir recours à une personne distante ayant le niveau de qualification requis. C'est à cela que tente de répondre la deuxième catégorie de systèmes.

6.2 Les systèmes « Supervisés »

Contrairement aux systèmes autonomes précédemment décrits, les systèmes supervisés²⁶ ne sont souvent qu'un moyen de télécommunication entre un opérateur et un expert, utilisant toutefois les concepts de la réalité augmentée. Autrement dit, l'opérateur est muni d'un système lui permettant de se faire assister par un expert distant dans l'exécution de sa tâche. L'objectif est donc d'ajouter à la modalité audio la modalité visuelle²⁷, et des capacités de présentation/interaction offertes par la réalité augmentée.

Les travaux de Zong, Boulanger et Georganas

Courant 2002, [ZBG02] présentent un prototype d'un système collaboratif pour la téléformation industrielle basé sur la réalité augmentée. Le principe de fonctionnement est de permettre à un opérateur de se faire assister dans sa tâche de maintenance par un expert éloigné en partageant sa vue avec celui-ci. Dans ce prototype, l'opérateur peut manipuler de façon interactive des objets virtuels grâce à des marqueurs (Fig.34).

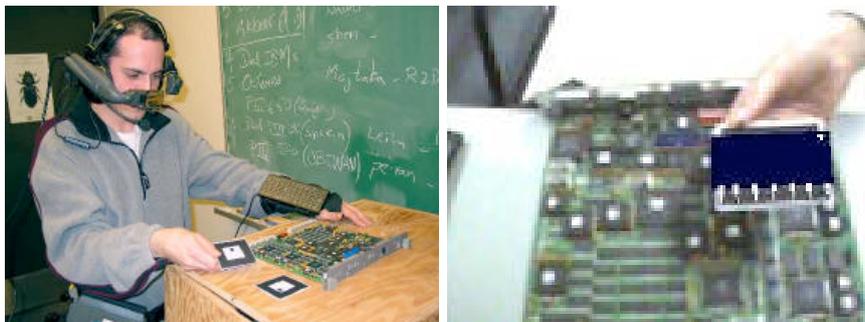


FIGURE 34: Tâche d'entraînement. L'opérateur manipule les objets virtuels via un marqueur.

L'objectif était d'explorer les possibilités d'un espace de travail mobile 3D partagé. Les caractéristiques de ce système sont :

- Présenter des informations virtuelles annotées ou 3D via un dispositif d'affichage tête porté.
- Mettre en place un espace de travail virtuel personnel entourant l'opérateur.
- Partager la vue de l'opérateur grâce à une caméra portée sur la tête de celui-ci.
- Utiliser des objets 3D virtuels plutôt que réels pour s'entraîner.
- Pouvoir discuter avec le collaborateur distant pour partager des conseils ou des informations sur la tâche.

Le système est donc intéressant à plus d'un titre, car il est, dans le domaine de la réalité augmentée, l'un des premiers à faire intervenir l'humain pour superviser une tâche. Les avantages sont multiples pour l'opérateur, l'expert peut partager son expérience en plus de s'adapter à l'avancement de la tâche grâce à la vue partagée. Cependant, l'expert ne peut que donner des directives audio pour guider l'opérateur.

26. Dans la littérature, cette catégorie peut être assimilée à une sous-classe du domaine CSCW (Computer Supported Collaborative Work) appelée « interaction à distance ».

27. Notons qu'il ne s'agit pas ici de visioconférence, même si les concepts sont quasi identiques, la différence étant au niveau de l'implémentation spécifique aux environnements de maintenance.

Le système WACL

Sakata et *al.* veulent aller plus loin que les précédents travaux en permettant à l'expert d'interagir dans l'espace physique de l'opérateur. Ils présentent dans [SKK⁺03] et [SKK06] le système WACL (Wearable Active Camera with a Laser pointer) pensé comme un dispositif pour supporter la télécommunication. WACL est un terminal porté sur l'épaule de l'opérateur et est composé d'une caméra « active », c'est-à-dire motorisée, et d'un laser la surmontant (Fig.35), le tout pilotable à distance par l'expert.

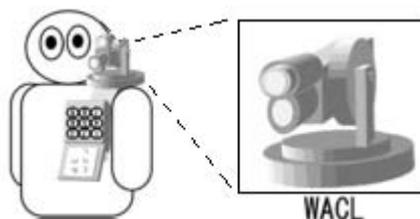


FIGURE 35: Le système WACL.

Grâce à la caméra pilotable, l'expert distant peut donc lui même « explorer » l'espace environnant de l'opérateur pour le guider de façon plus efficace dans sa tâche.

WACL dégage deux principaux avantages : pointer directement des objets réels pour les désigner, et pouvoir indiquer une direction de manière intuitive.

Vis-à-vis des précédents systèmes, il est à noter que dans ce cas précis, nous avons la possibilité d'avoir une interaction directe de l'expert vers l'opérateur. Toutefois, de par la nature du laser, celle-ci est limitée à la désignation.

Le système CAMKA

Développé par la société éponyme depuis 2005, le système CAMKA [Cou06] est présenté comme étant un système de maintenance vidéo assistée (VAM) qui permet à un expert de faire réaliser à distance une opération technique à l'aide d'outils multimédias par un intervenant présent sur le site. À l'origine, le système se présentait sous la forme d'un casque équipé d'un microphone, d'écouteurs, d'une caméra frontale et d'un afficheur vidéo (Fig. 36).

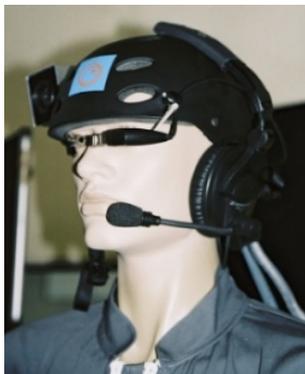


FIGURE 36: Dispositif porté par l'opérateur.

Dans son fonctionnement, ce que capte la caméra est envoyé à l'expert. Ensuite, celui-ci peut capturer une image du flux vidéo, ajouter des annotations puis renvoyer l'image ainsi enrichie sur le dispositif d'affichage de l'opérateur (Fig.37). Ici, l'expert a donc la possibilité d'enrichir des images du réel pour assurer une meilleure compréhension de l'action à accomplir par l'opérateur.

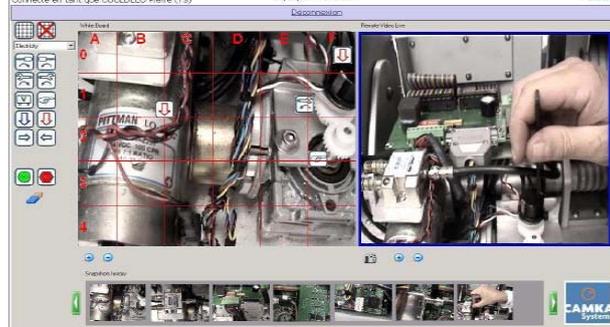


FIGURE 37: Interface de l'expert. À droite, le flux vidéo temps réel. À gauche, une image capturée en train d'être augmentée avant son envoi à l'opérateur.

Courant 2009, la société a gardé le même principe, mais s'est plutôt orientée vers une nouvelle implémentation du système, à savoir non plus sous forme de casque, mais sous forme de pistolet (Fig.38).



FIGURE 38: CAMKA Gun.

Même si le principe reste identique, nous pouvons noter ici que l'opérateur ne peut plus avoir les mains libres pour effectuer sa tâche. Comparée au système WACL, l'interaction pour l'expert est potentiellement plus riche même si les augmentations sont temporellement décalées.

Conclusion sur les systèmes « Supervisés »

Comme nous pouvons le constater, ces systèmes de réalité augmentée dévolus à la téléassistance dans le domaine de la maintenance font l'objet de moins d'attention vis-à-vis de leurs homologues autonomes. Nous venons de voir que les solutions proposées sont parfois assez différentes en termes d'interaction et d'implémentation, et pourraient donc être complémentaires vis-à-vis de l'aide à apporter à un opérateur. Mais il est intéressant de remarquer que toutes les solutions proposées sont très orientées vers la collaboration, ne permettant pas forcément d'apporter une assistance directe en rapport avec la tâche à accomplir. Les concepts de la réalité augmentée permettraient donc de « pousser » plus loin la capacité de l'expert à assister l'opérateur.

7 NAISSANCE D'UNE PROBLÉMATIQUE

Nous assistons ces dernières années à une attente forte du monde industriel pour intégrer de nouveaux outils d'assistance technologique à travers la réalité augmentée. Ce besoin exprime les difficultés rencontrées par les techniciens en maintenance actuellement confrontés à la grande variété de systèmes mécatroniques de plus en plus complexes et au renouvellement des gammes de plus en plus rapide.

La compression des durées de formations et la multiplicité des procédures de maintenance favorisent l'apparition de nouvelles contraintes liées à l'activité des opérateurs. On peut citer par exemple le manque de « visibilité » du système à maintenir et l'incertitude des opérations à réaliser. Ces constatations tendent à obliger les mécaniciens à se former « sur le tas », ce qui peut à terme impliquer un accroissement des erreurs de procédures et par ce biais augmenter les coûts de maintenance, occasionnant des pertes de temps considérables.

Nous avons vu au fil de ce chapitre les problèmes qu'engendre une aide classique. Mais dans un contexte mondial fortement concurrentiel, la demande des industriels à augmenter l'efficacité des tâches du support technique et de maintenance passe par l'intégration de nouvelles technologies de communication. En ce sens, nous avons passé en revue l'ensemble des prototypes d'aide du futur. Nous avons proposé une classification en deux catégories de ces systèmes. La première dite des « systèmes automatisés » regroupe l'ensemble des systèmes particulièrement pertinents lorsque les tâches sont régies par des règles attribuant une action à une situation précise, autrement dit dans le cadre d'une procédure opérationnelle standard. Ces systèmes font preuve d'une maturité technologique leur permettant de prétendre à une intégration industrielle relativement proche.

Mais lorsque surgit une situation imprévue, un opérateur travaillant seul peut avoir besoin d'une aide extérieure, et la personne ayant les compétences et connaissances nécessaires pour l'assister n'est généralement pas sur le même lieu. C'est dans ce cadre que s'inscrit la seconde catégorie dite des « systèmes supervisés ». Leur objectif est de faciliter la communication distante entre les individus. Cela a été rendu possible grâce à l'explosion des débits de communication et du World Wide Web. Cette collaboration entre un expert et un opérateur engendre de multiples bénéfices, mais encore faut-il avoir un système permettant de supporter des interactions distantes proches de celles disponibles lors d'une communication en face à face. Nous avons vu que cela n'est pas forcément le cas !

Notre problématique pourrait alors se formuler de la manière suivante :

« Pourquoi l'homme est-il encore un chaînon incontournable en termes d'assistance, et comment faire en sorte qu'il puisse à distance interagir avec un opérateur de façon presque aussi naturelle que s'il était coprésent ? »

Avec la réalité augmentée, nous pouvons désormais envisager un système de collaboration à distance permettant à l'expert d'être virtuellement coprésent avec l'opérateur. C'est ce que nous allons présenter au chapitre suivant.

Troisième partie

L'ASSISTANCE HUMAINE DISTANTE : PROPOSITIONS

Les technologies d'assistances profitant des concepts de la réalité augmentée sont donc bien trop souvent des outils que l'on pourrait qualifier de « déterministes », au sens où ils sont particulièrement bien adaptés dans le cadre d'une procédure opérationnelle standard. Lorsque la situation devient imprévue, ces outils deviennent inutiles et il est nécessaire de faire appel à une assistance extérieure.

En quoi l'humain est-il donc encore aujourd'hui un maillon indispensable en termes d'assistance? Qui plus est, la communication entre deux individus face à face est régie par des mécanismes complexes, alors que dire de la problématique engendrée par une communication distante! En ce sens, nous avons vu certains systèmes tentant d'améliorer celle-ci, mais est-ce que les critères pris en compte sont suffisants pour soutenir efficacement une collaboration entre individus physiquement éloignés? Comment soutenir les spécificités de la tâche tout en tenant compte du caractère générique d'une tâche quelconque?

C'est dans cette problématique que s'inscrivent les propositions dont nous allons détailler les concepts en fin de chapitre.

La première partie vise à exposer la complexité de ce que représente la connaissance, son acquisition et sa transformation pour justifier l'approche humaine en tant que système d'aide. La seconde partie quant à elle, a pour finalité de mettre à jour les difficultés et caractéristiques à prendre en compte lorsque l'on doit faire collaborer des individus physiquement éloignés. C'est à la lumière de ces arguments que nous allons proposer, grâce aux NTIC, et plus particulièrement à la réalité augmentée, un paradigme d'interaction visant à « simuler » la coprésence d'un expert auprès d'un opérateur dans le cadre d'une tâche de maintenance supervisée.

8 LE POURQUOI DE L'INTERVENTION HUMAINE

Si l'on se réfère au niveau KB²⁸ du modèle de résolution de problèmes de Rasmussen, il est difficile à l'heure actuelle d'automatiser le processus cognitif entrant en jeu lors d'une situation inconnue. Les technologies sont encore tâtonnantes à ce sujet, qu'il s'agisse de systèmes experts s'appuyant sur des moteurs d'inférences, ou de systèmes tentant de reproduire des interprétations proches de celles que le raisonnement humain est capable d'élaborer. Nous allons présenter ici à travers différents points pourquoi l'humain a toute sa place en tant que processus d'aide face aux systèmes informatiques. Nous allons tout d'abord détailler ce qu'est la connaissance, les différentes formes qu'elle peut prendre, avant d'énumérer les différentes théories sur la façon dont elle est acquise, puis transformée en d'autres connaissances. Nous verrons ensuite ce qu'est la compétence et pourquoi l'humain peut devenir un expert. Pour finir, nous mettrons en exergue les atouts d'une interaction humaine tournée vers le même objectif.

8.1 La connaissance

Alors que les théoriciens de l'erreur ont tenté de définir les mécanismes de résolution de problèmes, le lecteur se rendra compte que les méthodes proposées sont dépendantes des connaissances acquises, matières premières sur lesquelles les processus cognitifs agissent. Toutefois, il n'y a pas de définitions substantielles de la connaissance qui fasse l'unanimité, et pour cause, nous allons voir qu'il n'existe pas une connaissance, mais des connaissances, et que leur intégration, maîtrise, combinaisons, sont autant de conditions menant vers l'expertise. Si nous devons choisir une définition du terme connaissance, celle proposée par [Mer03] est intéressante, car elle définit ce qu'est un « contexte » pour une information, composante de la connaissance.

Contexte : le contexte d'une information est donné par une série d'informations permettant à son utilisateur de comprendre la signification en vue de satisfaire les objectifs qui lui sont assignés.

Connaissance : Une connaissance (pour un acteur) sera donc formalisée par un ensemble d'informations la caractérisant et par les informations décrivant le contexte dans lequel elle a été générée et/ou dans lequel elle est destinée à être exploitée.

Mais nous nous apercevons ici que la connaissance est formalisée, bien que cela ne soit pas toujours possible en l'état actuel du savoir.

8.1.1 Les différents types de connaissances

Si l'on se réfère aux sciences cognitives, il existe trois catégories fondamentales de connaissance : les connaissances déclaratives, les procédurales et les conditionnelles. Elles ont été identifiées suite aux travaux de [And83] et de [Tar92]. Chacune d'elle fait appel à des stratégies d'apprentissage différentes :

- Les connaissances déclaratives constituent le savoir théorique, c'est-à-dire les faits, les règles, les lois ou les principes. Ce sont des connaissances dites statiques, dans le sens où elles ne se suffisent pas à elles seules pour permettre l'action. Liées entre elles pour former des propositions,

28. « Knowledge Based », fonctionnement basé sur les connaissances déclaratives (voir 4.1.1 page 44).

elles devront être traduites en procédures et en conditions pour être utilisables. Exemples de connaissances déclaratives : définitions de mots, description de l'objet, connaissance des lois...

- Les connaissances procédurales correspondent au savoir-faire, c'est-à-dire la connaissance du comment de l'action, aux étapes pour la réaliser. Ce sont des connaissances dynamiques, car elles exigent de la pratique et de la répétition pour être maîtrisées. Exemples de connaissances procédurales : réaliser une expérience, faire un plan de travail...
- Les connaissances conditionnelles, concernent le quand et le pourquoi, c'est-à-dire qu'elles se réfèrent aux conditions de l'action. De type dynamique, elles permettent de déterminer le moment et le contexte dans lequel il est approprié d'utiliser telle ou telle stratégies ou démarches. Ces connaissances sont aussi celles responsables du transfert de l'apprentissage, autrement dit de la capacité d'utiliser des connaissances dans un contexte différent de celui pour, ou dans lequel elles ont été acquises. Exemples de connaissances conditionnelles : distinguer des formes différentes, reconnaître un problème exigeant une procédure précise...

L'une des caractéristiques de l'expert²⁹ est un équilibre entre ces trois types de connaissance, et plus particulièrement dans la maîtrise des connaissances conditionnelles. En effet, avoir de vastes connaissances déclaratives ou savoir appliquer de complexes procédures ne suffit pas. On devient expert dans la mesure où l'on peut, dans différents contextes, appliquer ces différentes connaissances et procédures [Tar92].

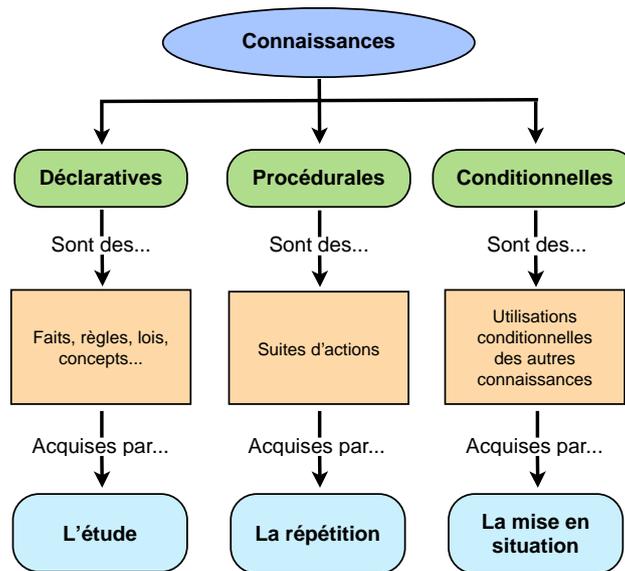


FIGURE 39: Les différents types de connaissance en science cognitive.

Les sciences cognitives ne sont pas le seul domaine à avoir défini différents types de connaissance. De façon pragmatique et à la croisée de diverses disciplines (sciences cognitives, théorie des organisations...), la gestion des connaissances³⁰ peut être définie comme l'ensemble des démarches, méthodes et outils permettant d'analyser, stocker, gérer et capitaliser les connaissances au sein d'une organisation. C'est dans ce contexte que la connaissance a été, grâce aux apports de la

29. Nous allons définir ce qu'est un expert au paragraphe 8.2.3.

30. Traduction de l'anglais « Knowledge Management ».

science cognitive, caractérisée en deux grandes catégories, à savoir les connaissances explicites et les connaissances tacites³¹.

Les connaissances explicites (ou tangibles) sont l'ensemble des connaissances qui peuvent être codifiées ou formalisées. Elles sont inscrites sur un support (physique ou numérique) permettant leur diffusion dans l'espace et le temps. Quant aux connaissances tacites (ou intangibles), elles représentent les connaissances non codifiables, car personnelles. Elles sont le fruit d'expériences, de représentations mentales ou plus généralement de ce que l'on nomme « savoir-faire ». C'est la partie non exprimable de notre savoir au sens de Polanyi [Pol58][Pol66].

Popularisée par Nonaka et Takeuchi [NT95], l'interaction entre ces deux connaissances serait à l'origine de la construction des connaissances générales chez l'individu. Ils ont d'ailleurs proposé une formalisation des différents modes de transformation de ces connaissances (Tab.9).

Du ↗	Vers	Tacite	Explicite
Tacite	Socialisation	<p>Par interaction forte, la connaissance tacite d'une personne peut devenir la connaissance d'autres personnes.</p>	Externalisation
	Intériorisation	<p>Par répétition, on enracine la connaissance explicite dans les séquences pouvant atteindre le stade du réflexe en adaptant le schéma explicite aux conditions spécifiques de l'exécution.</p>	Combinaison
Explicite			<p>Par le biais d'un langage commun et de mécanismes de communication variés, les connaissances explicites de plusieurs personnes sont rapprochées, combinées pour produire, par induction et déduction de nouvelles connaissances explicites.</p>

TABLE 9: Transformation de la connaissance [NT95].

Nous venons de voir, à travers l'énoncé des différentes propositions sur la définition de la connaissance, qu'elle peut être de nature variée. De manière globale, il existe une connaissance formalisable, extrêmement proche de la connaissance déclarative³². Ce type de connaissance peut donc sans mal être utilisé par un système informatique, mais quant à la compréhension du concept sous-jacent, il n'en va pas de même ! En ce qui concerne les connaissances relatives à l'expérience (les connaissances procédurales et conditionnelles), que l'on pourrait assimiler aux connaissances tacites, elles sont plus difficilement transposables vers un système informatique. En effet, est-ce que la capacité à manipuler des concepts permet pour autant d'en assimiler le sens et d'en définir la finalité ?

8.1.2 Théories d'apprentissage

La connaissance, ou plutôt les connaissances, sont donc des informations ou conglomérats d'informations que l'Homme ou la machine doivent utiliser en vue d'atteindre un objectif fixé par une

31. Initialement, c'est Michael Polanyi en 1966 dans [Pol66] qui fut à l'origine d'une telle distinction. Il formula alors le fait que les gens en savent plus qu'ils ne peuvent en dire.

32. [Dés97] propose d'ailleurs de remplacer le terme de connaissance déclarative par connaissance explicite, et connaissance procédurale par connaissance tacite.

tâche. Mais au-delà du système de traitement des connaissances (voir 4.1 page 44), se pose la manière dont toutes ces connaissances doivent être acquises. Nous allons voir, à travers les différentes approches proposées par les sciences cognitives, que ce dont l'Homme est naturellement capable peut se révéler très complexe à modéliser pour un système informatique.

Les sciences cognitives comportent de nombreux courants de pensée lorsqu'il s'agit d'aborder la question des mécanismes d'acquisition et de transformation des connaissances. Si l'on considère les principaux, alors nous retiendrons les cinq suivants : le béhaviorisme, le constructivisme, le cognitivisme, le connexionisme et l'énaclivisme. Mais ces courants ne sont pas traités avec les mêmes égards suivant les domaines où ils s'appliquent. En effet, le béhaviorisme, le constructivisme et le connexionisme sont des approches souvent comparées entre elles quand il s'agit de caractériser la pédagogie envers l'humain. Dans le domaine informatique, où la problématique peut être de modéliser un système complexe capable d'acquérir et de produire des connaissances, ce sont les approches cognitivistes, connexionistes et énaclivistes qui sont opposées.

- Le béhaviorisme (ou comportementalisme) ne se préoccupe pas des processus mentaux intervenant dans l'apprentissage, seuls sont intéressants les comportements observables découlant nécessairement d'un ou plusieurs facteurs (eux-mêmes observables). L'apprentissage est donc une réponse correcte à un stimulus donné. Il s'agit alors d'apprentissage par conditionnement. Cette théorie est en partie due à l'expérience d'Ivan Pavlov en 1889, qui montra un réflexe conditionnel de salivation chez le chien accoutumé à un stimulus sonore annonciateur de nourriture. Ce courant théorique, bien que paraissant dépassé aujourd'hui, est encore privilégié en aéronautique. En effet, l'enseignement y vise un apprentissage sans erreurs par répétition et renforcement des bonnes réponses. L'objectif avoué est de ne pas présenter aux futurs opérateurs les mauvaises façons de procéder, afin que ceux-ci ne puissent pas les confondre avec les bonnes.
- Le constructivisme, contrairement au béhaviorisme, ne considère pas l'être humain comme une boîte noire. Les connaissances acquises ne sont donc pas réduites à l'association stimulus — réponse. L'individu re-conceptualise les informations reçues en regard de ses propres expériences. L'apprenant ne transfère donc pas simplement le savoir dans sa mémoire, mais construit ses propres interprétations à partir des interactions qu'il a avec le monde. La pédagogie constructiviste aurait donc une tendance à transmettre le savoir en mettant l'apprenant au défi, en stimulant et en remettant en cause d'anciennes connaissances pour en produire de nouvelles (c'est la méthode du conflit cognitif). Ce processus d'acquisition des connaissances actif ne peut pas être sans problèmes dans une démarche de formation industrielle extrêmement codifiée. Laisser place à l'interprétation peut dans ce cas être source d'erreurs.
- Le cognitivisme, comme le constructivisme, rejette l'idée selon laquelle les processus d'apprentissage ne sont basés que sur le comportement observable. Mais contrairement au constructivisme, le cognitivisme s'intéresse à la modélisation des processus mentaux, à l'ensemble des phénomènes qui s'intercalent entre la stimulation du sujet par l'environnement et la réponse observable de celui-ci. La genèse du cognitivisme s'ancre dans les années 50 en même temps que se développèrent l'informatique et l'intelligence artificielle. L'humain traiterait donc l'information de manière symbolique, similaire au fonctionnement interne d'un ordinateur. Les différents types de connaissances proposées par Tardif [Tar92], partisan du cognitivisme, s'intègrent donc parfaitement dans ce courant de pensée. Notons cependant encore à l'heure actuelle les difficultés de modéliser en intelligence artificielle la cognition humaine, modélisation visant la manipulation et la production de nouvelles connaissances (systèmes experts, langage Prolog...).

Dans la réalité, l'apprentissage des connaissances n'offre pas une vision aussi tranchée, mais s'inscrit plutôt dans un continuum (Fig.40), comme le pensent certains auteurs [Jon91][EN93]. Si l'on prend le modèle éducatif par exemple, il se situerait entre le cognitivisme et le constructivisme, avec à l'occasion un peu de béhaviorisme [Jon91]. Mais c'est la nature et le contexte d'apprentissage qui déterminent l'approche à privilégier, ce qui est le cas en aéronautique comme nous l'avons vu précédemment.

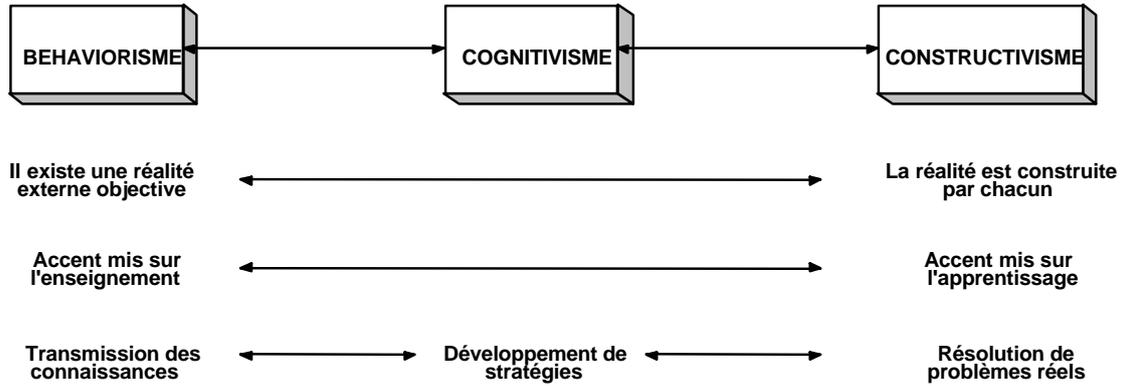


FIGURE 40: Continuum des approches théoriques de l'apprentissage [BRW98].

Dès les années 80, d'autres visions viennent contester l'approche cognitive. Le système cognitif ne serait pas construit à partir de règles, de processeurs logiques, d'informations stockées à une adresse précise, mais à partir d'unités interconnectées formant un réseau, ou suivant une approche d'action guidée par la perception.

- Le connexionisme défend donc l'émergence de processus mentaux ou comportementaux à partir d'interconnexions dynamiques de constituants basiques. Ainsi, chaque constituant fonctionne dans un environnement local, et leur interconnexion ferait apparaître de manière spontanée une coopération globale. La réalité physiologique du cerveau inspira le connexionisme³³, et les réseaux de neurones artificiels utilisés en intelligence artificielle sont certainement les méthodes les plus représentatives de ce mouvement.
- L'énactivisme est présenté comme considérant l'esprit inséparable de l'expérience avec l'environnement. La relation de l'individu avec le monde ne serait pas représentable en termes symboliques, comme le fait le cognitivisme, mais basée sur l'expérience et sur l'action guidée par la perception. L'énaction prône ainsi que le couple « action — interprétation » est à l'origine de la connaissance d'un individu. Chaque individu ou système se forme alors une représentation du monde qui lui est personnelle, forgée sur l'histoire de sa propre perception.

Ainsi, nous venons de voir une brève description des principaux courants tentant d'expliquer ou de modéliser le processus d'acquisition de connaissances. Mais ce qui semble naturel chez l'humain pose encore problème lorsqu'il s'agit de le transposer dans un système informatique. Nous voyons depuis quelques années que la recherche s'est orientée vers le connexionisme grâce aux réseaux de neurones. Mais l'énactivisme commence à prendre de plus en plus d'importance au sein de la

³³. Le cognitivisme et le connexionisme cherchent tous deux à représenter les mécanismes de fonctionnement interne du cerveau humain. D'un point de vue philosophique, l'objectif est commun, la séparation théorique étant donc principalement au niveau représentationnel.

communauté informatique, surtout dans le domaine de la réalité virtuelle. Pour ce qui est de la formation industrielle par exemple, [Oli10] s'était intéressé à la capitalisation de connaissances en environnements virtuels informés³⁴ « dans l'optique de transmettre une meilleure compréhension d'un processus ou du fonctionnement d'un système » à un futur opérateur. Toutefois, quelle que soit l'approche choisie, les comportements et les interactions des systèmes sont souvent prédéterminés (béhaviorisme, cognitivisme) ou tentants de « simuler la créativité » pour offrir un champ de solution paraissant infini (connexionisme, éactivisme).

8.2 De la compétence à l'expertise

8.2.1 La compétence

Si l'on se réfère à la définition de la compétence donnée par le CNRTL³⁵, il s'agit de la « capacité que possède une personne de porter un jugement de valeur dans un domaine dont elle a une connaissance approfondie ». Dès lors, l'humain en tant que ressource dans une situation inconnue serait à même de donner un avis éclairé en se fondant sur ses différents types de connaissances. La compétence s'appuie donc sur la connaissance, mais cela n'est pas suffisant. La compétence n'est pas seulement quelque chose que l'on sait faire, c'est-à-dire un simple « savoir-faire », mais un « savoir y faire ». Il faut donc être capable de se servir de ses connaissances avec pertinence lorsque la situation l'exige, tout en étant capable de juger le bien-fondé des règles à appliquer.

[LB98] donne en ce sens deux définitions de la compétence :

1. Somme de savoirs, de savoir-faire, savoir-être.
2. L'individu est constructeur de ses compétences en combinant et mobilisant des ressources incorporées (connaissance, savoir-faire, expérience...) et des réseaux de ressources dans son environnement (professionnel, documentation, bases de données...).

Ici, la deuxième définition étend la première. Mais toujours d'après [LB98], la compétence professionnelle est un savoir agir validé dans un contexte particulier et en vue d'une finalité.

Ce savoir agir peut prendre toutefois différents aspects. En effet, il faut distinguer la compétence de reproduction et la compétence de production. La première est la capacité, dans l'accomplissement d'une tâche donnée, à reconnaître et à exécuter un agencement connu de savoirs et de savoir-faire. Quant à la compétence de production, le « solutionneur » doit découvrir l'ordre d'utilisation du savoir et savoir-faire pour accomplir sa tâche. Cette recherche dans l'ordre d'exécution suppose alors une maîtrise dans la capacité d'élaboration d'une solution, autrement dit d'heuristique.

34. Autrement appelé EVI. Un EVI est un environnement virtuel doté de modèles à base de connaissances dans lequel il est possible à la fois d'interagir et de permettre des comportements par interprétation de représentations dynamiques ou statiques [Tho09].

35. Centre National de Ressources Textuelles et Lexicales.

8.2.2 Une représentation de la compétence

[Jed94] a proposé quant à lui une représentation (Fig.41) d'un ensemble cohérent des compétences, en faisant l'analogie avec une structure atomique où :

- Les compétences sont positionnées comme des électrons de la couche externe d'un atome.
- Les capacités sont positionnées comme les électrons des couches intermédiaires.
- Les connaissances constituent le noyau de l'atome.

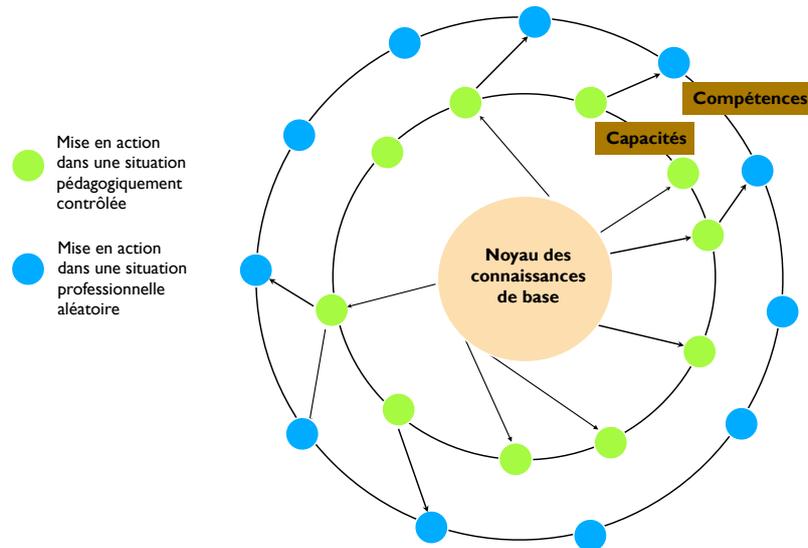


FIGURE 41: Représentation de la compétence sous forme « atomique » selon [Jed94].

Lorsque l'opérateur doit être formé sur une tâche, beaucoup de connaissances sont apprises (souvent des connaissances déclaratives) au prix d'efforts importants. Toutefois, l'intégralité des connaissances n'est pas retenue, d'autant que le système est complexe. Seul un noyau de connaissances est donc acquis. La connaissance initiale va s'enrichir au fur et à mesure de l'expérience (ou de la pratique). À ce moment-là, l'opérateur va se forger ses propres schémas mentaux qu'il utilisera de manière privilégiée, lui permettant ainsi d'être plus efficace (diminution de la charge cognitive). Il développera en quelque sorte des routines d'exécution (connaissances procédurales).

Mais il va également explorer de manière personnelle (en fonction de ses intérêts...) le domaine de la tâche, ce qui lui constituera de nouvelles compétences. En fonction des situations professionnelles rencontrées, l'opérateur pourra aussi acquérir de nouvelles compétences, mais ici de manière aléatoire. C'est ce phénomène qui explique notamment que les personnes ayant suivi la même formation n'ont pas forcément les mêmes compétences à l'arrivée.

Par voie de fait, l'expérience élargit les connaissances, et inéluctablement, celles qui sont ou seront peu utilisées s'estomperont (phase de rétraction des compétences). C'est à ce moment-là que fatalement les erreurs interviendront, et que les systèmes d'aides seront utilisés avec toute la problématique décrite au chapitre précédent.

8.2.3 Vers l'expertise

Pour LeBoterf [LB98], nous avons vu que la compétence (professionnelle) est un savoir validé dans un contexte particulier en vue d'une finalité. Mais alors qu'est-ce que l'expertise ?

Déjà dans les années 80, les frères Dreyfus de l'Université Berkeley ont montré [DD80] que justement c'est la prise en compte du contexte qui est l'une des caractéristiques essentielles de l'intelligence humaine. D'une manière plus intuitive, nous définissons l'expertise comme étant la compétence dans l'excellence. Mais qu'en est-il vraiment ?

Nous allons voir ce qu'il en est en présentant le modèle de Dreyfus d'acquisition de compétences. Il ne s'agit cependant pas à proprement parler d'un modèle, mais plutôt d'une théorie. Celle-ci expose les cinq étapes du passage de novice à expert d'un domaine :

Étape 1 : Le novice

Le novice est une personne avec peu ou sans expérience dans un domaine précis. Son objectif est de parvenir rapidement à des résultats sans comprendre comment les obtenir, la moindre difficulté pouvant alors mettre en échec ses actions. N'ayant aucune intuition sur la manière de résoudre les problèmes, il va s'appuyer sur des règles absolues, faisant abstraction du contexte.

Étape 2 : Le débutant avancé

La personne commence à assimiler les schémas d'action et les règles, et prend conscience du contexte d'application au travers d'exemples qu'un instructeur peut lui fournir. Encore dépendant des règles, il sait choisir une démarche, sans toutefois comprendre la stratégie globale d'application.

Étape 3 : Le compétent

Profitant d'expériences plus nombreuses, le compétent est capable de placer et trier les éléments importants, se faisant ainsi sa propre modélisation de la problématique. Face à une nouvelle situation, il est donc capable de réduire le nombre de paramètres à prendre en compte pour faciliter la résolution d'un problème. Il prend ainsi le contrôle de la situation et peut faire preuve d'initiative. Cependant, il réfléchit peu sur ses méthodes de travail, il n'est donc pas dans une logique d'amélioration continue.

Étape 4 : L'efficace (ou le spécialiste)

Le spécialiste essaie de comprendre le contexte général afin d'en extraire une vision d'ensemble. Son expérience s'enrichit de celle d'autrui pour affiner son « intuition ». En effet, à ce niveau, la théorie acquise (maîtrisée) fait davantage place à l'intuition pour reconnaître une situation et y réagir. Il comprend d'où viennent les règles et les maximes³⁶ et sait quand les appliquer (il ne se contente pas de recettes à suivre).

Étape 5 : L'expert

L'expert ne prend plus de décisions conscientes dans le sens où il sait ce qu'il faut faire, quel est le but à atteindre et le moyen d'y parvenir, le tout de manière intuitive. Il peut avoir du mal à expliquer ces choix qui lui semblent naturels ; il sent les choses comme par « magie ». Enfin, il cherche continuellement de nouvelles méthodes et manières de mieux faire les choses.

36. « Tester ceci », « Quel est le minimum pour... », « Quels sont les points critiques ? »,...

Comme on peut le voir, le novice ne fonctionne pas du tout comme l'expert. Pour qu'un expert puisse communiquer avec des personnes qui ne le sont pas, il doit se mettre au niveau requis. Pour un humain, cette adaptation peut se faire de manière aisée et évolutive au cours d'une discussion, d'une aide apportée... Il n'en va pas de même pour un système informatique. Pour faire un parallèle avec le domaine de l'IHM, on parle « d'adapter l'impédance » [Cae96], comme dans le domaine électrique. L'idée est qu'une interface doit s'adapter à une vitesse cognitivement compatible avec la vitesse de perception et de compréhension de l'utilisateur. L'objectif étant d'éviter la surcharge cognitive de l'opérateur (dépassement des capacités et des exigences nécessaires à l'accomplissement d'une tâche).

Selon Rasmussen, « le développement de l'expertise s'accompagne d'un déplacement des connaissances de l'explicite vers l'implicite », ce qui explique pourquoi l'humain est capable d'en savoir plus que ce qu'il peut en dire ! L'expert ne se résume donc pas à ses seules compétences, car bien que spécialiste, il est capable de moduler son niveau en fonction de l'apprenant. Nous avons donc tout intérêt à l'heure actuelle à remettre l'humain, ou plutôt l'expert, au coeur du système d'aide pour assister un opérateur lorsque la situation l'exige.

8.3 Travail collaboratif

Nous avons bien compris l'intérêt d'un expert comme « moteur d'aide » dans un système que nous avons qualifié de « supervisé ». Mais pourquoi au-delà de l'expertise (qui un jour sera certainement informatisée), l'assistance entre deux humains serait-elle plus profitable que l'assistance entre un humain et une machine ?

La réponse est certainement à trouver du côté de la collaboration. Le mot collaboration tire son origine du latin « com » et « laborare », qui signifient « action de travailler avec d'autres », sous-entendue à une oeuvre commune. Il existe toutefois une différence entre les notions de collaboration et de coopération.

[Pue03] en dresse les principales caractéristiques en tentant de les confronter point à point (Tab.10) :

Coopération	Collaboration
Objectifs organisationnels non pris en compte.	Des objectifs communs sont créés.
Interaction entre individus si besoin est uniquement.	Interactions permanentes entre les participants.
Relations informelles, chaque groupe/individu fonctionne de façon séparée.	Nouvelle structure organisationnelle avec des rôles corrélés.
Pas de définition commune du planning ou des objectifs à atteindre.	Un ou plusieurs projets sont créés avec une perspective de résultats à long terme.
Leadership unilatéral.	Chaque chef de projet est totalement engagé et convaincu de l'intérêt du projet. Le leadership est partagé et distribué entre les participants.
L'information est échangée si besoin est.	Différents niveaux de communication sont établis. La communication est vue comme un élément clef du succès.
Les ressources restent séparées.	Les ressources sont partagées dans une perspective d'effort à long terme tout en étant gérées par la structure collaborative.

TABLE 10: Comparaison des notions de coopération et de collaboration selon [Pue03] (résumé par [Ros04]).

Globalement, [Kva00] estime que la collaboration est une forme d'interaction beaucoup plus forte que la coopération, dans le sens où elle nécessite un esprit de groupe, une adhésion aux objectifs et aux résultats plus importants.

L'intérêt de la collaboration n'est pas accessoire, car d'après [DBBO95], « la collaboration implique la création d'une vision commune des questions à traiter [...]. Elle s'appuie sur un engagement mutuel des participants dans un effort coordonné pour résoudre ensemble le problème posé ». Ce point de vue est partagé par [Gra89] lorsqu'il dit que la collaboration se définit comme étant l'état « à travers lequel les différentes parties participantes, voyant les différents aspects d'un problème, peuvent explorer de façon constructive leur différence et chercher des solutions allant au-delà de leur propre vision limitée du domaine du possible ».

D'après les définitions, la collaboration serait alors un processus complexe que l'on pourrait qualifier de difficilement formalisable. Mais que ce soit une assistance portée par l'humain ou par la machine, il faut garder à l'esprit que le travail collectif est un moyen de surclasser les performances réalisées de façon individuelle, « la somme des résultats du groupe étant supérieure à la somme des résultats individuels » [Bar38].

8.4 Conclusion sur l'intervention humaine

Nous avons vu au chapitre précédent que les dispositifs d'aide dans l'industrie sont extrêmement formalisés (documents techniques, procédures opérationnelles...).

Cependant, nous savons maintenant que ce savoir est de type explicite, et que l'accès à cette connaissance n'est pas suffisant lorsque l'on tente de la mettre en oeuvre. Une personne experte dans un domaine sera alors capable d'aider un opérateur en tentant de transférer son savoir tacite à celui-ci, tout en s'adaptant à son niveau de compréhension. Qui plus est, de cette collaboration va naître un effort commun axé sur la résolution d'une situation imprévue.

9 COMMUNICATION DISTANTE : LES PARAMÈTRES

Nous venons de voir au paragraphe précédent l'intérêt pour un opérateur d'être encore à l'heure actuelle assisté par un expert humain. Qui plus est, nous avons vu que de leur collaboration naît un engagement complètement axé sur la résolution de la problématique qui leur est posée. Dans le cadre d'une collaboration distante, la communication entre les individus est de ce fait moins aisée, que ce soit par téléphone ou par les méthodes implémentées dans les systèmes de réalité augmentée jusque-là proposés (voir 5.4 page 57 et 6.2 page 66). Notre objectif est alors de favoriser cette collaboration distante, en particulier la collaboration informelle, entre un opérateur et un expert, afin de rendre la téléassistance plus simple et efficace.

Pour ce faire, nous nous sommes appuyés sur les principes technologiques des systèmes « Automatisés » et « Supervisés » pour introduire un nouveau paradigme d'interaction basée sur la réalité augmentée, et cela en accord avec les facteurs régissant la communication humaine. Nous allons tout d'abord étudier les problèmes qu'impliquent une collaboration à distance, puis identifier les facteurs essentiels de communication issus de la littérature, avant de décrire plus précisément les solutions que nous proposons pour soutenir l'assistance entre deux individus éloignés.

9.1 Collaborer à distance

9.1.1 Les types de collaboration

Grâce aux télécommunications, l'assistance à distance a pu devenir possible. Progressivement, les technologies permettant de collaborer à distance ont trouvé une désignation commune incarnée par le terme CSCW (Computer Supported Cooperative Work). Introduit pour la première fois en 1984 par Irene Greif et Paul Cashman, ce terme générique³⁷ s'est vu attribuer plusieurs définitions, nous ne donnerons ici que celle de Wilson [Wil91] qui à notre sens est la plus significative :

« CSCW is a generic term, which combines the understanding of the way people work in groups with the enabling technologies of computer networking, and associated hardware, software, services and techniques ».

Le domaine CSCW s'est intéressé à classifier les systèmes en considérant le contexte d'utilisation de celui-ci. La classification la plus connue repose sur les caractéristiques dimensionnelles de l'espace et du temps, à savoir où une action est exécutée et à quel moment elle l'est par un utilisateur. Cette matrice, nommée « CSCW Matrix » par Johansen en 1988 [Joh88] ou encore matrice/temps par Ellis en 1991 [EGR91], fut complétée par Grundin en 1994 [Gru94]. La figure 42 représente cette classification.

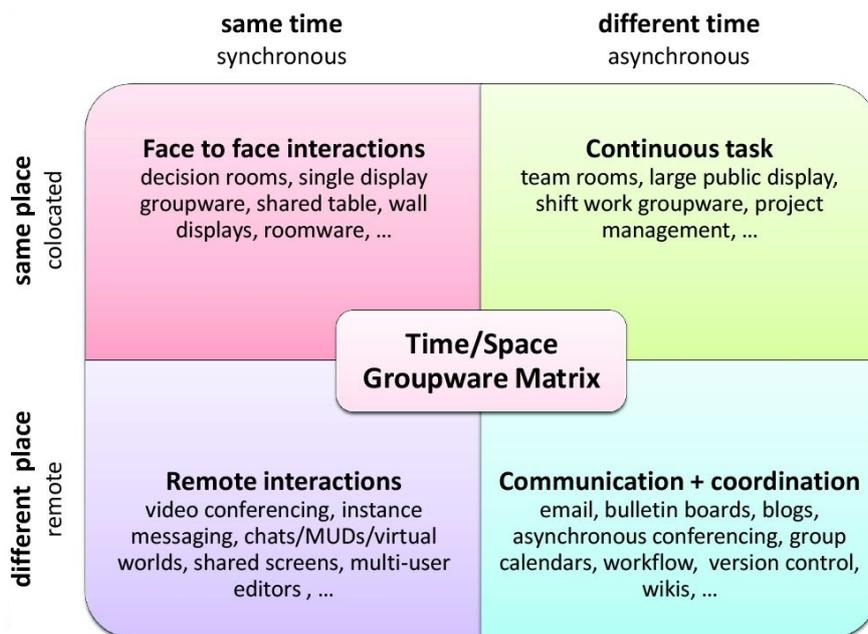


FIGURE 42: Matrice CSCW ou matrice Espace/Temps.

Sur l'ordonnée est représenté l'axe Espace qui considère la distance spatiale entre les utilisateurs (même lieu et lieux différents). En ce qui concerne l'abscisse, c'est-à-dire l'axe Temps, on considère la distance temporelle entre les utilisateurs (même moment et moments différents). Cette taxonomie a l'avantage de montrer la diversité des possibilités de collaboration entre individus. Supporter la collaboration pour le système est donc quelque chose de complexe. Qui plus est, les individus peuvent avoir besoin d'interagir de nombreuses façons, pas seulement limitées au contexte de la

37. La distinction entre collaboration et coopération n'est pas explicitement définie dans la définition de CSCW.

tâche. En ce sens, [LMF04] montre que le processus de collaboration est imprévisible et indéterminé, ce qui nécessairement doit influencer sur la conception et l'implémentation de systèmes tournés vers l'interaction expert — opérateur. Pour [WRSO07], l'état de l'art qu'il a effectué sur les technologies télécollaboratives fait ressortir l'aspect très limité, voire inexistant, de support naturel de communication. Wolf fait apparaître deux critiques majeures, à savoir le non-accès aux données environnantes pour les individus en raison de la séparation spatiale, et la limitation de la perception en termes de communication non verbale. Les technologies actuelles ne feraient donc que « connecter » les gens, au lieu de les « projeter » sur le site distant.

Cette conscience de la situation, c'est ce que Clark appelle le « Conversationnal Grounding », c'est-à-dire les fondations théoriques qui permettent aux individus de comprendre la relation entre la communication verbale et non verbale, puis d'entreprendre des actions en fonction de la perception de l'environnement physique de la tâche [Cla96]. Toute la difficulté est donc de faire partager ce « common ground » qui rend la communication interpersonnelle plus efficace lors des processus de collaboration distants. Le paragraphe suivant est justement consacré à la description des asymétries qui peuvent survenir de façon inopportune ou choisie lors de la conception d'un système de téléassistance.

9.1.2 Les asymétries

Lorsque plusieurs individus sont amenés à collaborer en situation de coprésence, les caractéristiques offertes par cette relation sont symétriques, c'est-à-dire que les interlocuteurs ont les mêmes possibilités d'action et de perception. Au contraire, les systèmes de collaboration à distance introduisent des asymétries, surtout dans la communication. Pour [BHKS98], les participants en situation de collaboration distante doivent de façon similaire bénéficier d'un accès direct aux « données sensorielles » des autres utilisateurs. [BKBB99] fait d'ailleurs ressortir cinq principales asymétries qui peuvent dégrader la collaboration distante :

1. L'asymétrie d'implémentation. Les propriétés physiques du matériel ne sont pas identiques pour les deux participants (ex. : en vidéoconférence, les résolutions des caméras ne sont pas identiques).
2. L'asymétrie fonctionnelle. Il existe un déséquilibre dans les fonctions de communication (ex. : une personne possède la vidéo, l'autre pas).
3. L'asymétrie sociale. Les capacités de communication humaine entre les personnes sont différentes (ex. : seule une personne voit le visage de l'autre, et donc interprète les expressions faciales).
4. L'asymétrie dans la tâche. Les protagonistes n'ont pas la même perception de l'environnement de la tâche (ex. : l'opérateur focalise sur la tâche physique tandis que l'expert distant doit se faire une représentation mentale de celle-ci, ce qui est typiquement le cas par téléphone).
5. L'asymétrie informationnelle. Les informations ou connaissances entre les individus sont très différentes (ex. : un expert tentant d'apporter son aide à un novice).

À la vue de ceci, il semble donc extrêmement difficile de concevoir un système permettant de soutenir la collaboration distante sans introduire d'asymétries de communication, tant matérielles

que relationnelles. Cependant, prendre connaissance de ces asymétries peut aider à minimiser les effets dommageables sur la communication.

En ce qui nous concerne, la collaboration distante entre un opérateur et un expert doit être considérée d'un point de vue du rôle de chaque intervenant, donc introduisant nécessairement des asymétries, par exemple du fait que l'opérateur n'a pas forcément besoin de voir ce que voit l'expert, l'implication contraire n'étant pas vraie.

9.2 Les facteurs importants de communication

Une partie de la communication humaine est verbale, mais en situation de coprésence, les expressions faciales, le regard, la désignation, les gestes et la distance physique entre les individus sont autant de canaux de communication additionnels [SWC93]. Grâce à cela, les personnes tendent à minimiser leurs efforts de communication en mixant les modalités, plutôt que de faire usage d'informations verbales redondantes [CB91]. Dans le cadre d'une tâche de maintenance de visu, la conversation entre les individus se focalise sur l'identification des objets de la scène, sur la description des actions et sur la confirmation que ces actions ont été effectuées correctement. Ceci met en lumière toute l'importance que revêt le contexte dans ladite conversation. Les mécanismes de communication et le contexte sont donc étroitement liés et jouent un rôle important lorsque deux interlocuteurs visent à atteindre un même but.

9.2.1 Partage de l'espace visuel

En situation de coprésence, les protagonistes partagent le même espace visuel. L'objet de la tâche leur apparaît alors de façon identique. Nous avons vu dans le chapitre précédent que le défaut des systèmes téléphoniques était justement l'obligation pour l'expert de devoir se représenter de façon mentale l'espace physique de l'opérateur.

Dans la collaboration à distance, l'expert n'a donc pas forcément de relation spatiale avec les objets environnants de la tâche [KFS03]. Cette contrainte se trouve plus ou moins levée dans les systèmes introduisant la modalité vidéo, des études ayant démontré que le processus de Conversationnal Grounding est amélioré quand les collaborateurs ont accès à l'espace visuel de la tâche [CK04][GKF04][KGF02][KFS03]. Par plus ou moins nous entendons faire remarquer qu'il ne suffit pas de proposer un flux vidéo pour régler le problème. En effet, l'angle d'ouverture de la caméra filmant ce que voit l'opérateur joue un rôle important. Plus l'angle d'ouverture est élevé, plus l'expert aura une vision élargie de la scène. Ceci est un paramètre important lors d'une communication distante, car l'expert doit pouvoir disposer d'un « espace visuel périphérique » afin de faciliter la prise de conscience de la situation. Cela aurait un effet direct sur la coordination avec les actions de l'opérateur, et favoriserait également la connaissance permanente du statut des travaux [FSK03]. D'ailleurs, [HL91] avait déjà constaté que le manque de vision périphérique dans la collaboration à distance faisait perdre l'efficacité de la communication dans l'accomplissement d'une tâche. Toutefois, il est difficile d'offrir un angle de vue approchant la vision humaine, et nous verrons que cela n'est pas forcément nécessaire, principalement à cause des différentes fonctions relatives aux différents angles de vues de l'œil humain (voir 13.1.2 page 105).

Mais partager seulement une vision périphérique de la scène avec l'expert n'est pas sans poser un autre problème. Plus l'angle de prise de vue est grand, moins le flux vidéo contiendra de

détails³⁸, et ceci impactera directement sur la façon de percevoir de manière précise les informations habituellement perçues par le centre de l’œil humain³⁹. Une vidéo ne contenant que peu de détails obligera l’expert à multiplier les demandes de rapprochement et d’éloignement à l’opérateur.

Pour résumer, il est important de rendre compte à l’expert de l’environnement dans lequel évolue l’opérateur, et ceci passe par la modalité vidéo suivant deux aspects : offrir une vision précise de ce que voit l’opérateur, et lui faire bénéficier d’une vision périphérique de la scène. Comme l’affirme Fussel dans [FSY⁺04], plus l’objet de la conversation auquel on se réfère est visible et partagé par les télécommunicants, meilleures sont les performances en termes de collaboration.

9.2.2 Les références ostensives

Le succès d’une communication lors d’une collaboration distante ne dépend cependant pas uniquement de l’accès partagé à la vue de l’espace de l’opérateur. Lors d’une communication face à face, l’intégrité de la collaboration passe aussi par les actions gestuelles, le discours et le geste étant alors étroitement liés. Cette relation a été largement mise en avant dès les années 90 avec les travaux de [Tan91],[BOO95], [HL96] ou encore [Fra00]. Mais c’est le travail de Clark avec sa théorie du Conversationnal Grounding qui a défini les causes de cette interdépendance [CS89, CB91, Cla96, CK04].

Lorsqu’ils parlent, les protagonistes utilisent des gestes pour clarifier ou améliorer leur message [BOO95]. Le geste de désignation est alors utilisé pour se référer à l’objet de la tâche et à sa localisation. D’après [Lan05], « on attend du geste qu’il complète une expression référentielle à laquelle aucun référent ne pourrait être attribué sans sa participation. Le geste est ainsi vu comme désignant directement le référent ciblé, et l’on ne s’imagine pas qu’il puisse introduire une ambiguïté supplémentaire ». Mais justement, dans certains cas, comme cela l’est souvent dans l’industrie sur des systèmes généralement complexes et denses⁴⁰, il est possible que le geste de désignation puisse introduire une ambiguïté, notamment à cause de l’imprécision de la direction de pointage, à cela s’ajoutant le décalage du point de vue entre les interlocuteurs. Nous verrons dans nos propositions comment contrecarrer cet effet, grâce à l’action combinée de notre paradigme d’interaction et du mode de visualisation destinée à l’opérateur lors d’une collaboration distante⁴¹.

Comme nous l’avons précisé, dans une situation de communication, nous retrouvons le geste et le discours qui lui est associé. Généralement, le discours contient alors des termes déictiques en rapport à la désignation. Les deixis sont les termes ayant recours à la situation de communication. Dans l’exemple « Prends cet objet ! », le groupe nominal « cet objet » constitue une forme de deixis. À cet énoncé est associé le geste de désignation, que l’on appelle alors ici geste déictique (ou geste ostensif). L’association d’un geste déictique⁴² et de deixis est alors nommée « référence ostensive ».

Il existe plusieurs formes de deixis, les plus courantes étant les deixis de personne (« je », « tu »,...), les deixis spatiales (« ici », « là »,... et les marqueurs déictiques « -ci », « -là »,...), et les deixis

38. Quelle que soit la résolution du capteur. Bien entendu, l’évolution technologique fera qu’un jour le niveau de détail perceptible surpassera les capacités de l’œil humain, même pour une prise de vue grand-angle.

39. Durant nos différentes recherches dans la littérature, nous n’avons pas relevé d’études traitant ce problème en réalité augmentée.

40. Les cibles potentielles peuvent avoir une telle ressemblance que les mots employés pour les différencier sont insuffisants.

41. Il est paradoxal de remarquer que c’est justement les systèmes de collaboration à distance qui peuvent, grâce à la réalité augmentée, régler un problème de communication face à face.

42. Le geste déictique peut aussi prendre la forme d’un regard directif ou d’un mouvement de tête.

temporelles (« tout de suite », « maintenant », « après »,...). Ce sont les deixis spatiales qui ont fait l'objet de nombreuses études, car majoritairement utilisées lors d'une tâche physique. Kramer a même étudié la sensation de présence lors d'une collaboration distante en fonction des références linguistiques utilisées par les personnes [KOF06]. Il conclut que l'utilisation de deixis locales (« celui-ci », « ici »,...) et de deixis distantes (« celui-là », « là-bas »,...) ne reflète pas la sensation de présence, mais que leur utilisation varie durant la conversation en fonction de la tâche. L'étude de Bao met en avant toutefois une différence notable en fonction de la taille de l'afficheur vidéo [BG09], les deixis locales étant plus utilisées pour des afficheurs de grande taille (60"x40").

Notons qu'en plus des deixis locales et distantes, Bauer avait proposé une terminologie [BKS99] listant les « Referential Speech Act » en trois catégories : les références absolues (ex. : « le deuxième trou sur la droite du coin gauche »), les références relatives (ex. : « le trou au-dessus du câble jaune ») et justement les références déictiques (sans toutefois faire la distinction entre les locales et les distantes). Lors de ces études sur l'utilisation de la désignation en vidéoconférence, il est mis également en avant le fait qu'utiliser un télépointeur permettrait à l'expert de guider et diriger efficacement l'opérateur, ce que confirme également Fussell dans [FSY⁺04]. Les références ostensives sont donc un puissant moyen de communication en face à face, et ont démontré leur intérêt lors d'une communication distante à travers les études que nous venons d'exposer.

Résumé des facteurs importants de communication :

Pour s'assurer qu'une communication distante puisse se rapprocher, en termes d'efficacité de collaboration, d'une situation de communication face à face, il faut tenir compte de deux aspects majeurs durant celle-ci :

- Les protagonistes partagent le même espace visuel. Dans la collaboration à distance, l'expert n'a pas forcément de relation spatiale avec les objets. Celui-ci doit donc pouvoir disposer d'un espace visuel périphérique afin de lui faciliter la prise de conscience de la situation. Ceci aura un effet direct sur la coordination des actions avec l'opérateur, avec l'avantage de connaître en permanence le statut des travaux. Le manque de vision périphérique ferait donc perdre l'efficacité de la communication dans l'accomplissement d'une tâche. En termes d'espace visuel, il faut également tenir compte de la vision de précision qui est utilisée pour faciliter l'examen de détails.
- Les protagonistes ont la possibilité d'utiliser des références ostensives, c'est-à-dire des deixis (« ça! », « là! ») associées à des gestes de désignation. De nombreuses recherches suggèrent l'importance de la désignation d'objet dans un travail collaboratif. Ce type d'interaction est directement en rapport avec la notion d'espace visuel partagé cité ci-dessus.

10 PROPOSITIONS POUR SOUTENIR LA COLLABORATION DISTANTE

Dans le chapitre précédent, nous avons pu constater que les systèmes utilisant la réalité augmentée sont soit très orientés vers la maintenance avec un opérateur seul face à son dispositif, soit orientés vers la collaboration, avec des possibilités ne permettant pas forcément d'apporter une assistance directe de la tâche et avec des interactions de communication limitées. Nous venons de voir au cours de ce chapitre l'intérêt qu'il pouvait y avoir à l'heure actuelle de faire encore confiance à l'expertise humaine pour assister un opérateur lors de situations prévues, mais surtout imprévues. Faire communiquer des individus est un processus naturel lorsque nous sommes en situation face à face, les choses se compliquant à distance. En ce sens, nous avons analysé les éléments qui nous semblent être capitaux pour soutenir une communication distante, et donc améliorer la collaboration. C'est dans ce cadre que doit s'inscrire le futur des systèmes d'assistance à distance soutenant la collaboration sur des tâches de maintenance industrielle, voire des tâches plus générales.

Nous proposons donc le principe général de fonctionnement d'un tel système (Fig.49) :

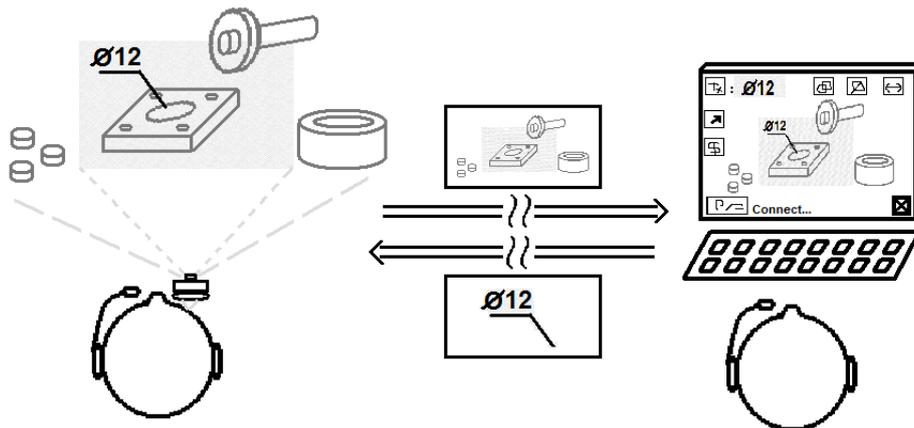


FIGURE 43: Notre proposition du principe de fonctionnement d'un système d'assistance à distance.

L'opérateur doit être équipé d'un dispositif de visualisation tête-porté. Il aura donc les mains libres, et doit pouvoir visualiser des augmentations en provenance de l'expert dans son champ visuel, et ce, de manière « See-Through » directe ou indirecte (voir 2.3 page 33). Ce dispositif devra être capable de retranscrire à l'expert ce que voit précisément l'opérateur (Flux O), puis retranscrire l'espace visuel périphérique de la tâche (Flux P). Parmi les deux flux que recevra l'expert en temps réel (en plus de la modalité audio), il aura la possibilité d'augmenter le Flux O grâce à notre paradigme d'interaction (voir paragraphe 10.2 page 90). Les augmentations devront alors être envoyées en temps réel sur l'afficheur de l'opérateur.

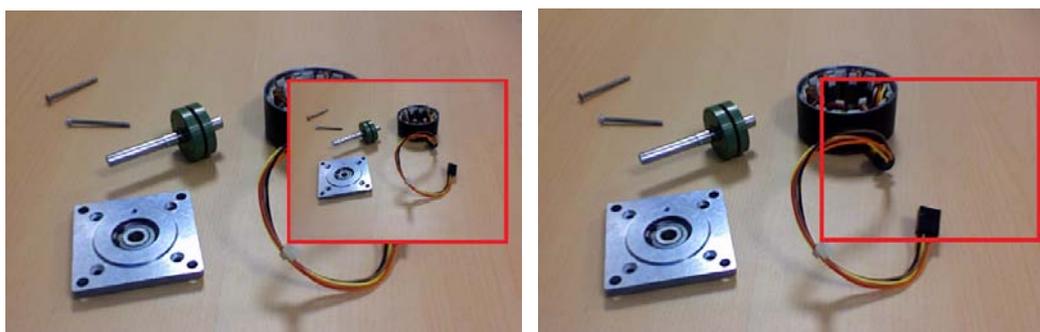
Nous allons décrire plus précisément la manière de percevoir l'espace physique pour l'expert, la manière de percevoir les augmentations pour l'opérateur, et le paradigme d'interaction mise à disposition de l'expert pour permettre l'utilisation de références ostensives.

10.1 La façon de percevoir

Pour l'opérateur

Alors que la réalité virtuelle requiert une immersion totale, un opérateur utilisant un système de visualisation pour la réalité augmentée ne devra pas perdre la perception de son environnement réel et immédiat. En effet, dans l'industrie, les situations potentiellement accidentogènes sont nombreuses sur les systèmes à maintenir, et l'opérateur doit à chaque étape appréhender de façon naturelle son environnement. Les contraintes de manipulation des systèmes d'aides classiques (ordinateurs, livres...) sont aussi un frein pour l'opérateur, qui doit en permanence pouvoir trouver l'usage de ces deux mains. L'aide apportée via les augmentations virtuelles doit donc être aussi transparente que possible pour l'activité usuelle de l'opérateur.

Le fait d'utiliser un système de visualisation tête-porté répondrait alors à ces deux critères ; avoir les mains libres et avoir les informations virtuelles directement devant le regard. Mais pas seulement ! Les implications d'un système tête-porté sont plus profondes. En effet, dans le premier chapitre, nous avons vu que le système tête-porté était de type See Through, ce qui peut supposer le maintien d'un champ visuel orthoscopique⁴³ pour l'opérateur (Fig.44).



(a) Affichage classique d'un Video See Through. (b) Affichage orthoscopique (Video ou Optical).

FIGURE 44: Représentation du champ visuel de l'opérateur portant un dispositif See Through sur l'oeil droit (intérieur du rectangle rouge).

Sans augmentations virtuelles, le champ visuel de l'opérateur paraît alors aussi naturel que possible. Qui plus est, de par la technologie actuelle, l'inconvénient du faible angle d'affichage de ce type de système devient ici un avantage certain. La surcharge du champ visuel par des éléments virtuels est ainsi évitée pour l'opérateur (dans une moindre mesure). Seule une « fenêtre » de son champ visuel est augmentable, ce qui facilite et assure l'appréhension de son environnement immédiat. L'argument est d'autant plus vrai si l'on fait le choix d'un système monoculaire, laissant toujours ainsi au moins un oeil dans la réalité.

Enfin, l'orthoscopie apportée par une visualisation See Through connaît deux autres avantages majeurs. Le premier est qu'elle évite le phénomène de « distraction » vue en 4.3 (p. 49). Ne pas avoir à soustraire son regard de la tâche, même un court instant, limite fortement le détournement d'attention de l'esprit vis-à-vis de l'activité mentale et/ou manuelle. Deuxièmement, en ajoutant des

43. L'orthoscopie implique un point de vue similaire à la prise de vue effectuée par une caméra. Dans le cas d'une personne portant un dispositif de visualisation pour la réalité augmentée, celle-ci doit voir exactement ce qu'elle aurait dû voir si elle n'avait pas eu le système devant les yeux.

informations virtuelles de manière orthoscopique pour l'opérateur, nous supprimons le phénomène de surcharge mentale liée à la transposition (la gymnastique intellectuelle entre la scène représentée sur une image par exemple, et sa relocalisation dans la scène réelle, est un processus cognitif fatiguant à terme et potentiellement générateur d'erreurs).

Pour l'opérateur donc, les impératifs que nous venons de décrire orientent vers un système d'affichage See Through monoculaire.

Pour l'expert

En ce qui concerne l'expert, nous venons de voir précédemment que lors d'une communication distante, celui-ci n'a pas le même rapport spatial que l'opérateur vis-à-vis de l'espace physique de la tâche. Il faut donc lui en rendre compte. En ce sens, le système d'affichage de réalité augmentée de l'opérateur va jouer un rôle essentiel. Nous sommes partis du fait que le meilleur moyen pour l'expert de simuler sa présence était encore de se mettre virtuellement à la place de l'opérateur. En voyant ce que voit exactement l'opérateur, l'expert s'affranchit du décalage de point de vue habituellement présent lorsque deux individus sont côte à côte, problèmes évoqués au paragraphe 9.2.2. Il peut ainsi « jeter un coup d'oeil » à travers la vision de l'opérateur et le guider⁴⁴ de façon plus efficace.

Bien entendu, pour ce faire, il faut impérativement avoir accès au point de vue orthoscopique de l'opérateur. Bien que conceptuellement simple, nous verrons au chapitre suivant les contraintes techniques qui s'appliquent à cet accès.

Cette vision orthoscopique est la vision que nous avons qualifiée de vision précise lorsque nous avons évoqué la problématique du partage de l'espace visuel. C'est une vision restreinte, mais exacte et détaillée, utilisée lorsque l'on porte attention sur l'action à accomplir. Cette vision correspond au Flux O du principe de fonctionnement général que nous avons décrit plus haut, et donc au flux augmentable par l'expert.

Pour ce qui est de la vision périphérique nécessaire à la prise de conscience de la situation, l'idée est, toujours grâce au système d'affichage de l'opérateur, de pouvoir capturer, via une caméra grand-angle, l'environnement immédiat de celui-ci (le Flux P). Ainsi, l'expert aurait accès à une représentation panoramique de l'espace de la tâche, représentation évoluant avec les mouvements de l'opérateur. Ce mode de visualisation possède également l'avantage de pouvoir resituer la vision précise dans le contexte de la tâche⁴⁵.

Nous proposons alors de rendre compte de la situation à l'expert via deux flux vidéo partagés de la vision de l'opérateur, grâce au système tête-porté de celui-ci. Cela se traduirait du côté expert par une interface (Fig.45) représentant la vidéo périphérique de la scène dans laquelle serait incrustée la vidéo orthoscopique (principe du PiP ou Picture in Picture).

44. Nous nous risquons à faire l'analogie que l'opérateur devient ainsi « pilotable » comme le ferait un téléopérateur avec un robot, l'avantage ici étant un humain en termes d'intelligence.

45. On évite ainsi l'effet tunnel qu'implique la vision précise à cause de son angle de vue restreint.

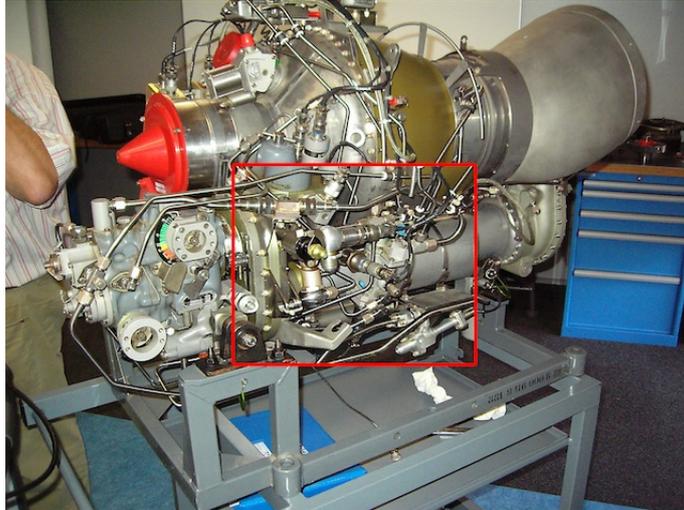


FIGURE 45: Proposition de visualisation de l'espace de l'opérateur pour l'expert. La vue orthoscopique (intérieur du rectangle rouge) est placée dans la vue panoramique. Nous voyons ici l'intérêt de pouvoir restituer la vue orthoscopique par rapport à la vue panoramique.

10.2 Paradigme d'interaction : P. O. A.

Les références ostensives régissent en grande partie la communication humaine lors de l'exécution d'une tâche en situation de coprésence. Mais celles-ci ne peuvent avoir de sens seulement si l'expert distant est capable de partager de manière visuelle le contexte de l'opérateur. Grâce à la proposition précédente, il est rendu compte à l'expert de l'environnement de l'opérateur. Ce partage permet alors de réduire l'asymétrie de la tâche au sens de Billinghurst, et ouvre ainsi la perspective d'user de références ostensives, mais pas seulement ! Les deixis étant directement reliées aux gestes, il convient alors de s'intéresser aux différents types de gestes qui facilitent et clarifient le discours [BOO95][McN92].

Bien qu'il existe une multitude de classifications de gestes établis durant différents travaux [McN92] [Efr41] [EF69], tous sont d'accord pour faire une distinction entre les gestes déictiques et les gestes représentant le mouvement des mains.

Voici leur définition et leurs fonctions possibles lors d'une utilisation dans une tâche collaborative physique :

- Les gestes déictiques : orienter le doigt ou la main vers un point de l'environnement. Généralement utilisés en concordance avec des deixis, ils réfèrent vers un objet ou indiquent une direction.
- Les gestes iconiques : utiliser les mains pour représenter des formes. Par exemple, une personne peut former un cercle avec les mains pour indiquer la forme d'un objet rond.
- Les gestes spatiaux : servent à définir des distances en utilisant une seule main, voire les deux. Ils sont souvent utilisés pour spécifier à autrui de quelle distance un objet doit être déplacé, généralement pour des distances inférieures à l'envergure des deux bras.
- Les gestes kinétiques : spécifiques à la description des mouvements. Par exemple, un expert reproduira avec ses mains la façon dont deux pièces viennent s'encastrent l'une dans l'autre.

Nous avons vu au chapitre précédent quelques exemples de systèmes « Supervisés » tournés vers l’assistance en maintenance. Ces systèmes ne tiennent que peu compte de l’importance des gestes dans la communication, et l’implémentation de la désignation se révèle souvent limitée ou asynchrone. D’autres systèmes, purement collaboratifs, comme DOVE [OFc⁺03], intègrent ces différentes capacités gestuelles à travers l’utilisation de tablettes graphiques. L’expert est alors plutôt face à ce que l’on nomme des « digital sketches », reproduisant des gestes en les dessinant.

Les expérimentations sur ce genre de système ne concluent pas quant à réduire le temps d’exécution de la tâche, vis-à-vis de l’utilisation du téléphone par exemple. Mais les bénéfices sont certains en termes de compréhension mutuelle [FSY⁺04]. Notons que l’expert ici dessine sur des images provenant de points de vue statiques. Il existe donc un décalage entre ce que voient les deux protagonistes, avec les différents problèmes que nous avons déjà évoqués.

Nous citerons également les travaux de [LWAS07], où l’objectif fut de présenter à l’opérateur la surimpression des bras et des mains de l’expert sur un moniteur affichant l’espace de la tâche de ce même opérateur. Cela a impliqué un point de vue fixe, et donc difficilement transposable dans le cadre d’une tâche de maintenance. L’expert avait le choix d’utiliser ses mains pour faire des gestes, ou un curseur en guise de pointeur. Bien que la désignation soit intuitive avec les mains, les utilisateurs ont utilisé plus souvent le curseur pointeur pour ce faire, le pointeur s’étant révélé suffisant et très précis.

Mais en dehors de la désignation, cette étude confirme que la gestuelle des mains est un moyen adéquat, ayant un impact sur la perception de l’interaction.

Au regard des faits exposés, nous devons tenir compte pour soutenir la communication distante des éléments suivants :

- La désignation peut être symbolique, c’est-à-dire qu’il n’est pas nécessaire d’avoir une représentation fidèle du geste humain.
- Les gestes iconiques et spatiaux n’ont trouvé que peu d’usage dans les systèmes étudiés.
- Les gestes kinétiques eux, ont été largement utilisés, soit par le biais de dessin sur un plan de vue statique, soit en mimant le mouvement avec des mains en surimpression. Dans le premier cas, l’inconvénient majeur est qu’il peut être difficile de représenter le mouvement en 2D, sans compter le temps pour le dessiner ! Le deuxième cas est plus intéressant, mais tout aussi statique en termes de point de vue.

C’est à la lumière de ces indications, du fait que nous rendons compte à l’expert de la situation de façon dynamique, et des principes de la réalité augmentée, que nous proposons un paradigme d’interaction pour l’expert. Notre paradigme, nommé P.O.A. pour « Picking Outlining Adding », s’inspire de tout cela en se basant donc sur trois points :

Picking

Cette simple interaction est représentative du geste déictique. L'expert peut rapidement pointer ce qu'il veut montrer en cliquant simplement sur le flux vidéo représentant la vision orthoscopique de l'opérateur. Celui-ci verra donc apparaître devant ses yeux un modèle iconographique (flèche, cercle...) à l'endroit exact où il doit porter son attention, levant ainsi toute ambiguïté (Fig.46). Qui plus est, les termes déictiques deviennent pleinement utilisables, évitant alors de se perdre en explications descriptives.

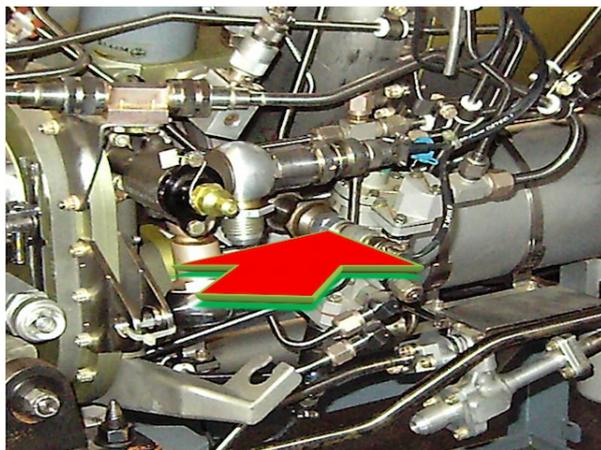


FIGURE 46: Vue augmentée de l'opérateur après une interaction de type « Picking ». Ici, l'élément désigné est clairement discriminé parmi tous les autres. Inutile donc de le décrire, ce qui ici peut se révéler ardu !

Outlining

Grâce à la réalité augmentée, nous avons la possibilité d'aller plus loin que le simple Picking. Nous pouvons ajouter des éléments virtuels au regard de l'opérateur de manière à souligner les informations vocales données par l'expert, ou pour augmenter la compréhension d'une situation donnée. Il peut s'agir alors de mettre en valeur un élément simplement en affichant ses contours⁴⁶ visibles, ou cachés (principe de la réalité à compréhension ou visibilité augmentée). Nous estimons qu'un simple rendu non photoréaliste est suffisant, évitant ainsi les informations inutiles et potentiellement dangereuses en termes de perception (confondre réalité et virtualité).

Pour cette seconde interaction, l'expert, sur le principe du Picking, clique sur un élément du système à maintenir. L'opérateur verra alors apparaître devant lui la représentation 3D de l'élément considéré, et ce de manière complètement recalée par rapport à la réalité. Clairement orienté vers la tâche, l'Outlining oblige en amont à posséder les modèles 3D adéquats, et d'être en mesure de recalibrer l'information.

Dans le cadre d'une utilisation plus générale, il nous est donné grâce à la réalité augmentée de pouvoir afficher n'importe quel autre type d'information (textuelle...), information ayant pour objectif de soutenir le discours de l'expert (Fig.47).

46. On « identifie » ainsi l'objet, la pièce mécanique ou un sous-système.

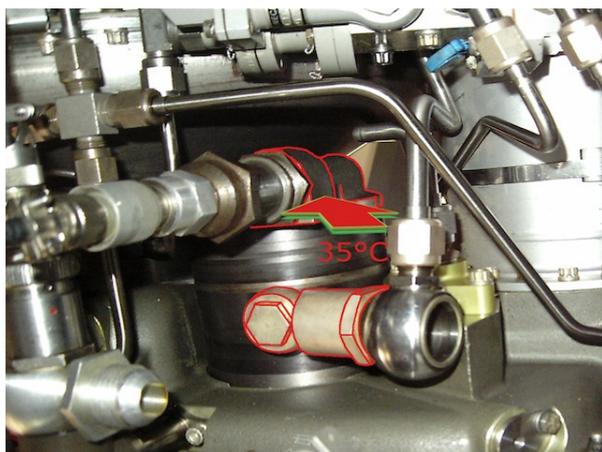


FIGURE 47: Vue augmentée de l'opérateur après une interaction de type « Outlining ». L'expert a souhaité mettre en avant l'objet de la discussion (pièces détournées en rouge) tout en laissant apparaître une indication pouvant se révéler importante pour la suite des opérations.

Adding

Calqué sur le principe des gestes kinétiques, l'Adding substitue ce geste en utilisant des animations 2D/3D adéquates. Complètement orienté vers la tâche, l'Adding est alors une gradation de l'Outlining à laquelle on combine des animations 2D ou 3D. En fonction de la situation, d'un mouvement particulier à accomplir ou de l'avancement de la tâche, l'expert peut sélectionner l'animation désirée et pointer l'élément auquel elle se rapporte dans le flux vidéo. L'opérateur visualisera, à travers sa vision orthoscopique, l'action à accomplir (Fig.48). Comme pour l'Outlining, l'Adding nécessite au préalable de posséder un catalogue d'animations se rapportant directement au système à maintenir.

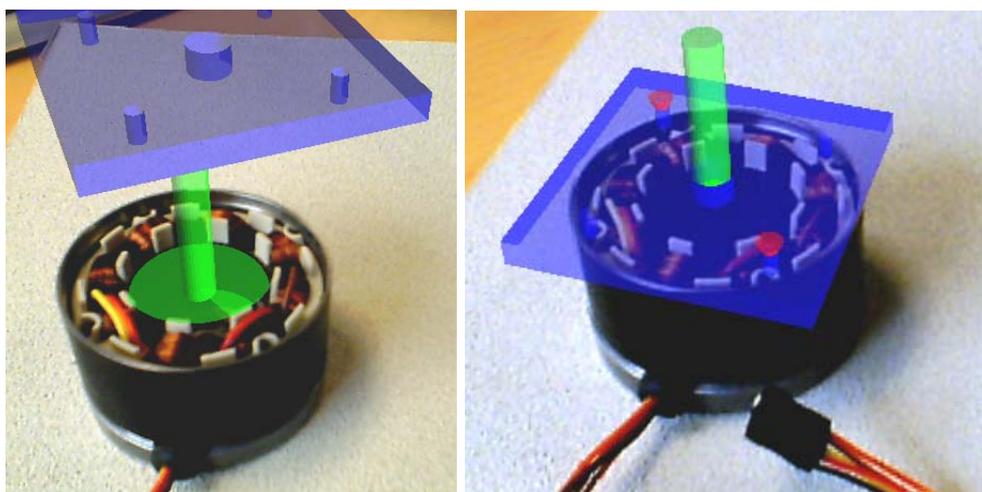


FIGURE 48: Vue augmentée de l'opérateur après une interaction de type « Adding ». L'expert montre, par une animation virtuelle 3D recalée sur l'élément réel, l'enchaînement des opérations.

11 CONCLUSION

Au cours de ce chapitre, nous avons voulu justifier l'intérêt qu'il peut y avoir à requérir une aide humaine au travers des NTIC. Nous avons exposé les faits qui argumentent l'approche humaine en termes d'expertise et de collaboration.

Nous avons alors orienté notre réflexion sur la façon dont on pourrait soutenir une assistance à distance, permettant à l'expert d'interagir de façon aussi efficace et naturelle qu'en situation de coprésence, tout en profitant des possibilités offertes par la réalité augmentée dans le domaine de la maintenance. En ce sens, les propositions que nous avons faites se veulent fédératrices du meilleur des principes des systèmes « Automatisés » et « Supervisés ». Pour cela, nous avons montré les mécanismes importants qui interviennent lors d'une communication interpersonnelle lorsque des individus tentent de résoudre ensemble un problème posé pendant une intervention de maintenance. Intégrer ces mécanismes nous a permis de proposer pour l'expert une façon de se projeter dans l'espace distant de la tâche et d'y « interagir » grâce à notre paradigme P.O.A., et pour l'opérateur une façon de percevoir les indications de l'expert.

Le chapitre suivant a pour objectif de décrire l'implémentation en termes matériel et logiciel des propositions que nous venons de faire. Nous y décrirons également les expérimentations que nous avons menées afin de tester leur viabilité.

Quatrième partie

SYSTÈME T.A.C : TÉLÉ-ASSISTANCE COLLABORATIVE (Conception et Expérimentation)

En regard de la téléassistance, la collaboration doit être considérée d'un point de vue du savoir. L'approche développée par [NT95] montre que la gestion du savoir ne peut être réduite à la gestion d'informations, mais doit plutôt considérer le rôle des protagonistes. Ils sont engagés dans un processus établissant une succession de tâche où ils doivent interagir mutuellement : à chaque étape ils doivent percevoir l'environnement de la tâche, l'analyser et décider quoi faire, le faire, suivant ainsi un cycle de perception, décision et action [Gib77].

L'expert distant doit alors être capable de voir ce que voit l'opérateur pour percevoir correctement la situation. Il doit analyser cette situation et prendre une décision basée sur ses connaissances et son expérience. Bien qu'il ne puisse agir physiquement, il doit être en mesure de transférer à l'opérateur sa compréhension afin que celui-ci effectue l'action appropriée. L'expert a donc besoin d'outils interactifs performants pour compléter l'information audio, comme des possibilités multimodales (visuelles et gestuelles) offertes en situation de coprésence, et ce en temps réel. Grâce à cela, il pourra « transférer » son savoir tacite en savoir explicite à l'opérateur.

C'est dans cette optique que nous avons conçu notre propre système d'aide nommé T.A.C., acronyme de Télé-Assistance Collaborative. Celui-ci est l'implémentation du principe général de fonctionnement d'un système d'assistance à distance que nous avons énoncé au chapitre précédent, tout en y intégrant nos propositions de visualisation et d'interaction.

Nous allons dans ce chapitre présenter le système T.A.C. dans son fonctionnement interne en justifiant nos choix de conception et d'implémentation. Qui plus est, durant nos travaux, nous avons été en mesure d'expérimenter le principe des références ostensives lors de tâches de maintenance. Nous en présenterons les premiers résultats.

12 LE SYSTEME T.A.C.

Les systèmes d'aide intègrent des fonctionnalités toujours plus riches, nombreuses et complexes. Pour que les utilisateurs puissent pleinement en profiter, leur niveau d'utilisabilité est un élément essentiel. L'utilisabilité correspond à « l'adaptation de la technologie aux caractéristiques de l'utilisateur » [BB03] et est définie par la conjonction de :

- L'efficacité : produire l'effet attendu.
- L'efficience : produire une tâche donnée avec un minimum d'efforts.
- La satisfaction : niveau de confort ressenti en utilisant le système d'aide.

C'est guidé par ces recommandations que nous avons conçu le système T.A.C. Nous allons tout d'abord rappeler son principe de fonctionnement avant de présenter son diagramme de conception général.

12.1 Rappel de fonctionnement

Pour proposer une solution permettant d'allier collaboration à distance et maintenance grâce à la réalité augmentée, nous avons retenu deux aspects fondamentaux :

- Le mode d'interaction entre les interlocuteurs : c'est la façon qu'aura l'expert de pouvoir « simuler » une gestuelle.
- L'espace visuel partagé : il s'agit de pouvoir rendre compte à l'expert de l'environnement de l'opérateur et la façon dont l'opérateur peut visualiser les informations promulguées par l'expert.

De par ces aspects, nous sous-entendons également que notre système puisse supporter une collaboration synchrone entre les intervenants. Nous avons alors proposé le fonctionnement d'un tel système (Fig.49) :

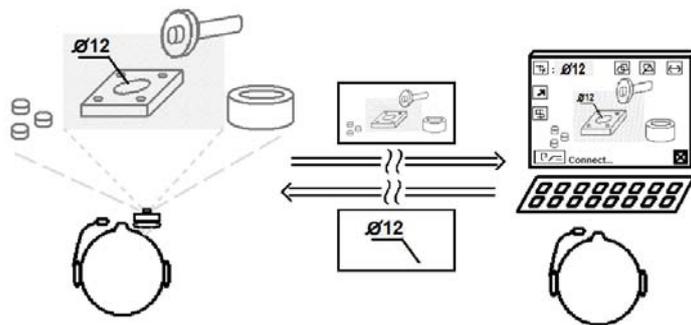


FIGURE 49: Principe de fonctionnement de T.A.C.

L'opérateur est équipé d'un dispositif particulier de visualisation. Sa conception particulière (que nous allons décrire dans ce chapitre) lui permet de capturer un flux vidéo de ce que voit exactement l'oeil du porteur (flux orthoscopique ou flux O) et un flux vidéo grand-angle (flux périphérique ou flux P). Parmi les deux flux que recevra l'expert, il aura la possibilité d'augmenter le flux O grâce à notre paradigme P.O.A. Les augmentations seront alors envoyées en temps réel sur l'afficheur de l'opérateur.

12.2 Diagramme de déploiement

Le diagramme de déploiement de la figure 50 représente une vue statique de l'infrastructure physique des ressources matérielles qui composent T.A.C., la manière dont sont répartis ces composants ainsi que leurs relations⁴⁷.

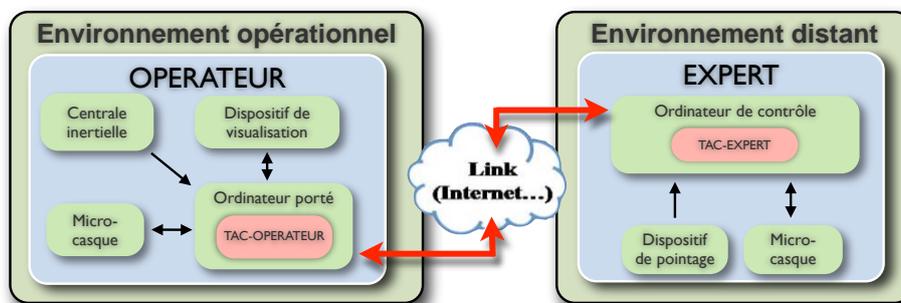


FIGURE 50: Design général de T.A.C.

Le système T.A.C. se découpe en deux parties distinctes, l'une relative à l'opérateur, l'autre relative à l'expert.

Côté opérateur, celui-ci sera équipé des éléments suivants :

- De lunettes de visualisation capables de présenter les augmentations en provenance de l'expert. La seconde fonctionnalité de ces lunettes étant de retranscrire l'environnement de l'opérateur pour l'expert.
- D'une centrale inertielle en charge d'évaluer les mouvements en vue de gérer les augmentations.
- D'un système de calcul porté implémentant l'application supportant P.O.A. (nommée TAC-Opérateur) et gérant la communication distante.
- D'un micro-casque pour les communications audio.

Côté expert, nous retrouvons :

- Un micro-casque pour les communications audio.
- Un ordinateur implémentant l'application supportant P.O.A. (nommée TAC-Expert) et gérant la communication distante.

Nous avons fait en sorte que l'opérateur soit le moins instrumenté possible pour ne pas perdre de temps à s'équiper et être gêné dans son activité. Le fonctionnement de T.A.C. sera le plus transparent possible, ne nécessitant aucune manipulation de la part de l'opérateur durant l'intervention, à l'exception de la phase d'initialisation qui sera simplifiée au maximum (connexion avec l'expert).

Pour ce qui est de l'expert, celui-ci n'a que peu de contraintes en termes d'équipement, un ordinateur de bureau ou un ordinateur portable (équivalent en termes de puissance) étant suffisants. Nous recommandons tout de même un système de pointage type souris plutôt que trackpad pour une meilleure précision et une plus grande rapidité.

47. Nous précisons que nous n'avons pas utilisé le formalisme UML pour cette représentation.

12.3 Choix d'implémentation

Il existe dans le domaine de la réalité augmentée de nombreuses solutions architecturales qui ont été développées avec l'objectif de rendre le prototypage d'application rapide et flexible. Cela a permis dans un premier temps de proposer une couche d'abstraction facilitant l'intégration entre les diverses technologies relatives à ce domaine. Ces architectures sont principalement des architectures modulaires. Elles sont donc composées de modules qui, suivant le paradigme de conception et de programmation utilisé, sont appelés *composant*, *service* ou encore *objet*. Cela permet une réutilisation aisée du code déjà écrit.

Parmi les plus connues et les plus représentatives, nous pouvons citer COTERIE de l'université de Columbia [MF96], StudierStude de l'université de Vienne [SFH⁺02], TINMITH de l'université d'Australie du Sud [PT03], DWARF de l'université de Munich pour les environnements distribués [BBK⁺01] ou encore DART, initié par l'université de Columbia [MGDB05] en remplacement de COTERIE. Il en existe de nombreuses autres et toutes ont leurs avantages et leurs défauts. Dans [Did05], Didier en dresse un tableau comparatif montrant que le choix de l'une ou l'autre de ces architectures ne se révèle pas forcément adapté au prototypage d'un système expérimental.

En ce qui nous concerne, nous avons fait le choix d'une programmation orientée composants. Un composant est un module cohérent d'un programme qui peut être développé séparément et distribué en tant qu'unité indépendante. Il offre une interface par laquelle il peut être connecté, tout en restant inchangé, avec d'autres composants pour former un système de taille plus importante [DW99]. Il s'agit donc d'une boîte noire, renfermant du code compilé, devant être capable de fournir une interface, lui permettant ainsi d'être sujet à une composition avec d'autres composants. Opter pour une programmation orientée composants nous est alors apparu comme une solution permettant une grande modularité dans le code, adaptée à la conception d'un prototype expérimental.

Pour ce faire, nous avons porté notre choix de développement de composants en nous appuyant sur la bibliothèque multiplateforme Qt de Nokia⁴⁸. Développée en C++, elle offre des composants d'interfaces graphiques, d'accès aux données, de connexion réseau... Elle offre également un puissant moyen de communication entre les composants à travers le mécanisme signal-slot, implémentation du patron de conception dit observateur⁴⁹. L'intégration qui a été faite dans la bibliothèque Qt de ce mécanisme possède à notre sens deux atouts majeurs : pouvoir connecter et déconnecter des composants de manière dynamique, mais surtout la réception d'un signal de traitement par un composant implique une exécution immédiate du processus associé.

Enfin, de manière plus classique, nous avons opté pour un développement en C/C++, le langage C ayant été utilisé pour l'écriture d'algorithmes dont la rapidité d'exécution est critique (traitement d'images...). Pour ce qui est du traitement 3D, nous avons choisi la bibliothèque OpenGL, qui en plus d'être aussi performante que la bibliothèque Direct3D, a l'avantage d'être multiplateforme et facilement interfaçable avec la bibliothèque Qt.

48. Notons qu'il existe d'autres systèmes de programmation par composants comme CORBA, .COM, XPCOM ou encore JavaBeans.

49. « pattern observer » est le terme plus communément utilisé dans la communauté informatique.

12.4 Vue d'ensemble de l'architecture interne

Les programmes TAC-opérateur et TAC-expert sont donc le coeur de traitement du système T.A.C. Ces deux programmes sont chargés de gérer les niveaux de perception relatifs aux protagonistes ainsi que les augmentations qui y seront associées. Il s'agit par conséquent de l'application de réalité augmentée en elle-même, divisée en deux modules. Nous allons préalablement exposer une vue générale de l'architecture de l'application avant de décrire le rôle de chacun des composants constituant. Certains d'entre eux, présentant une conception particulière, verront leurs fonctionnalités détaillées à la section 14 (p. 114).

12.4.1 Diagramme de relation

Ce diagramme permet de mettre en évidence les dépendances et les relations entre les différents éléments constituant l'application T.A.C.

Basés composants, les modules TAC-opérateur et TAC-expert sont architecturés comme suit (Fig.51) :

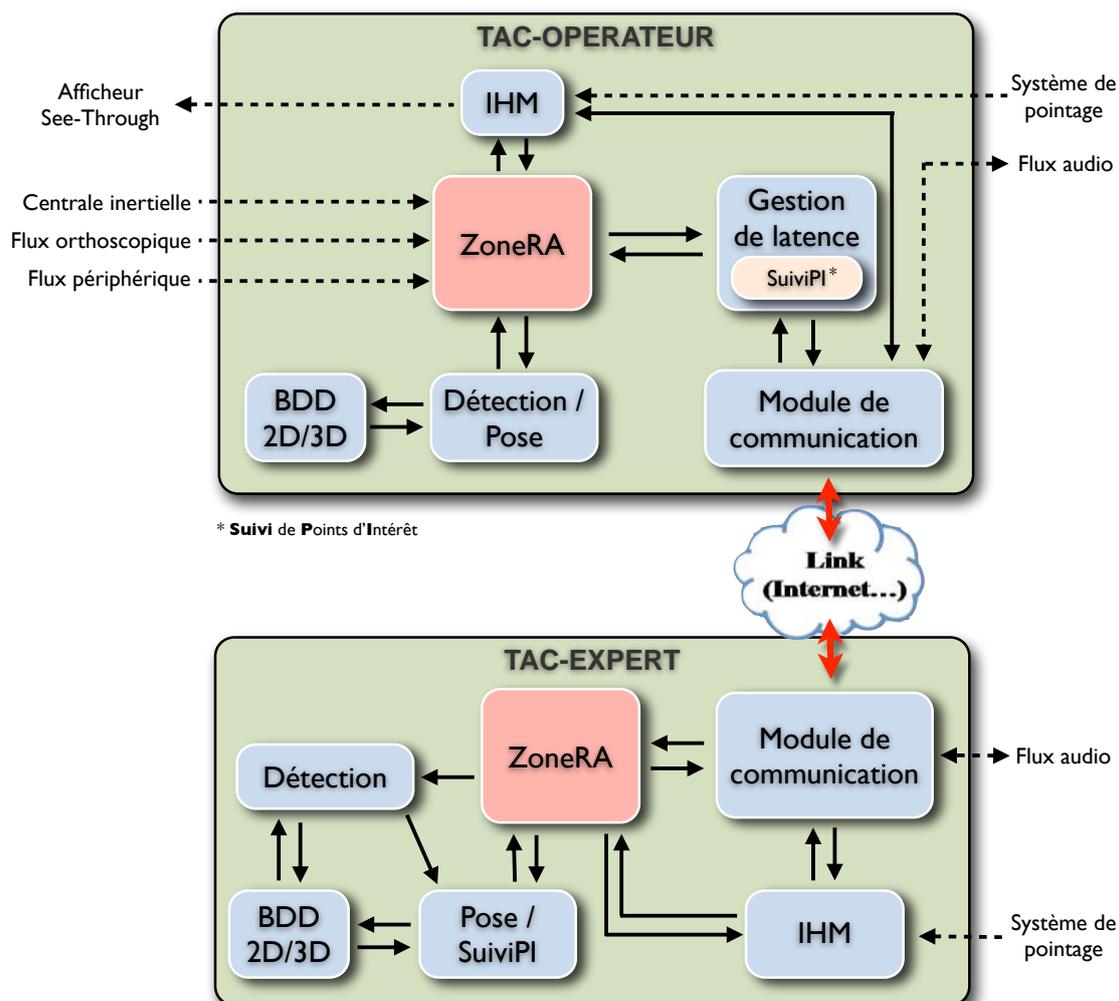


FIGURE 51: Vue d'ensemble de l'architecture basée composants du système T.A.C.

12.4.2 Description du rôle des composants de TAC-Opérateur

IHM Il s'agit du composant chargé d'afficher l'image orthoscopique sur le système de visualisation See-Through (voir la section 13 page 103 pour les détails de la technologie mise en oeuvre). Grâce à un système de pointage (souris ou autre), l'opérateur devra initialement régler de façon précise l'orthoscopie, et ce, de manière à s'affranchir des problèmes de parallaxe entre l'oeil et le dispositif See Through. Ces paramètres seront alors transmis au composant chargé d'en appliquer les modifications pour chaque image du flux vidéo. Toujours grâce au système de pointage, l'opérateur pourra choisir de se connecter au service d'assistance désiré parmi plusieurs préalablement enregistrés. Si la connexion avec l'expert est interrompue, l'IHM préviendra l'opérateur en affichant le message adéquat.

ZoneRA Ce composant est chargé de collecter et synchroniser les flux vidéo (orthoscopique et périphérique) puis les informations provenant de la centrale inertielle. Chaque image du flux orthoscopique se verra ainsi associer les valeurs inertielles en correspondance avec la dynamique de l'utilisateur au moment où cette même image est prise. Ceci fait, ces images sont transmises au composant « Gestion de latence », de même que celle du flux périphérique. Pour inhiber les problèmes de parallaxe ajustés grâce au composant IHM, ZoneRA se charge également d'appliquer les transformations géométriques sur les images provenant du flux orthoscopique. Enfin, lorsqu'une demande d'augmentation est émise par l'expert, la nature de celle-ci est analysée par ZoneRA pour y associer le traitement adéquat (nous ne traitons pas les augmentations de type « Picking » de la même manière que celle de type « Outlining-Adding », leur nature différant en termes de procédé algorithmique, voir section 14 page 114). Le flux orthoscopique étant ainsi augmenté, il est transmis vers le composant IHM en vue d'être affiché.

Détection/Pose Constitué de deux sous-composants « Pose » et « Détection », ce composant est utilisé lors d'un « Outlining » ou d'un « Adding ». Dès qu'une augmentation de type 3D en provenance de l'expert est demandée, « Détection » détectera dans l'image en cours l'élément auquel il est fait référence. S'en suivra un calcul par « Pose » afin de déterminer la position-orientation de cet élément par rapport au point de vue de l'opérateur. Nous détaillerons les techniques employées pour ce faire au paragraphe 14.3 (p. 118).

BDD 2D/3D Ce composant intègre une base de données des éléments virtuels susceptibles d'être utilisés en tant qu'augmentation. Ces éléments peuvent être aussi bien des éléments 2D, notamment en ce qui concerne le principe du « Picking », comme des flèches par exemple, ou peuvent être des éléments et animations 3D pour le principe de « l'Outlining » ou de « l'Adding ». Pour ce qui est des éléments 3D, la base de données référence tant leur modèle géométrique que leur modèle d'indexation (ensemble de descripteurs caractérisant l'élément pour effectuer une analyse basée vision, voir le paragraphe 14.3 page 118).

Gestion de latence Ce composant traite les problèmes relatifs à l'introduction du délai engendré par la transmission distante lorsque des augmentations en provenance de l'expert devront s'afficher dans le champ visuel de l'opérateur. L'objectif est d'en gérer l'affichage en tenant compte aussi bien des mouvements de l'opérateur que du temps de transmission des données entre les protagonistes. Pour cela, avant d'être transmise à l'expert, chaque image du flux

vidéo orthoscopique est horodatée (timestamp), puis sauvegardée dans la mémoire tampon (buffer) du composant avec les valeurs inertielles associées. Le retour d'information en provenance de l'expert pourra être ainsi traité par un sous-composant (SuiviPI) en fonction de ces paramètres et du type d'augmentation utilisé (cet aspect sera détaillé au paragraphe 14.2 page 114).

Module de communication Il s'agit du composant gérant la communication audio-vidéo distante proprement dite entre l'ordinateur de l'expert et l'ordinateur de l'opérateur sur un réseau IP. Pour cela, nous avons tiré parti de la bibliothèque OpenSource Live555 spécialisée dans le streaming multimédia. Nous en utilisons le protocole SIP⁵⁰ pour établir une session audio-vidéo, cette session étant décrite via le protocole SPD⁵¹. Le transfert de données quant à lui est assuré par le protocole RTP⁵². Dans notre cas, la connexion vidéo sera unilatérale, de l'opérateur vers l'expert, la connexion audio, elle, étant bien évidemment bilatérale. Le composant « module de communication » aura également pour rôle de recevoir les ordres d'augmentation de l'expert ainsi que les paramètres de réglage du débit des flux vidéo.

12.4.3 Description du rôle des composants de TAC-Expert

Note : Bien que l'on retrouve dans le module TAC-expert des composants ayant le même nom que dans le module TAC-opérateur, certains n'ont pas un fonctionnement identique, soit parce qu'ils sont en interne assemblés entre eux de manière différente, soit parce qu'ils ont un comportement adapté au module dans lequel ils se trouvent.

IHM Il s'agit de l'interface graphique à travers laquelle l'expert va visualiser et interagir avec l'opérateur (voir paragraphe 14.4 page 119). L'interface prend la forme d'une fenêtre graphique classique affichant la visualisation PiP, l'expert ayant la possibilité de cliquer avec la souris sur le flux orthoscopique. C'est également ce composant qui affichera les informations relatives à la qualité de la connexion. Ainsi, l'expert aura la possibilité de dégrader le débit d'envoi des flux vidéo provenant de l'opérateur si la situation l'exige.

Détection Les fonctionnalités de ce composant seront appelées lorsque l'expert aura cliqué sur le flux vidéo orthoscopique en vue de l'augmenter. Suivant le type d'augmentation désiré, l'algorithme de détection basé image adéquat sera utilisé. S'il s'agit d'une augmentation 3D, c'est-à-dire de type « Outlining-Adding », un algorithme détectera dans l'image en cours l'élément auquel il est fait référence, et ce, en relation avec les données fournies par le composant « BDD 2D/3D ». Si l'augmentation est de type « Picking », alors des points caractéristiques liés à l'image et proches des coordonnées cliquées seront recherchés. Une fois détectées, ces informations sont envoyées au composant « Pose/SuiviPI » (voir la section 14 page 114 pour plus de détails).

50. « Session Initial Protocol », implémenté dans la bibliothèque SophiaSIP.

51. « Session Description Protocol ».

52. « Real Time Protocol ».

Pose/SuiviPI Ce composant est constitué de deux sous-composants, à savoir « Pose » et « SuiviPI ». Le sous-composant « Pose » a pour objectif de calculer la position-orientation de l'élément préalablement détecté par le composant « Détection » dans le cadre d'une augmentation 3D. Pour ce qui est du sous-composant « SuiviPI⁵³ » intervenant pendant une augmentation de type « Picking », il sera chargé de suivre dans l'image en cours les points d'intérêt détectés autour de la zone de clic. La position-orientation et/ou les points d'intérêts suivis seront en suite transmis au composant « ZoneRA ».

ZoneRA C'est ce composant qui recevra les flux vidéo en provenance de l'opérateur. Il procédera alors à l'incrustation du flux orthoscopique dans le flux périphérique avant de le transmettre au composant IHM pour affichage. Il recevra également, en provenance du composant IHM, le type d'augmentation désiré ainsi que la zone de clic. Il commandera alors au module « Détection », soit de détecter l'élément à reconnaître dans l'image en cours, soit de trouver les points d'intérêt autour de la zone de clic. En retour, il augmentera le flux orthoscopique en vue de l'afficher sur l'interface de l'expert. En parallèle, ce composant transmettra au composant « module de communication » les informations relatives à l'augmentation (points d'intérêt détectés et/ou position-orientation + type d'augmentation).

Module de communication En ce qui concerne la communication audio-vidéo, ce composant possède les mêmes spécificités que son homologue de TAC-opérateur. De plus, sa tâche sera également de transmettre les ordres et caractéristiques des augmentations à afficher côté opérateur. Enfin, il transmettra au module distant les débits à appliquer pour l'envoi des flux si tel doit être le cas.

BDD 2D/3D Le fonctionnement de ce composant est identique à celui de TAC-opérateur. Se référer ci-dessus pour la description de son rôle.

Ainsi, nous avons, dans cette section 12, présenté de manière globale la composition logicielle interne de T.A.C. puis l'infrastructure matérielle nécessaire à son fonctionnement. Dans les sections 13 et 14 nous allons décrire de manière plus concise les éléments clés du système T.A.C., à savoir côté opérateur le dispositif chargé de la visualisation des augmentations et de retranscrire l'environnement de celui-ci, puis côté expert la façon dont est implémenté P.O.A. dans l'application.

53. « SuiviPI » pour **Suivi** de **Points d'Intérêt**.

13 PERCEPTION : LES LUNETTES MOVST

Pour chaque système de réalité augmentée développé, il convient de choisir son type d’affichage de façon spécifique. En ce sens, nous avons vu au premier chapitre les recommandations faites par [KC98] en fonction du domaine d’application. Mais bien entendu, les critères jugés acceptables ne peuvent tenir compte de toutes les contraintes qu’impose le travail d’un opérateur dans le cadre de la maintenance.

Qui plus est, le marché des lunettes de réalité augmentée est aujourd’hui un marché encore naissant, les grands constructeurs comme Sony ou Ucuiti s’étant désintéressés de ce segment pour cause de non-rentabilité. C’est pour cela que durant la décennie écoulée, nombres de laboratoires et chercheurs ont développé leur propre système d’affichage, souvent basé sur d’anciens modèles du marché (ex. : [AMR02], [PHMG06] ou encore [SKF05]). La situation est analogue lorsque les systèmes existants ne proposent pas les fonctionnalités attendues, comme la gestion des occultations pour les systèmes optiques [CHR04] ou encore la perte de la vision stéréoscopique lors d’un travail en vision très proche [SAH⁺01].

Il en est de même en ce qui nous concerne. À la vue des propositions de perception, tant pour l’opérateur que pour en rendre compte à l’expert, nous avons dû concevoir notre propre système de visualisation. Avant d’en présenter les caractéristiques et l’aspect, nous allons décrire les capacités de la perception visuelle humaine pour comprendre les choix qui ont été effectués.

13.1 La perception visuelle humaine

Le système visuel est certainement le sens humain qui est le plus utilisé pour percevoir l’environnement spatio-temporel. Celui-ci peut se découper en deux parties distinctes, à savoir les yeux et le système visuel cérébral (Fig.52) :

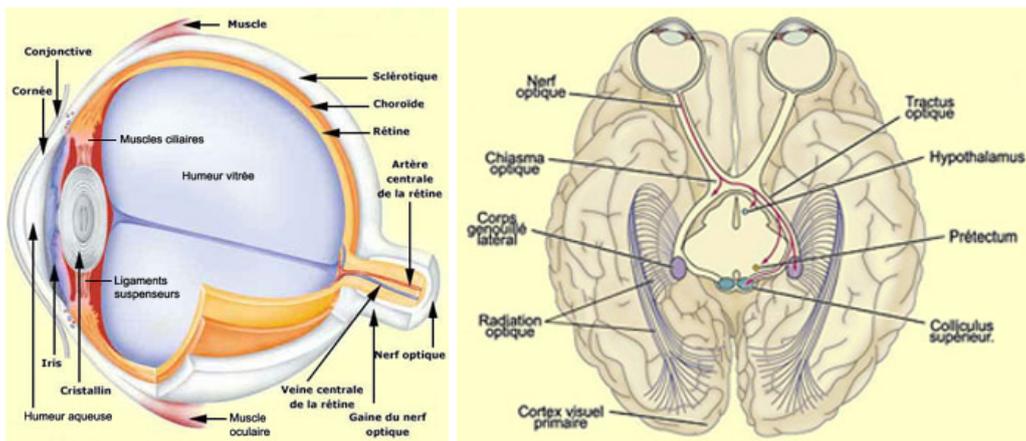


FIGURE 52: Détails de l’œil et du cortex visuel cérébral.

D’un point de vue fonctionnel, l’œil est le capteur de la perception. Son rôle est de concentrer sur la rétine une image nette et bien contrastée grâce aux actions conjuguées de la cornée, du cristallin et de l’iris. La rétine est le capteur de lumière à proprement parler, constituée de cônes (vision photonique) et de bâtonnets (vision scotopique).

Les cônes sont des cellules photoréceptrices, localisées en grande partie au centre de la rétine, réagissant aux couleurs du spectre visible par l'oeil (Fig.53). Il existe trois types de cône permettant de différencier les couleurs : les cônes sensibles aux longueurs d'onde d'environ 580 nm (rouge), ceux sensibles à 540 nm (verts) et les derniers à 450 nm (bleu).

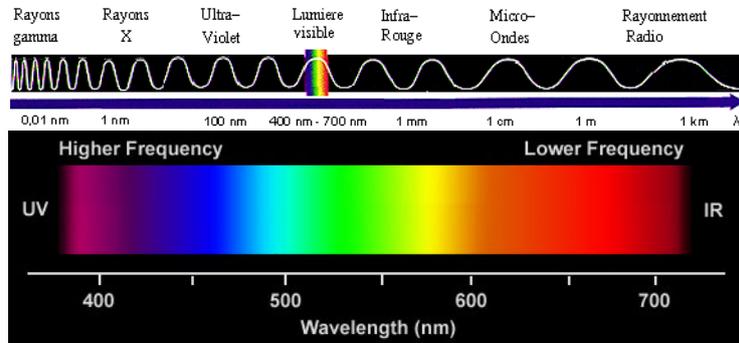


FIGURE 53: Spectre électromagnétique (haut) et spectre visible par l'oeil humain (bas).

Quant aux bâtonnets, plutôt situés en périphérie de la rétine, ils interviennent lorsque la luminosité est faible, leur sensibilité étant maximale pour une longueur d'onde de 500 nm. En vision périphérique, ils interviennent également dans la détection de mouvement.

Le cortex visuel cérébral lui, reçoit les informations en provenance des yeux via le nerf optique. Ces informations visuelles, traduites en influx nerveux, seront propagées jusqu'aux différentes structures cérébrales responsables de l'analyse des signaux. C'est dans ces régions que seront « interprétées » ces informations en vue de recréer une image de la réalité. Dans les paragraphes suivants, nous avons choisi de présenter de façon succincte les fonctionnalités clés de la vision humaine. L'objectif étant de mettre en avant les capacités d'*acquisition* d'une scène et l'*interprétation* qui en est faite par le cerveau.

13.1.1 La luminance

La luminance est l'intensité lumineuse⁵⁴ d'une source en provenance d'une direction donnée, divisée par la surface apparente de cette source. Ainsi, pour une source lumineuse quasi ponctuelle, la luminance et l'intensité lumineuse définissent la même notion. Avec une source lumineuse ayant une certaine aire, pour que la luminance continue à représenter l'intensité lumineuse d'un point, on doit diviser l'intensité lumineuse totale par l'aire apparente.

Dans une scène naturelle, la variation de la luminance (Tab.11) est due aux variations d'éclairement et aux variations de réflectance⁵⁵. Mais à intensité lumineuse égale, la perception de la luminance n'est pas constante, et ce, suivant la longueur d'onde. En effet, c'est pour une longueur d'onde de 555 nm (couleur jaune-vert) que la luminance semble atteindre son maximum (Fig.54).

⁵⁴. L'intensité lumineuse est une mesure de l'éclat d'une source lumineuse perçue par l'oeil. L'unité SI est le candela.

⁵⁵. La réflectance est le rapport entre l'énergie lumineuse réfléchie et l'énergie lumineuse incidente sur une surface. Une réflectance de 100 % implique une réflexion totale de l'énergie.

Conditions lumineuses de la vie courante	Eclairement lumineux (lux)	Conditions de travail	Eclairement lumineux (lux) recommandé par les autorités européennes
Plein Soleil	50 000 à 100 000	Tâche de précision	600
Ciel couvert	20 000 à 30 000		
Eclairage de bureau	200 à 500	Mécanique moyenne	200
Nuit de pleine Lune	0.2 à 0.5		
Limite pour l'appréciation des formes	0.02 à 0.003		

TABLE 11: Quelques gammes d'éclairement.

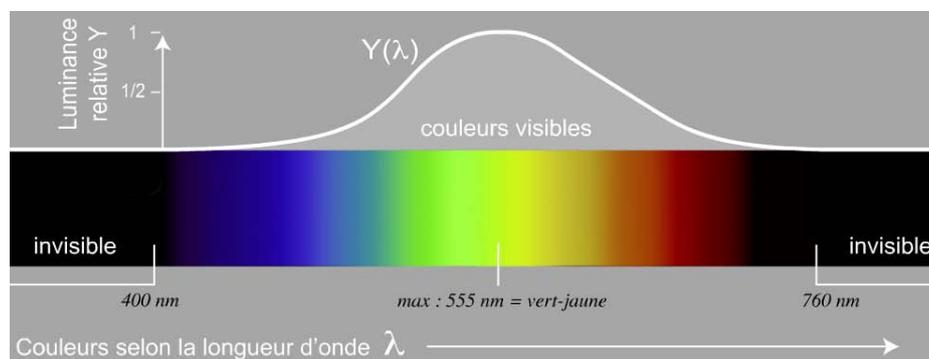


FIGURE 54: Luminance relative en fonction de la longueur d'onde visible par l'oeil humain (à intensité lumineuse égale).

Vu la diversité d'éclairement, de contraste et de couleur présente dans une scène naturelle, l'oeil s'adapte à la *luminance moyenne* par le biais de deux processus :

1. En faisant varier le diamètre de la pupille.
2. En faisant varier la sensibilité moyenne de la rétine (Effet Stiles-Crowford). Cette variation est faite de manière progressive⁵⁶.

Mais cette adaptation à la luminance moyenne a des limites, surtout si les deux yeux ne perçoivent pas la même. Dans ce cas, les luminances moyennes en provenance de chaque oeil peuvent provoquer des problèmes de perception. [HV96] a d'ailleurs mis en avant ce phénomène. En présentant à des sujets différents verres de différentes teintes devant un seul oeil, les résultats ont montré que la perception est fortement dégradée pour une opacité supérieure à 50 %.

13.1.2 Le champ visuel

Le champ visuel est l'espace délimité par la perception de l'oeil, sans bouger la tête. Il est défini suivant un champ de portée horizontale et un champ de portée verticale (Fig.55). En ce qui concerne

⁵⁶. Par exemple, lors d'une variation brusque de l'éclairage moyen, comme lorsque l'on passe d'une pièce sombre vers l'extérieur ensoleillé.

le champ de portée vertical, il couvre un angle d'environ 130° , limité vers le haut par les arcades sourcilières et vers le bas par les joues. Pour ce qui est du champ horizontal, il est en moyenne de 180° pour les deux yeux et de 150° pour chaque oeil (limité par le nez).

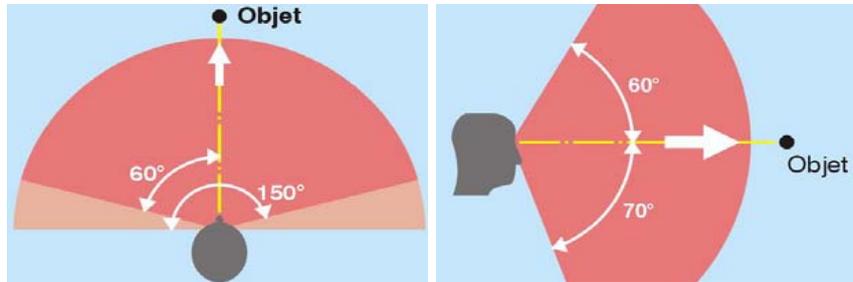
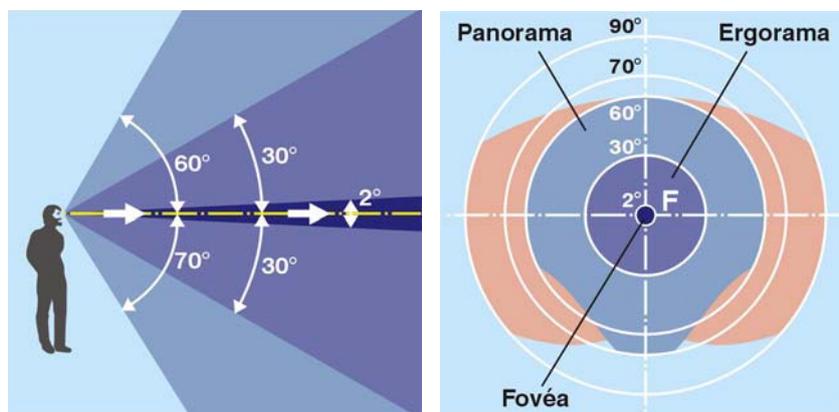


FIGURE 55: Champ de vue horizontal et champ de vue vertical.

La capacité des yeux à saisir l'information visuelle dépend de leur position relative dans le champ visuel global. On dénombre trois zones en tout (Fig.56) :

- La Fovéa. C'est le champ central de la vision, avec un domaine d'ouverture extrêmement réduit, de l'ordre de 2° à 3° environ. L'acuité visuelle dans la zone fovéale est extrêmement élevée, ce qui la prédestine à être utilisée pour observer des détails très fins. En dehors de cette zone, l'acuité est jugée médiocre. Le regard est donc en perpétuel mouvement pour diriger l'axe fovéal sur une zone précise, reconstituant ainsi la sensation globale d'une image nette.
- L'Ergoma. Domaine de $2 \times 30^\circ$ d'ouverture, cette zone proche de la fovéa constitue une zone de surveillance qui permettrait à l'observateur, de manière consciente, d'orienter son regard en vue d'une analyse fine. De façon plus générale, cette zone est destinée à percevoir les formes dans leur ensemble.
- Le Panorama est la zone d'environ $2 \times 60^\circ$ d'ouverture, renseignant sur les mouvements relatifs des grandes masses structurant l'image. Une détection de mouvement dans cette zone peut induire une orientation volontaire du regard par un déplacement conjugué de la tête et des yeux.



(a) Coupe du champ visuel de profil.

(b) Champ visuel en coupe frontale perçu par les deux yeux (bleu) et par chaque oeil (rose).

FIGURE 56: Zones d'action du champ visuel.

13.1.3 La vision stéréoscopique

À l'endroit où le champ visuel des yeux se recouvre, l'Homme a une vision binoculaire. C'est très certainement l'une des capacités les plus importantes du système visuel. Grâce à la superposition des champs dans la zone médiane, un même objet est vu simultanément, mais sous deux angles différents⁵⁷. C'est le cortex visuel qui sera en charge de « recréer » l'impression de relief. Cependant, il est important de faire remarquer qu'une vision en relief est possible avec un seul oeil. Nous allons voir les facteurs responsables de cette vision en relief, basée soit sur des facteurs monoculaires, soit binoculaires [Gao92].

En vision monoculaire, il est parfaitement possible de reconstruire l'espace tridimensionnel. Pour cela, nous utilisons de nombreuses informations :

- L'accommodation. Capacité du cristallin de l'oeil à changer de forme pour permettre une mise au point de l'image sur la rétine (en particulier sur la fovéa). C'est également un facteur important dans l'appréciation des distances.
- La parallaxe monoculaire. Relation au déplacement en translation ou en rotation d'un objet devant l'oeil de l'individu. Grâce à ce mouvement, l'oeil perçoit à deux instants très proches deux images différentes d'un même objet, le cortex visuel reconstituant ainsi la notion de profondeur.
- Les perspectives géométriques sont ce que l'on pourrait définir des facteurs psychologiques, dans le sens où il est fait appel à l'expérience de l'individu pour évaluer l'environnement. On peut ainsi obtenir une sensation de fortes profondeurs à partir d'images bidimensionnelles. Ces effets ont été classés en six principales catégories :
 1. La superposition des contours. C'est le fait d'éliminer les parties cachées par des objets plus proches.
 2. La taille de l'objet est également une information importante qui permet d'évaluer les distances.
 3. Les perspectives linéaires, c'est-à-dire la convergence des lignes vers l'infini, sont également des éléments importants dans l'évaluation de la profondeur d'une scène.
 4. La perception des textures. C'est la réduction progressive d'une texture dans une certaine direction (ex. : réduction des fenêtres d'un immeuble vers l'infini).
 5. L'ombrage et ses effets sur l'illumination des objets environnants nous renseignent sur l'agencement des différents objets entre eux.
 6. La perspective de surface. C'est le jeu des atténuations de contrastes pour des objets lointains. Un objet lointain est souvent moins net qu'un objet rapproché.

⁵⁷. C'est la distance entre les deux yeux qui est responsable du décalage du point de vue. Cette distance, appelée distance interne pupillaire (IPD en anglais) est généralement comprise entre 6 et 7 cm.

En vision binoculaire, deux facteurs importants interviennent dans la perception spatiale, d'une part la convergence binoculaire, d'autre part la disparité binoculaire :

- La convergence binoculaire est le mouvement permettant aux yeux de s'orienter vers la direction de l'objet observé. De par l'éloignement séparant l'objet des yeux, ceux-ci doivent déterminer une accommodation correcte pour la distance à laquelle les axes visuels convergent. Accommodations et convergences sont donc étroitement liées [SO58].
- La disparité binoculaire est le décalage horizontal⁵⁸ sur les deux images rétiniennes d'un même objet. C'est grâce en partie à ce décalage de point de vue que la vision tridimensionnelle est possible⁵⁹. Plus l'objet est près, plus la « sensation de profondeur » sera présente. En revanche, pour les objets très éloignés, les images rétiniennes sont quasi identiques, la perception tridimensionnelle étant alors quasi nulle.

La simple vision monoculaire peut donc donner une sensation de relief. Mais pour obtenir une perception fine du relief, la vision stéréoscopique est indispensable (condition binoculaire).

13.1.4 Interprétation de l'information

Comme nous allons le voir, les informations en provenance des yeux vont permettre au cortex visuel et au cerveau d'analyser les éléments importants en vue de recréer la scène observée.

En effet, nous avons accès à la réalité à travers nos sens. Mais ce que nous « ressentons » ne correspond pas forcément au réel, mais plutôt à une représentation mentale de ce que nous pensons être le réel. Ainsi, pour la vision, les images que nous voyons ne nous sont pas données comme une photographie, mais sont traitées et interprétées par notre système visuel et notre cerveau.

Ce phénomène prend naissance dès la rétine, avec une réduction de la quantité d'informations transmises au nerf optique. À ce stade, les informations des 125 millions de photo-récepteurs⁶⁰ convergent vers dix fois moins de cellules ganglionnaires⁶¹. Le cortex visuel et le cerveau se chargent alors de compenser les informations perdues en complétant des éléments de la réalité avec des informations abstraites, dans le but de recréer une image cohérente. Notre cerveau interprète donc en permanence les informations en provenance des yeux pour construire une image qui a du sens pour nous. Cette interprétation est tellement performante que nous avons une impression de cohérence là où il n'y en a probablement pas. Les illusions d'optique en sont la manifestation⁶².

Ces illusions peuvent avoir été créées artificiellement ou provenir d'un arrangement particulier de la scène réelle. Dans tous les cas, notre cerveau est influencé par le contexte en cherchant des indices lui permettant d'effectuer des déductions, en utilisant par exemple la mise en relation de grandeur entre elles (Fig.57).

58. Il existe également la disparité verticale, celle-ci se produit lorsque l'individu connaît un problème d'alignement entre les deux yeux.

59. En fait, [Jul60] a montré qu'il est possible d'obtenir une sensation de relief (ex. : stéréogrammes) uniquement grâce à la disparité binoculaire, c'est-à-dire sans que les facteurs monoculaires interviennent.

60. Environ 5 millions de cônes et 120 millions de bâtonnets.

61. Type de neurones chargés de recueillir l'information visuelle via des cellules bipolaires, des cellules amacrines et des cellules horizontales.

62. Le mécanisme des illusions a été identifié un peu partout le long des voies visuelles, même parfois dans la rétine.

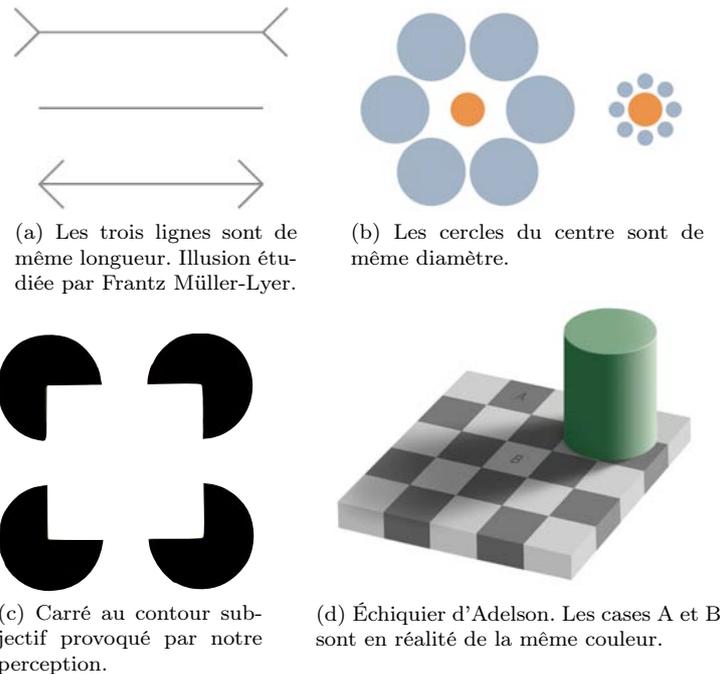


FIGURE 57: Les illusions d'optique sont la manifestation de notre cerveau à recréer de l'information en étant influencé par le contexte et l'expérience.

Nous venons de voir quelques caractéristiques fondamentales constituant notre perception visuelle. Bien que les technologies d'affichage des HMD actuels ne soient pas encore matures, nous pouvons utiliser les formidables capacités de notre cerveau à ajouter, corriger et recréer l'information visuelle.

13.2 Caractérisation du prototype

13.2.1 Objectifs

Pour satisfaire nos propositions de perception, l'utilisation d'un système de visualisation de réalité augmentée doit répondre aux exigences du domaine d'application. En ce qui nous concerne, les critères sont les suivants :

- L'opérateur doit avoir les mains libres.
- L'opérateur doit pouvoir appréhender facilement son environnement immédiat.
- L'opérateur doit pouvoir maintenir un champ visuel le plus naturel possible.
- Transmettre à l'expert ce que voit exactement l'opérateur.
- Transmettre à l'expert une vision globale de ce que perçoit l'opérateur.

C'est éclairé par le fonctionnement du système visuel humain et par la description des technologies d'affichage en réalité augmentée (voir 2.3 page 33) que nous allons expliquer les choix effectués pour caractériser notre propre système d'affichage respectant les critères ci-dessus.

13.2.2 Choix

L'opérateur doit avoir les mains libres.

Dans notre cadre de téléassistance pour les systèmes mécaniques, l'opérateur ne doit pas avoir les mains contraintes par le système, ce qui exclut l'utilisation des Handheld-Displays. De plus, nous devons avoir un dispositif simple à mettre en œuvre et peu intrusif pour son porteur et son environnement. Cela exclut donc l'utilisation des Projective-Displays. *Nous nous sommes alors orientés vers l'utilisation d'un dispositif d'affichage type Head-Worn-Displays.*

Nous avons cependant conscience des problèmes induits par l'utilisation des HWD. Nombre d'études rapportent les phénomènes qui peuvent intervenir durant le port de celui-ci. [HGAB08] par exemple, met en avant la fatigue visuelle provoquée par un conflit entre la distance à laquelle les yeux accommodent sur l'affichage (écran ou miroir) et la distance virtuelle à laquelle l'objet affiché est perçu⁶³. D'autres problèmes comme la sous-estimation des distances ont été signalés. Bien que ce phénomène soit encore mal compris, [KL04] a écarté le paramètre « champ de vue restreint », et le mode de vision binoculaire ou monoculaire n'en serait également pas la cause [Cut03].

Il existe de nombreux autres problèmes, mais leur étude dépasse le cadre de ce document et leurs possibles causes dépendent des conditions physiologiques propres aux individus. Pour information, nous signalons la diplopie, la rivalité rétinienne, la phorie ou encore certains troubles musculo-squelettiques dus au poids des HWD.

L'opérateur doit pouvoir appréhender facilement son environnement immédiat.

Un opérateur en maintenance ne peut s'affranchir du monde réel qui l'entoure, il ne peut être en immersion totale à l'instar de la réalité virtuelle.

Nous ne cherchons pas non plus à surcharger la vue de l'opérateur. En effet, une mauvaise perception dans la zone visuelle augmentée est potentiellement dangereuse lors d'une tâche de maintenance. *Nous avons donc opté pour un dispositif d'affichage monoculaire*, offrant ainsi uniquement une fenêtre augmentée sur le monde réel. Le faible angle d'affichage pour les augmentations des HWD actuels se révélant ici un atout !

En choisissant un système monoculaire, nous limiterons alors au maximum l'effet d'immersion et l'effet tunnel, et ce, afin de préserver les zones fonctionnelles de la vision humaine précédemment décrites.

Mais là encore, il existe de possibles phénomènes dégradant l'expérience du port d'un dispositif monoculaire. L'anomalie majeure est certainement la rivalité binoculaire. Lorsque l'on présente séparément à chaque œil des images contrastées, mais très différentes entre elles, le cerveau réagit de façon instable en nous faisant apparaître une alternance entre les deux images. Ce mécanisme n'est pas effectué de manière consciente, par conséquent il n'est pas contrôlable. Nous conseillons donc, même si le mode monoculaire offre une zone d'augmentation limitée, de ne pas la surcharger d'informations virtuelles.

63. Problème connu sous l'appellation « rivalité accommodation-vergence ».

L'opérateur doit pouvoir maintenir un champ visuel le plus naturel possible.

Afin de conserver un champ visuel « naturel », le dispositif monoculaire devra impérativement donner l'illusion à son porteur de voir ce qu'il aurait dû voir à l'œil nu. Un tel système See Through est dit orthoscopique.

Mais un dispositif See Through n'est pas forcément orthoscopique. Ainsi, pour un dispositif en vision indirecte comme les Video See Through (VST), le monde réel n'est perçu que par le biais d'une caméra. L'orthoscopie n'est alors pas obligatoire dans le cas d'une vision binoculaire⁶⁴. Il en va tout autrement en vision monoculaire si l'on veut respecter la cohérence entre les deux yeux (l'un est en prise directe avec la réalité, l'autre l'apercevant de manière détournée). Un dispositif de type Optical See Through (OST) par contre, apporte une vision directe du monde réel à travers un miroir semi-transparent. Ces dispositifs sont donc obligatoirement orthoscopiques, tant en vision monoculaire qu'en vision binoculaire.

A priori, il est donc plus facile d'obtenir un système orthoscopique grâce à un dispositif OST, la littérature comparative entre les VST et les OST y faisant d'ailleurs référence [RF00][Azu97]. *Nous avons pourtant fait le choix d'un dispositif VST*, principalement pour les raisons suivantes :

- Le défaut majeur des miroirs semi-transparents composant les OST est l'opacité. En monoculaire, cela devient capital. En effet, ces miroirs, de par leur conception, transmettent un pourcentage assez faible de la lumière provenant du monde réel, 30% seulement pour certains modèles [Hol92], environ 50% pour les meilleurs à l'heure actuelle⁶⁵. Cela peut donc engendrer un problème de rivalité binoculaire, la différence de luminosité perçue par le système visuel étant trop importante pour qu'il puisse la compenser. En ce sens, les expériences de [LW02] ont montré qu'en monoculaire, une opacité de 50% ou plus dégrade fortement la perception, engendrant des effets négatifs sur les performances d'une tâche.
- De façon idéale, la luminosité et le contraste du monde réel et du monde virtuel doivent être semblables. Grâce au couple « caméra-afficheur vidéo », il devient numériquement possible d'ajuster ces paramètres. Bien que cette solution soit limitée par la dynamique et les spécificités des technologies employées, nous pouvons nous approcher des caractéristiques d'une image perçue directement par l'œil. À ce moment-là, nous pouvons compter sur la capacité du système visuel humain à s'adapter à luminance moyenne et à recréer une vision cohérente de l'ensemble des informations perçues par les deux yeux, compétences que nous avons décrites précédemment.
- Loin d'atteindre l'acuité visuelle⁶⁶ de l'œil, les résolutions des caméras et des afficheurs vidéos deviennent pourtant relativement importantes. À résolution constante, la taille angulaire pour chaque pixel augmente en même temps que le champ de vision, la conséquence étant une perte de la définition de l'image. En ce qui nous concerne, le fait de vouloir limiter la zone d'augmentation nous amène à limiter également le champ de vue de la caméra, et donc à augmenter l'acuité du système. Ainsi, encore une fois, nous souhaitons jouer sur la capacité du système visuel humain

64. La perception de la réalité dans ce cas est alors substituée à notre regard, l'immersion étant semblable à celle des casques de réalité virtuelle.

65. Pour avoir une idée de ce que représente une opacité de 50%, le lecteur pourra, en pleine journée, tenter de regarder son environnement à travers un rétroviseur intérieur de voiture.

66. Le test d'acuité visuelle de Snellen en optométrie donne une acuité V définie par : $V = \frac{20}{D}$ avec D représentant la distance à laquelle un observateur ayant une vision normale identifie une lettre de l'alphabet. Une acuité normale est de 20/20, le seuil de non-voyance étant de 20/200 (extrait de [Cha02]). Pour un visiocasque ayant un FOV de 26° et une définition horizontale de 480 pixels, l'acuité visuelle est d'environ 20/47.

à traiter l'information pour des images ayant une définition différente (les porteurs de lentilles correctrices sont souvent coutumiers du fait!).

- Les états de l'art sur la comparaison VST-OST évoquent le problème de la parallaxe. Ce problème de précision de l'alignement optique entre la caméra, l'affichage et l'oeil peut se corriger de manière informatique (en termes d'image affichée devant l'oeil).
- Un dispositif VST à notre sens possède des modes d'affichage plus intéressants pour notre application. En effet, nous avons ici le choix de pouvoir l'utiliser soit comme un système augmentant la vue de l'utilisateur, soit comme un afficheur classique permettant de présenter de l'information sans rapport direct avec le point de vue (vidéo explicative, schéma technique...), ce que ne permet pas de faire un OST de façon confortable.

Vis-à-vis d'un OST, le choix d'un VST peut avoir des conséquences sur la vision stéréoscopique en champ proche⁶⁷. La caméra étant fixe, elle ne respecte pas le principe de convergence binoculaire. Aussi, pour limiter ce phénomène, nous suggérons de légèrement excentrer l'afficheur vers l'extérieur de l'axe de l'oeil. L'objectif est d'éviter que l'image de l'affichage ne soit trop dans la zone de chevauchement des yeux, la capacité de détection stéréoscopique diminuant rapidement en direction du champ de vision périphérique. Pour une éventuelle prise en charge de ce problème, [SAH⁺01] présente en ce sens une technique de sélection dynamique des parties d'une image.

Transmettre à l'expert ce que voit exactement l'opérateur.

Le fait d'avoir opté pour un dispositif VST va nous soustraire au problème de l'acquisition du point de vue exact de l'oeil du porteur. En effet, un dispositif OST ne permet pas de faire cela de manière aisée, obligeant généralement à mettre en place un système optique complexe à fabriquer [Cak06][SKF05] pour amener une partie des rayons lumineux vers la caméra. Dans notre cas, nous pouvons alors utiliser directement le flux de la caméra du dispositif VST pour retranscrire à l'expert la vision orthoscopique de l'opérateur.

Transmettre à l'expert une vision globale de ce que perçoit l'opérateur.

Pour ce dernier point, aucune difficulté particulière, tant pour un dispositif OST que VST. En utilisant une caméra grand-angle fixée au HMD, il devient possible de saisir une partie du champ visuel de l'opérateur. De façon courante, nous pouvons trouver dans le commerce des caméras ne nécessitant aucun objectif particulier pour des angles d'ouverture allant de 90° à 120°. Nous faisons remarquer également au lecteur qu'il est envisageable, pour un dispositif VST, d'utiliser la caméra en charge de reproduire la vision orthoscopique pour capturer la vision périphérique, en jouant de façon astucieuse sur le choix de la résolution du capteur de la caméra et sur sa focale⁶⁸. Ainsi, à partir du même flux vidéo, la partie centrale de l'image serait utilisée pour la vision orthoscopique (après correction informatique), l'image totale pour la vision périphérique.

67. [Cut03] propose une division de l'espace perceptuel en trois régions : champ proche (espace manipulable par les mains), champ moyen (jusqu'à 30m, limite pour laquelle une conversation peut être tenue), champ lointain (au-delà de 30m, espace visuel profond).

68. Pour cela, afin de ne pas dégrader l'acuité du système, il faut augmenter la résolution de la caméra et diminuer la focale de l'objectif.

Note : Nous sommes conscients que la conception d'un HWD s'avère extrêmement complexe si l'on veut tenir compte de la multitude de paramètres entrant en jeu. Néanmoins à l'heure actuelle nos besoins ne sont que partiellement couverts par les systèmes existants. Afin que nous puissions expérimenter nos propositions de visualisation et d'interaction, nous avons alors orienté nos choix en fonction de nos aspirations.

13.2.3 Prototype MOVST

Nous avons donc créé notre propre système de visualisation Video See Through monoculaire orthoscopique. Nous l'avons baptisé **MOVST**, acronyme anglophone de « **M**onocular **O**rthoscopic **V**ideo **S**ee **T**hrough ».

N'utilisant pas de dispositif optique pour la correction de la parallaxe, sa mise en oeuvre a pu être entièrement conçue avec des éléments commerciaux (Fig.58).



FIGURE 58: Nos lunettes MOVST.

Pour la caméra orthoscopique, nous utilisons une caméra dont nous avons modifié le « packaging » afin de la rendre la moins encombrante possible. Elle possède un capteur couleur haute résolution Sony de type HQ 1/3'' Ex-View HAD PAL associé à un objectif d'ouverture F2.0, et un angle de vision d'environ 30°, le tout pour un poids inférieur à 30 grammes. La caméra périphérique quant à elle, possède un angle de vision de 90° et un capteur CMOS de 307 200 pixels d'une résolution de 640x480 provenant d'une webcam QuickCam Pro 5000 de chez Logitech.

Le système d'affichage est un ICUITI M920 pesant moins de 100 grammes, équipé d'un écran LCD VGA couleur 24 bits de 307 200 pixels d'une résolution de 640x480 et d'un champ de vision de 26° environ.

La caméra est fixée directement sur l'afficheur en alignement avec celui-ci afin de minimiser la parallaxe entre les deux. Nous avons par la suite créé un programme informatique en charge de corriger plus précisément la parallaxe, assurant ainsi l'orthoscopie du système. Pour les calculs, ceux-ci sont effectués par un ordinateur équipé d'un processeur Intel Core 2 Duo à 2GHz avec 2Go de RAM, nous ayant assuré un nombre de 25 images par seconde en termes d'affichage. Nous avons également testé les calculs de correction de la parallaxe sur une carte graphique NVidia GeForce 9400M d'un ordinateur portable Apple MacBookPro équipé du processeur Intel Core 2 Duo à 2.26GHz avec 2Go de RAM. L'implémentation de l'algorithme a été réalisée grâce au logiciel QuartzComposer et nous a permis d'atteindre les 70 images par seconde.

14 INTERACTION : IMPLÉMENTATION DE P.O.A.

Lorsque nous avons présenté le rôle des divers composants constituant l'application T.A.C. au paragraphe 12.4.2, nous avons présenté le fait que nous ne traitons pas de façon similaire les augmentations de type « Picking » et celles de type « Outlining-Adding ». Cela est dû à la nature même de ces augmentations. C'est ce que nous allons décrire plus précisément dans cette partie, puis nous terminerons en présentant l'interface graphique par laquelle justement l'expert sera en mesure de « se mettre à la place » de l'opérateur et d'interagir avec lui.

14.1 L'opérateur : un être capable et en mouvement

Durant une tâche de maintenance, les déplacements physiques dans la zone de travail et les mouvements de la tête de l'opérateur sont inhérents à son activité. Si au cours d'une phase collaborative il est naturel pour les individus de se mouvoir et de reconnaître les constituants de leur environnement immédiat, il n'en va pas de même lorsque l'on désire transposer cette capacité à un ordinateur. Faire un acte de désignation tout en étant en mouvement par exemple, implique de correctement suivre la cible quelle qu'elle soit ! *Reconnaître* un objet implique justement de connaître au préalable les caractéristiques de celui-ci, quel que soit le point de vue adopté par l'observateur ! La diversité des couleurs, formes, illuminations et autres caractéristiques physiques que peuvent prendre les éléments nous obligerait à modéliser entièrement l'espace de travail, ce qui, bien que possible, n'est pas réaliste à l'heure actuelle.

Pour pallier ces difficultés, nous allons voir quelles sont les solutions que nous avons choisies et décidé de mettre en oeuvre pour retranscrire notre paradigme P.O.A.

14.2 Picking : Suivi et Latence

Le « Picking » est un acte de désignation, qu'il cible un objet tangible ou qu'il veuille exprimer une direction. Il peut alors s'effectuer sur un élément quelconque de l'environnement dans lequel est immergé l'opérateur. Nul besoin donc d'identifier ce qui se cachera sous le pointeur de la souris lorsque l'expert cliquera dans le flux orthoscopique. Mais l'augmentation associée, apparaissant devant le regard de l'opérateur, doit être recalée en accord avec les mouvements qu'il serait susceptible d'effectuer pendant cette phase. En vision par ordinateur, des techniques sont particulièrement adaptées à ce genre de situation : les méthodes de suivi basées sur des points d'intérêt.

14.2.1 Le suivi de points d'intérêt

Lorsque des objets *non définis* doivent être suivis, il faut alors s'intéresser au mouvement relatif qu'ils peuvent avoir vis-à-vis de l'observateur. Ces techniques de suivi de mouvements impliquent de façon typique le suivi de points visuels « significatifs », autrement appelés points d'intérêt.

L'idée principale des points d'intérêt est de regarder la fonction d'intensité dans une image pour en extraire les points de discontinuités, provoquées par exemple par des variations de contraste. La littérature est riche de techniques permettant de définir des points d'intérêt dans une image. Parmi les plus connus, nous pouvons citer le détecteur de coin de Harris [HS88], le détecteur SIFT (Scale-Invariant Feature Transform) de [Low04] ou encore GoodFeatureToTrack de [ST94].

Une fois ces points d'intérêt déterminés dans l'image, leur suivi d'une image à l'autre est assuré par des algorithmes dits de flux optique (optical flow). Les algorithmes de Lucas-Kanade⁶⁹ [LK81][TK91], de Horn-Schunck [HS81] ou de Block-Matching [HZ95] en sont les représentants les plus populaires.

En ce qui nous concerne, nous avons choisi le couple GoodFeatureToTrack — Lucas-Kanade (KLT) principalement pour sa rapidité reconnue et son implémentation optimisée [Bou02] au sein de la bibliothèque de vision par ordinateur OpenCV⁷⁰.

Cependant, l'utilisation de KLT repose sur trois hypothèses :

1. La constance d'illumination entre deux images consécutives. Si nous faisons l'hypothèse d'un flux vidéo à 25 images par seconde, l'écart entre deux images est de $40ms$. Dans des conditions « normales » de maintenance, une variation de luminosité brusque dans ce laps de temps est improbable.
2. La cohérence spatiale. Les points immédiatement voisins des points d'intérêt doivent appartenir à une même surface et doivent avoir un mouvement similaire à ce même point d'intérêt. Dans notre cas, l'environnement de l'opérateur n'est pas hautement « déformable », la cohérence spatiale sera respectée.
3. La persistance temporelle. Deux images consécutives doivent varier très peu suite à un mouvement. En ce qui nous concerne, les mouvements de la tête sont principalement des mouvements de rotation, bien que le cou soit capable d'effectuer des translations, mais celles-ci sont limitées. La vitesse angulaire de la tête (suivant l'axe du cou) peut atteindre au maximum $800^\circ/s$, néanmoins nous n'effectuons en moyenne que des rotations de l'ordre de $110^\circ/s$ [FMP01]. En temps normal lors d'un acte de maintenance, les mouvements de la tête sont encore plus lents et ont déjà été identifiés comme tels par [KC98], la cause étant que nous concentrons notre regard principalement dans la zone d'intérêt de la tâche. [HF09] en a mesuré la vitesse angulaire lors d'expérimentations sur des sujets portant des HMD. Il ressort de cela que la vitesse angulaire moyenne suivant l'axe du cou est d'environ 7 à $8^\circ/s$, et de 4 à $5^\circ/s$ suivant l'axe parallèle aux épaules. Ces faibles vitesses angulaires sont certainement le fait d'avoir directement les instructions virtuelles dans le champ visuel. En termes de vision par ordinateur, ces mouvements traduiront de petites variations entre deux images consécutives d'un flux vidéo. Si l'on considère un nombre de 25 images par seconde ($\Delta_t = 40ms$), alors le décalage entre les deux images sera approximativement de l'ordre du degré.

14.2.2 Implémentation du « Picking »

Nous évoquons au chapitre I que « l'exigence de recalage parfait entre le réel et le virtuel n'est pas justifiée si l'on s'attache par exemple, à présenter des informations utiles à la tâche (consignes, instructions...). Bien qu'imparfaitement recalés, le système cognitif de l'utilisateur associera alors aisément la zone réelle ciblée par l'action aux indices virtuels ». Cela est vrai dans la mesure où la zone réelle qui est ciblée par l'indice virtuel le reste. Tout décalage, par exemple pour une flèche indiquant une vis parmi plusieurs contiguës, peut amener une ambiguïté dans la désignation.

69. Algorithme plus connu sous le nom de KLT (Kanade-Lucas-Tomasi).

70. Bibliothèque initialement développée par INTEL. Disponible sous licence BSD à l'adresse : <http://opencv.willowgarage.com/wiki/>

Dès lors que l'expert cliquera dans le flux orthoscopique, le pixel sous le pointeur ne sera pas forcément un point d'intérêt au sens de l'analyse d'image. Il se peut alors que le suivi de ce pixel entre deux images consécutives dérive, ou échoue dans le pire des cas. Pour éviter cela, nous proposons la méthode suivante :

1. Quand l'expert désigne l'élément souhaité dans l'image, nous récupérons les coordonnées du pixel cliqué. Pour respecter l'hypothèse de la cohérence spatiale de l'algorithme de suivi KLT, nous définissons une zone de 5 pixels de large⁷¹ de part et d'autre du point cliqué (Fig.59a).
2. Dans la fenêtre entourant le pixel cliqué, nous lançons une recherche de points d'intérêt via GoodFeatureToTrack (Fig.59b). L'avantage induit par une recherche limitée dans cette zone de 11 x 11 pixels est le faible coût calculatoire.
3. Pour chacun des points d'intérêt, nous calculons le vecteur défini par le point d'intérêt concerné et le pixel cliqué (Fig.59c). Nous détaillons l'avantage d'un tel procédé ci-dessous.
4. Il ne reste plus qu'à associer l'élément virtuel ayant pour base le pixel cliqué (Fig.59d).
5. Enfin, à l'image suivante, nous lancerons l'algorithme KLT garantissant le suivi de ces points.

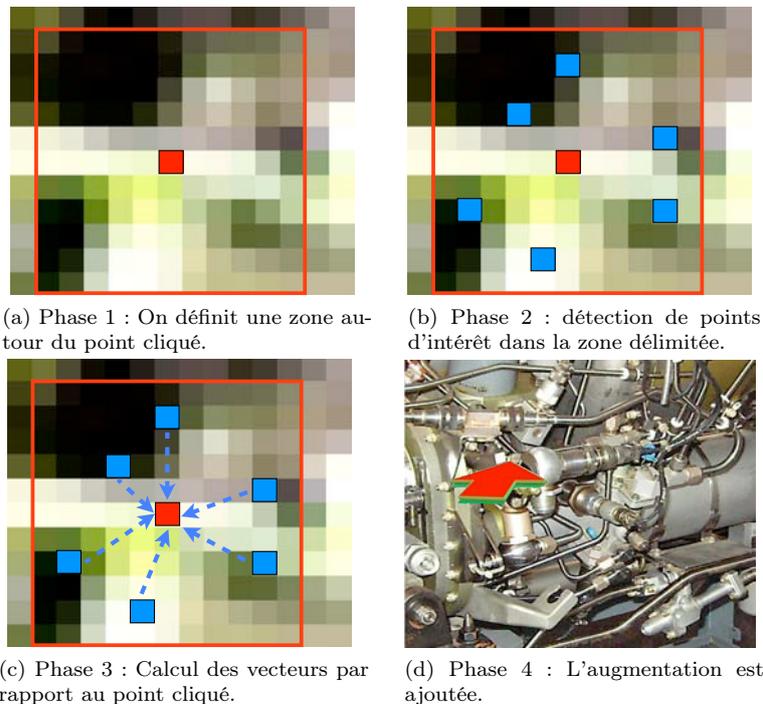


FIGURE 59: Méthode proposée pour l'implémentation du « Picking ».

C'est lorsque l'on doit assurer le suivi que notre méthode met en exergue tous ses avantages. En effet, KLT ne va pas suivre le point cliqué, potentiellement non porteur d'information, mais va suivre les points d'intérêt l'entourant. Dans le flux vidéo, le mouvement de la tête de l'opérateur va induire, pour un *petit déplacement*, une transformation (*quasi*) affine entre deux images consécutives. Nous n'avons alors plus qu'à calculer la transformation affine associée⁷², et cela en nous basant sur les points d'intérêt suivis. En appliquant la transformation à chaque vecteur, il nous est ainsi aisé, dans l'image $n+1$, de retrouver la position (au pixel près) du point cliqué.

71. La taille de la zone a été définie empiriquement.

72. La bibliothèque OpenCV propose, entre autres, une fonctionnalité permettant de calculer la transformation affine entre deux ensembles de points appartenant à deux images différentes.

14.2.3 La latence entre l'expert et l'opérateur

Le système T.A.C. est avant tout un système composé de deux applications (TAC-expert et TAC-opérateur) qui doivent communiquer entre elles à distance. Cet éloignement est générateur de retard dans la boucle de traitement de l'information. En effet, il existe un décalage entre le moment où le flux orthoscopique perçu est envoyé à l'expert, où celui-ci décide de l'augmenter, et le moment où cette augmentation sera affichée devant le regard de l'opérateur. L'opérateur ayant pu effectuer des mouvements⁷³ entre-temps, il faut alors être capable de correctement retranscrire la désignation de l'expert. Grâce à la méthode de suivi de points d'intérêt et aux données inertielles fournies par la centrale, il devient alors possible de proposer une solution simple permettant de gérer correctement les augmentations. La méthode que nous proposons est la suivante :

1. Avant d'être envoyée à l'expert, l'image orthoscopique est horodatée et associée aux valeurs inertielles⁷⁴ disponibles au moment de sa prise à l'instant t . Cette image est ensuite stockée dans une mémoire tampon côté opérateur.
2. À l'instant $t + \Delta t$, l'expert reçoit l'image ayant transitée par le réseau. Il décide alors d'y désigner un élément. En résultent une recherche de points d'intérêt et un calcul des vecteurs. Ces informations calculées et l'horodatage associé à l'image sont alors envoyés à l'opérateur. Pendant la phase que nous venons de décrire, la mémoire tampon de l'opérateur s'est remplie avec de nouvelles images.
3. À l'instant $t + n\Delta t$, l'opérateur reçoit les informations d'augmentation. À ce moment-là, nous comparons les valeurs inertielles (différence Δ_{inertiel}) entre l'image ayant cours dans l'afficheur de l'opérateur avec les valeurs inertielles de l'image ayant été augmentée par l'expert. Trois cas de figure se présentent à nous (Fig. 60) :
 - (a) Si la différence Δ_{inertiel} est inférieure à 5° sur chaque axe de rotation⁷⁵, alors le point de vue de l'opérateur a très peu bougé. Les deux images respectent ainsi l'hypothèse de persistance temporelle. Nous appliquons alors l'algorithme KLT pour retrouver les points d'intérêt dans l'image en cours en vue de l'augmenter.
 - (b) Si $5^\circ < \Delta_{\text{inertiel}} < 20^\circ$, alors l'hypothèse de persistance temporelle n'est plus respectée, la différence entre les deux images est trop importante pour KLT. Nous sommes donc contraints de rechercher dans la mémoire tampon l'image augmentée par l'expert, puis de suivre de proche en proche les points d'intérêt pour chaque image du tampon jusqu'à l'image en cours⁷⁶. À l'issue, celle-ci sera augmentée.
 - (c) Si $\Delta_{\text{inertiel}} > 20^\circ$, alors le point de vue actuel de l'opérateur n'a plus rien à voir avec le point de vue de l'image augmentée par l'expert. Il est donc inutile de vouloir retrouver les points d'intérêt. Cependant, même si l'augmentation ne peut s'effectuer, nous indiquons dans le champ de vue de l'opérateur, par le biais d'une flèche clignotante, la direction dans laquelle celui-ci doit porter son attention. S'en suivra un échange verbal entre les protagonistes pour réitérer l'acte de désignation.

73. Comme nous l'avons évoqué précédemment, durant une phase de maintenance, l'attention de l'opérateur est portée sur une zone précise, cette attitude impliquant principalement des mouvements de rotation de la tête.

74. La centrale inertielle nous fournit des valeurs de rotation type Roulis, Tangage et Lacet. Pour éviter le phénomène de blocage de cardan (*gimbal lock*), nous utilisons dans nos calculs une représentation sous forme de quaternion.

75. Les valeurs de Δ_{inertiel} ont été déterminées de manière empirique pour notre cas d'étude, pour une distance entre la tête de l'opérateur et de l'objet comprise dans la zone manipulable (champ proche, au sens de [Cut03]).

76. Nous faisons remarquer que le tampon reçoit une nouvelle image toutes les $40ms$, et que le temps de calcul pour KLT est négligeable en comparaison ($< 1ms$ dans nos conditions d'implémentation). La phase de suivi peut donc remonter le tampon sans risque de ne point atteindre l'image courante.

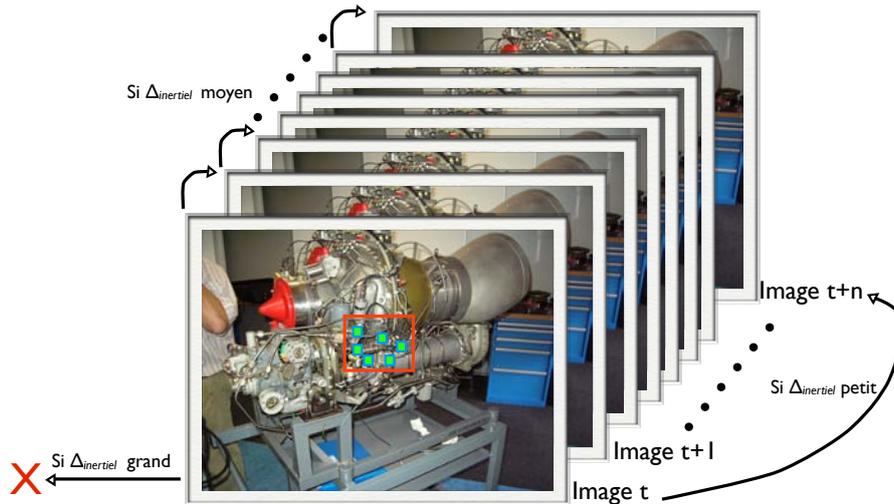


FIGURE 60: Procédé de suivi des points d'intérêt dans la mémoire tampon en fonction de la différence inertielle $\Delta_{inertiel}$.

14.3 Outlining-Adding : Reconnaissance et Localisation

Dans le cadre des travaux de cette thèse, nous n'avons pu terminer le développement de la partie implémentant le principe de « l'Outlining » et de « l'Adding ». En effet, la mise en oeuvre de ces principes nécessite un suivi temps réel robuste d'objets 3D en ayant une connaissance a priori du modèle. Nous avons toutefois commencé la conception des sous-composants associés « Détection » et « Pose ». De ce fait, nous allons de façon succincte présenter les grandes étapes de notre algorithme. Nous nous sommes inspirés de la méthode décrite par [PHMG06] que nous avons adaptée. Cela nécessite une phase d'initialisation manuelle préalable pour construire un « modèle 3D de points d'intérêt ». Cette phase est la suivante :

1. Nous prenons une image de référence de l'objet réel en connaissant le point de vue associé.
2. Nous extrayons de cette image les points d'intérêt (blob detection) grâce à l'algorithme SURF [BETVG08]⁷⁷.
3. En utilisant les fonctionnalités de la bibliothèque OpenGL⁷⁸, nous projetons ces points d'intérêt sur le modèle 3D CAO pour obtenir une correspondance 2D-3D. Nous obtenons ainsi un « modèle 3D de points d'intérêt ».

Par la suite, lorsque nous voulons reconnaître cet objet dans la scène réelle, pour chaque image du flux orthoscopique, nous cherchons les points d'intérêt SURF. Si suffisamment de points de l'image en cours sont associés avec ceux du modèle 3D de points d'intérêt, alors l'objet recherché a été trouvé. Ceci fait, il nous reste à calculer la position-orientation de cet objet grâce à l'algorithme POSIT [DD95]. Nous pouvons alors augmenter le flux vidéo.

À l'heure actuelle, le développement de ces composants est toujours en cours, la description de notre algorithme fera donc l'objet d'un détail dans de futurs travaux.

⁷⁷. Celui-ci présente les mêmes avantages (invariance à la rotation et à l'échelle) que SIFT [Low04] sur lequel il est basé, la rapidité en plus.

⁷⁸. En particulier la fonction `gluUnproject()`.

14.4 L'interface « Expert »

Pour visualiser et interagir avec l'opérateur, le composant « IHM » de l'application TAC-expert prend la forme d'une interface graphique fenêtrée (Fig.61). Celle-ci a été développée grâce à la bibliothèque Qt fournissant tous les widgets nécessaires à son élaboration. Pour respecter un principe d'utilisation de l'outil efficace et efficient au sens de [BB03], c'est-à-dire une utilisation tournée vers l'utilisateur final, l'interface ne présentera pas directement les paradigmes technologiques [Coo95], c'est à dire les multiples facettes représentant les réglages et possibilités avancées de l'application.

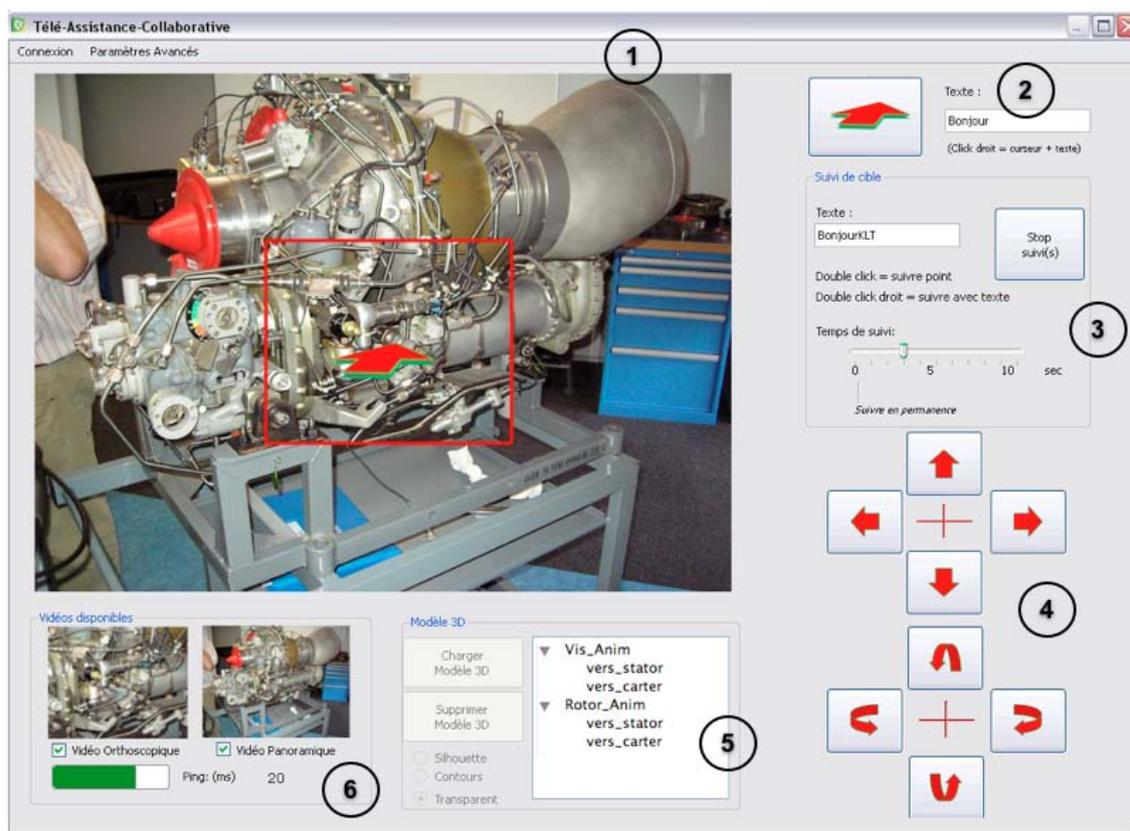


FIGURE 61: Interface de l'expert.

Description des différentes zones de l'interface graphique :

1. **Zone de visualisation PiP.** C'est l'affichage à proprement parler incrustant le flux orthoscopique de l'opérateur (intérieur du cadre rouge) dans le flux périphérique représentant l'environnement immédiat de celui-ci. L'expert aura donc la possibilité d'enrichir le flux orthoscopique en cliquant simplement sur celui-ci.
2. **Interaction Picking.** Avant de pouvoir cliquer dans le flux orthoscopique, l'expert, s'il désire faire une désignation, devra au préalable activer cette fonctionnalité en appuyant sur la flèche rouge symbolisant cet acte. Ceci étant fait, un clic gauche dans la zone considérée entraînera une augmentation du flux via une flèche pendant une durée d'une seconde. La possibilité lui est également donnée de pouvoir associer une information textuelle qui apparaîtrait alors au côté de la flèche de désignation.

3. **Suivi de cible.** Dans le même esprit que « l'Outlining », l'expert peut avoir besoin de procéder à de multiples désignations en souhaitant que celles-ci ne disparaissent pas au bout d'une seconde, gardant ainsi les éléments d'intérêt dans le champ visuel de l'opérateur. Pour ce faire, il lui suffit de régler la durée de l'augmentation et de faire un double-clic gauche sur les parties désirées (avec ou sans texte associé).
4. **Sélection rapide d'icônes.** Toujours en ce qui concerne le principe du « Picking », l'expert peut choisir rapidement ici le type d'augmentation en fonction de ce qu'il souhaite exprimer (ex. : un sens pour tourner une pièce, donner une direction...).
5. **Interaction Outlining-Adding.** Pour que l'expert puisse utiliser une interaction de type « Outlining », il devra au préalable charger le modèle 3D de l'élément considéré (ex. : ici le turbomoteur). Ceci fait, lorsqu'il effectuera un clic annexe (ex. : clic droit + touche espace) sur la pièce de son choix dans le flux orthoscopique, celle-ci apparaîtra « surlignée », mettant soit en évidence sa silhouette, soit ses contours (visibles et cachés). La pièce étant reconnue par le logiciel, une liste d'animations associée sera présentée, l'expert n'ayant plus qu'à sélectionner celle désirée pour jouer l'augmentation, effectuant ainsi une interaction de type « Adding ».
6. **Modes vidéo.** L'incrustation orthoscopique dans le flux périphérique apparaît comme un « zoom » d'une zone considérée. Ainsi, nous donnons la possibilité à l'expert de rapidement désactiver cette vue orthoscopique pour n'avoir accès qu'à la vue périphérique. De même, l'incrustation orthoscopique peut sembler trop petite dans certains cas d'utilisation. L'expert pourra désactiver la vision périphérique, laissant alors place entièrement à la vision orthoscopique. De plus, pour que l'expert ait une idée de la qualité de la connexion, et donc conscience du décalage potentiel dans la communication, nous lui en indiquons le niveau grâce à une barre à progression de couleur (du rouge au vert).

15 ÉVALUATION EXPÉRIMENTALE

Dans le cadre de ces travaux de thèse, nous avons pu développer une partie du système T.A.C., notamment les lunettes MOVST, ainsi que l'interface logicielle de l'expert, application intégrant le paradigme du « Picking » et la façon de percevoir l'environnement de l'opérateur. Pour tester en partie la pertinence de nos propositions, nous avons donc voulu expérimenter le principe des références ostensives, mécanisme clé du succès d'une communication interpersonnelle.

Dans ce qui suit, pour établir une évaluation conforme respectant les procédures, analyses et termes utilisés, nous avons suivi les directives communément employées dans les méthodologies des sciences humaines (ressources : [Gou03], Professeur au Collège québécois d'Ahuntsic).

15.1 Ce que l'on cherche à évaluer

Dans le cadre d'une assistance à distance pour un opérateur, nous souhaitons comparer différents moyens de communication entre eux. Pour cela, nous testerons l'opérateur exécutant des tâches a priori simples : des opérations de maintenance sur une imprimante et sur un ordinateur. Le téléphone étant le moyen le plus largement utilisé pour effectuer des tâches de maintenance, nous avons voulu tester la pertinence d'une communication audio seule face à un système comme T.A.C. De plus, nous avons voulu tester l'intérêt de la vision See Trough de nos lunettes MOVST vis-à-vis d'une vision déportée sur un écran ordinaire.

De façon classique, nous nous sommes principalement intéressés dans cette étude à répondre à la question suivante :

« Est-ce que les différents modes de communications⁷⁹ permettent d'exécuter une tâche avec la même rapidité ? »

Cet aspect, bien que réducteur, est un point de vue fondamental pour une application à vocation industrielle. En effet, l'introduction d'un nouvel outil technologique ne doit pas faire perdre plus de temps en comparaison des méthodes conventionnelles. Si tel est le cas, l'outil doit alors justifier de contributions additionnelles (un meilleur contrôle qualité par exemple!).

En fin de test, nous avons également questionné les utilisateurs sur leurs impressions (facilité d'utilisation, niveau de conversation avec l'expert...) en vue d'établir un classement des systèmes.

15.2 Description de l'évaluation

15.2.1 Descriptions des tâches et des moyens de téléassistance

Nous avons utilisé deux types de produits industriels à maintenir. Pour chacun de ces produits, nous avons déterminé trois opérations de maintenance. Toutes les opérations, entre elles, ont une difficulté équivalente en termes de manipulation (dextérité requise) avec un temps d'exécution relativement proche.

Le premier produit industriel à maintenir est un ordinateur type PC. Les trois opérations sont les suivantes :

⁷⁹. Sous entendu « mode de communication soutenant une collaboration distante ».

- Changer une carte d’extension par une autre avec un branchement spécifique.
- Changer une barrette mémoire (RAM) défectueuse.
- Effectuer les branchements d’un lecteur DVD interne (nappe de données, prise son et prise d’alimentation).

Le deuxième produit industriel est une imprimante laser noir et blanc. Les trois sous-tâches sont les suivantes :

- Enlever le toner, le bac de récupération d’encre et la résistance chauffante.
- Changer le toner, le bac de récupération d’encre et la résistance chauffante (la procédure n’est pas totalement identique à la première).
- Changer physiquement le réglage d’impression papier (enlever le bac papier, positionner trois curseurs à trois endroits différents, remettre le bac papier et positionner un autre curseur dans l’imprimante).

Pour chacune des six tâches, des procédures linéaires strictes ont été établies (avec l’aide des manuels d’entretien de chacun des systèmes). En ce qui concerne les moyens de téléassistance utilisés pour les tests, nous avons donc voulu comparer la pertinence de T.A.C. face au moyen le plus largement utilisé qu’est le téléphone. Nous avons testé trois configurations :

- TEL : un téléphone mains libres pour pouvoir converser de manière classique entre l’opérateur et l’expert.
- VISIO : de la réalité augmentée en vision déportée, c’est-à-dire sans les lunettes MOVST. L’opérateur est équipé d’une caméra cravate grand-angle et d’un dispositif d’affichage (écran 20”) placé sur le plan de travail. L’expert a toujours la possibilité de pouvoir désigner sur le flux vidéo dit « orthoscopique ». Une communication audio mains libres est bien sûr utilisée.
- T.A.C. : c’est l’utilisation du système T.A.C. avec les lunettes MOVST, donc en vision See Through orthoscopique. Nous rappelons que les protagonistes peuvent communiquer oralement entre eux.

15.2.2 Sujets et plan de recherche

Sujets

Onze sujets ont participé à l’étude en tant qu’opérateur. Tous sont des hommes avec une moyenne d’âge de 22 ans. Ils n’ont aucun lien de parenté entre eux. Ils ne sont pas non plus familiers des applications collaboratives et n’ont aucune expérience de téléassistance, ce qui malgré leur faible nombre en fait un échantillon représentatif⁸⁰. Ils n’ont jamais effectué de manipulation sur les produits industriels utilisés durant l’expérience.

En ce qui concerne l’expert, le sujet est un homme de 26 ans. Il est expert en maintenance informatique et possède une expérience confirmée de la hotline. Il a préalablement été formé sur l’interface logicielle du système T.A.C. et n’a eu aucun contact avec les opérateurs qu’il ne connaissait pas.

⁸⁰. Un échantillon est dit représentatif lorsqu’il possède les mêmes caractéristiques que la population que l’on souhaite étudier. Cette représentativité doit surtout s’établir à partir de caractéristiques pouvant influencer les réponses.

Plan de recherche ⁸¹

Chaque participant a expérimenté les trois moyens de communication (TEL, VISIO, T.A.C.), l'ordre d'utilisation étant aléatoirement tiré. Pour chacun des trois modes, une tâche à accomplir est composée de deux opérations de maintenance (Ordinateur + imprimante) tirées aléatoirement parmi les six proposées. Les différents tirages aléatoires ont pour finalité de neutraliser (limiter) les variables parasites comme par exemple ici le phénomène d'apprentissage (l'accoutumance induite par l'apprentissage). Pour un moyen de communication donné, les deux sous-tâches sont exécutées de façon séquentielle.

Les tests se sont déroulés dans une salle contenant une table de travail (180x80 cm) avec tout le nécessaire (ordinateurs, imprimantes, tournevis...). La salle était correctement éclairée (lumière artificielle). Les participants n'ont pas été préalablement familiarisés aux nouveaux modes de communication. De par leurs natures (parler au téléphone, chausser des lunettes...), aucun apprentissage n'a donc été nécessaire. Pour ce qui est du lieu où se trouvait l'expert, celui-ci était équipé d'un système téléphonique mains libres et d'un ordinateur contenant l'application T.A.C. en connexion directe avec la salle de l'opérateur.

Nous avons mesuré le temps de réalisation pour chacune des sous-tâches, puis le temps total. Aucune limite de temps n'a été imposée. L'unique consigne que les participants aient reçue était celle de se laisser guider par l'expert distant. Toutes les manipulations ont été enregistrées pour un visionnage ultérieur. Afin de ne pas introduire de biais induit par la caméra fixe, nous avons utilisé la technique du « cheval de Troie » préconisée par [JMMP08]. Cela consiste à présenter l'usage de la caméra comme faisant partie du système. L'usage véritable leur est dévoilé en fin d'expérience (nous avons ensuite obtenu leur accord pour exploitation). Pour finir, nous leur avons demandé de remplir un questionnaire jugeant différents critères en vue d'établir un classement.

16 ANALYSE ET INTERPRÉTATIONS

Nous allons présenter les résultats en deux parties. Premièrement, nous allons examiner le temps de réalisation des tâches en fonction des systèmes utilisés. Deuxièmement, nous examinerons le résultat du questionnaire d'évaluation. Les analyses statistiques ont été effectuées à l'aide du logiciel SPSS (Statistical Package for the Social Sciences) de la société IBM.

Pour finir, nous discuterons les résultats de ces deux parties.

16.1 Méthode d'analyse des résultats

16.1.1 Pour le temps de réalisation

Nous avons souhaité étudier l'influence de 3 facteurs (TEL, VISIO, T.A.C.) sur la variable dépendante ⁸² quantitative qu'est le temps de réalisation d'une opération de maintenance. Pour cela nous avons donc soumis les sujets à chacun des niveaux de la variable indépendante ⁸³ (i.e le mode

⁸¹. Ensemble des procédures qui déterminent comment les sujets d'une recherche seront soumis aux différentes modalités ou niveaux de la variable indépendante. Il doit également décrire la façon dont seront neutralisées les variables parasites.

⁸². Variable observée/décrite et mesurée qui peut être modifiée par la variable indépendante.

⁸³. Facteur présumé responsable du phénomène étudié.

de communication) conformément au plan de recherche précédemment décrit. Nous sommes donc dans un plan de test à mesures répétées. Afin d'inférer une relation entre les modes de communication et le temps de réalisation, nous avons utilisé une analyse de la variance (ANOVA) à mesures répétées en ayant choisi un niveau de risque alpha⁸⁴ de 0.05.

Notre variable indépendante qualitative X est donc le mode de communication (à 3 niveaux de valeur), la variable dépendante quantitative Y étant le temps de réalisation d'une opération de maintenance.

Notre hypothèse statistique H0 (hypothèse nulle⁸⁵) est :

« *Le temps de réalisation de l'opération de maintenance est égal pour tous les modes de communication ($t_{TEL} = t_{VISIO} = t_{TAC}$)* ».

L'hypothèse bilatérale⁸⁶ H1 correspondant à notre objectif de recherche est donc :

« *Est-il vrai de dire que le temps de réalisation de l'opération varie en fonction des modes de communication utilisés ?* ».

16.1.2 Pour les données du questionnaire

À la fin des tests, chaque opérateur s'est vu remettre un questionnaire en six points sur la façon dont a été ressentie l'expérience. Pour chaque question, nous leur avons demandé d'évaluer les trois modes de collaboration sur une échelle ordinale de 0 (faible) à 14 (fort).

Pour examiner les réponses, nous avons utilisé le coefficient de concordance W de Kendall [KS39] en respectant les recommandations de Legendre [Leg04]. Ce coefficient va nous permettre de statuer sur le degré d'accord entre les différents sujets sur le classement des modes de collaboration.

Pour établir le questionnaire, nous avons sélectionné les questions que l'on retrouve le plus souvent dans la littérature⁸⁷ ainsi que l'échelle de notation utilisée (de 0 = faible à 14 = fort). Les questions ont été les suivantes :

1. « *La conversation avec l'expert est-elle facile ou difficile (effort de conversation) ?* »
2. « *À quelle vitesse estimez-vous avoir réalisé votre tâche ?* »
3. « *Est-il facile de faire le lien entre les indications de l'expert et le monde réel ?* »
4. « *Quel confort d'utilisation avez-vous ressenti ?* »
5. « *Est-il facile ou difficile de corriger ses erreurs durant la tâche ?* »
6. « *Quel niveau de stress avez-vous ressenti pour chaque mode de communication ?* »

16.2 Résultats et interprétations

16.2.1 Pour le temps de réalisation

La table 12 récapitule les statistiques descriptives des tests utilisateurs. Il y est indiqué pour chaque mode le temps de réalisation moyen et l'écart-type (standard deviation en anglais, abrégé STDEV).

84. Le risque α représente le risque de rejeter à tort une hypothèse alors quelle est vraie. Il est défini en termes de probabilité, et son seuil de signification en science humaine est de 5%.

85. Hypothèse qui postule qu'il n'y a pas de différence entre les moyennes des trois groupes (ou des trois mesures).

86. Hypothèse qui postule qu'il existe une différence entre les moyennes des trois groupes (ou des trois mesures).

87. Nous nous sommes aidés de l'état de l'art de [DGB08].

La variable F correspond à la variable du test de Fisher-Snedecor et donc au résultat de l'ANOVA. Quant à la variable p, elle représente l'erreur alpha, variable permettant de confirmer ou d'infirmier l'hypothèse bilatérale H1.

		Avg. time (s)	STDEV	F	p
Mode	TEL	441	83	0.699	0.509
	VISIO	444	123		
	T.A.C	404	91		

TABLE 12: Effet inter-sujet sur le temps de complétion.

L'analyse des données indique que les opérateurs, en utilisant le mode collaboratif TEL, ont mis un temps moyen de 441s ($\sigma = 83s$), un temps moyen de 444s ($\sigma = 123s$) pour VISIO, comparativement à 404s ($\sigma = 91s$) pour T.A.C. Nous observons donc un gain d'environ 10% en faveur de T.A.C. en comparaison des temps moyens quasi identiques de VISIO et TEL.

Le résultat de l'ANOVA à mesures répétées ($F(2,20) = 0.699$, $p = 0.509$, $\eta_p^2 = 0.065$) va nous permettre de statuer quant à la dépendance réelle du temps de réalisation en fonction des modes de communication. La valeur de F obtenue est de 0.699 soit une valeur inférieure à la valeur critique de 3.49 donnée par la table de Fisher-Snedecor. L'hypothèse nulle H0 doit donc être conservée. La probabilité de commettre l'erreur alpha sur le résultat de F est de $p = 0.509$, soit supérieure au seuil de 5%. La différence entre les trois modes de communication n'est alors pas statistiquement significative, bien qu'il apparaisse un gain de 10% en faveur de T.A.C.

On peut donc en conclure que statistiquement le mode collaboratif n'influence pas sur le temps de réalisation de l'opération de maintenance. Nous commenterons ce résultat par la suite.

Remarque 1 : Les valeurs 2 et 20 dans l'expression $F(2,20)$ représentent les degrés de liberté des variables inter et intra groupe. Ces paramètres servent à déterminer la valeur de F dans la table de Fisher-Snedecor.

Remarque 2 : L'effet de taille partiel (η_p^2) indique que 6.5% de la variation observée dans le temps d'accomplissement s'explique par l'utilisation du mode collaboratif, ce qui sur l'échelle de Cohen [Coh88] est un effet de taille moyen.

16.2.2 Pour les données du questionnaire

À la fin de l'expérience, les utilisateurs ont donc évalué les modes de collaboration en leur attribuant une note en fonction des paramètres que nous avons cherché à estimer. La table 13 en résume les notes moyennes (avec l'écart type associé) de chaque mode et pour chaque question (Q1 à Q6). La table 14 quant à elle présente le degré d'accord de ces notes entre les utilisateurs. Pour chaque question, il y est indiqué dans l'ordre : le classement moyen de chaque mode puis le degré d'accord (entre 0 et 1) sur ce classement avec la probabilité de commettre l'erreur alpha (niveau de risque de 0.05).

	TEL		VISIO		T.A.C	
	<i>Average</i>	<i>STDEV</i>	<i>Average</i>	<i>STDEV</i>	<i>Average</i>	<i>STDEV</i>
Q1	8.4	1.33	9.5	2.55	11.3	1.41
Q2	7.7	2.77	7.4	2.06	9.6	2.29
Q3	6.7	2.27	10.3	1.5	11.3	1.58
Q4	10	3.57	7.4	3.53	10.7	2.22
Q5	4.3	1.8	10	2.64	11.2	1.98
Q6	9.8	2.02	9.2	2.22	11.5	1.5

TABLE 13: Notes moyennes pour chaque mode en fonction des questions.

	Average Rank ^a			Kendall's W	
	<i>TEL</i>	<i>VISIO</i>	<i>T.A.C</i>	<i>W^b</i>	<i>p</i>
Q1	1.33	1.94	2.72	0.498	0.011
Q2	1.67	1.72	2.61	0.374	0.034
Q3	1	2.33	2.67	0.824	0.001
Q4	2.28	1.39	2.33	0.297	0.069
Q5	1	2.33	2.67	0.875	0.000
Q6	1.83	1.67	2.5	0.259	0.097

a: From 1 to 3 (Higher is better)

b: Accordance (Higher is better)

TABLE 14: Degré d'accord entre les différents sujets sur le classement des systèmes.

La question 1 (Q1) porte essentiellement sur la facilité de pouvoir s'exprimer et de se faire comprendre par l'expert. À la question : « *La conversation avec l'expert est-elle facile ou difficile (effort de conversation) ?* » (0=Très difficile à 14=Très facile), nous obtenons le classement TEL<VISIO<TAC. Cependant, le degré d'accord entre les sujets est de 49.8%. En regardant les rangs moyens, on remarque que le désaccord porte surtout entre TEL et VISIO. Les moyennes permettent toutefois de voir que la communication ne présente pas de difficulté majeure. Les sujets n'ayant pas eu connaissance de leur temps de réalisation, nous nous sommes intéressés à savoir comment ils ont perçu leur performance.

À la question 2 (Q2) : « *À quelle vitesse estimez-vous avoir réalisé votre tâche ?* » (0=Très lente à 14=Très rapide), nous obtenons le classement TEL<VISIO<T.A.C. Là encore le degré d'accord n'est que de 37.4%. Les rangs moyens nous indiquent que le désaccord porte encore sur TEL et VISIO. L'examen des moyennes nous indique cependant que les sujets n'ont pas vraiment eu l'impression d'accomplir plus rapidement leur opération de maintenance même s'ils accordent un gain de temps en faveur de T.A.C.

À la question 3 (Q3) : « *Est-il facile de faire le lien entre les indications de l'expert et le monde réel ?* » (0=Très difficile à 14=Très facile), nous obtenons là encore, avec un degré d'accord de 82.4%, le classement TEL<VISIO<T.A.C. Les moyennes indiquent clairement l'infériorité de TEL comparé à VISIO et à T.A.C. Ce résultat nous permet de conclure quant à l'efficacité de la désignation par l'expert sur la vision See Through orthoscopique ou déportée de l'opérateur. La différence entre VISIO et T.A.C. vient principalement de la charge mentale supplémentaire⁸⁸ qu'induit un mode de visualisation déporté comme VISIO.

La question 4 (Q4) traite plutôt du degré de confort à utiliser l'un ou l'autre des systèmes. À la question : « *Quel confort d'utilisation avez-vous ressenti ?* » (0=Très inconfortable à 14=Très confortable), nous obtenons cette fois le classement VISIO<TEL<T.A.C. Cependant le degré d'accord est seulement de 29.7%. En examinant de plus près les rangs et les moyennes, nous voyons que le désaccord porte surtout sur T.A.C. et TEL, néanmoins le confort d'utilisation ressenti n'est pas à remettre en cause. Toutefois, tous ont jugé VISIO très peu confortable à utiliser. Cela est certainement à mettre en relation avec la charge mentale supplémentaire précédemment évoquée et le fait que l'expert demandait souvent à l'opérateur de modifier sa position pour mieux percevoir la scène (via la caméra cravate).

La question 5 (Q5) concerne la gestion des erreurs. Quand nous leur avons demandé : « *Est-il facile ou difficile de corriger ses erreurs durant la tâche ?* » (0=Très difficile à 14=Très facile), nous obtenons le classement TEL<VISIO<T.A.C. avec un degré d'accord de 87.5%. En se penchant sur les rangs et les moyennes, on déduit une large supériorité de VISIO et T.A.C. Avec ces deux systèmes, l'expert se rend immédiatement compte des erreurs commises et en informe aussitôt l'opérateur qui peut les corriger d'autant plus facilement. Ce résultat illustre parfaitement l'utilité des modes de communication basés sur l'image.

Enfin, à la question 6 (Q6) : « *Quel niveau de stress avez-vous ressenti pour chaque mode de communication ?* » (0=Très stressé à 14=Très détendu), le classement est VISIO<TEL<T.A.C. avec un degré d'accord de 25.9%. Là encore, nous remarquons d'après les données statistiques que le désaccord porte sur TEL et VISIO qui, sans pouvoir les départager, ont tout de même été jugé comme n'induisant pas un niveau de stress important.

16.2.3 Discussion des résultats obtenus

Les premiers résultats semblent montrer que statistiquement parlant, il n'y a pas de différence significative en termes de temps d'exécution des tâches, et ce malgré un gain apparent de 10% en faveur de T.A.C. Ce résultat peut être mis en balance avec le nombre de participants impliqué dans l'expérimentation. Plus l'effectif (ou le nombre de degrés de liberté) est petit et plus la variance sans biais est incertaine. Néanmoins, les sujets ont été sélectionnés pour leur représentativité et ces premiers résultats nous permettent de dégager une première tendance.

Toutefois, après analyse des vidéos, il est intéressant de remarquer que les sujets prennent davantage leur temps pour réaliser la tâche avec les modes VISIO et T.A.C. Ce constat est en accord avec les résultats de la question 2 où les sujets n'ont pas eu l'impression d'aller plus vite avec l'un ou l'autre des modes. Sachant que l'expert voit ce qu'ils voient, ils désignent volontiers du doigt pour avoir une validation de la part de celui-ci. Les opérateurs ont donc moins d'hésitation au

⁸⁸. C'est à dire une transposition mentale entre la scène perçue sur l'écran et la scène réelle. Voir chapitre précédent.

final pour exécuter une action, et les remarques que nous avons relevées montrent qu'ils sont plus détendus lorsqu'ils savent que l'expert peut corriger directement leurs fautes. Ils entament ainsi plus volontiers une relation élève-enseignant.

Par contre, en ce qui concerne TEL, nous constatons beaucoup d'erreurs d'interprétation qui engendrent de l'hésitation. Certains sujets persistent même dans leurs erreurs tandis que d'autres ne savent plus quoi faire. Les sujets qui s'en sortent le mieux sont ceux qui instaurent un protocole de communication avec l'expert (confirmation vocale de l'ordre à exécuter, description de la tâche en cours puis acquittement de fin). Dès que l'on passe sur VISIO et T.A.C., les erreurs commises par TEL disparaissent ou sont très vite corrigées. Ces observations sont conformes aux résultats des questions 3 et 4. Les sujets comprennent immédiatement où ils doivent agir et sont rapidement interpellés par l'expert en cas d'erreurs.

Bien que l'on vienne de voir que l'attitude des sujets est positivement affectée par VISIO et T.A.C., une différence entre les deux persiste sur tous les classements effectués par les sujets. En effet dans la majorité des cas, nous avons eu des remarques comme quoi il est plus facile de se laisser guider par T.A.C. que par VISIO. Ceci est à mettre en rapport directement avec le type d'affichage des différents modes (résultat de la question 3). T.A.C., grâce aux lunettes MOVST, permet une vision mentale directe contrairement à VISIO qui utilise une vision déportée. Dans ce dernier cas, il faut alors faire l'effort de regarder les instructions à l'écran sur le plan de travail pour ensuite faire le lien avec la réalité, ce qui a souvent été source d'erreurs. Parfois, les sujets enlevaient même leur caméra cravate pour la pointer plus précisément sur une zone afin que l'expert puisse mieux y voir. Cependant dans un cas comme dans l'autre, les sujets ne semblent pas plus « appréhender » la tâche (question 6).

En revanche, nous avons relevé beaucoup de remarques concernant T.A.C. sur l'ergonomie des lunettes MOVST. Malgré un poids inférieur à 150 grammes, cela semble être encore trop lourd. Cette charge physique peut alors poser problème sur des tâches de longue durée. Il est important de souligner que les sujets ont toutefois apprécié la non-immersivité de ce type de lunette, ayant ainsi une meilleure perception de leur environnement.

Côté expert, le fait d'utiliser VISIO et T.A.C. est jugé appréciable, surtout la capacité de pouvoir visualiser ce que l'opérateur exécute, et la possibilité qui lui est offerte de pouvoir montrer rapidement là où il faut intervenir. En termes de perception, l'expert a jugé plus intéressant le système T.A.C. par rapport à VISIO pour deux raisons :

- La première est de pouvoir visualiser exactement ce que voit l'œil de l'opérateur, et donc de « s'identifier » à son point de vue. C'est aussi semble-t-il son plus gros défaut, car les mouvements de la tête de l'opérateur affectent beaucoup la stabilité de l'image. Quand l'opérateur bouge trop vite, il devient alors difficile de donner des indications en cliquant sur le flux vidéo orthoscopique.
- La deuxième est la présence du second flux vidéo représentant une vision plus globale de l'environnement immédiat de l'opérateur. L'expert s'est alors plus facilement fait une représentation mentale de l'espace de travail de l'opérateur. Chose intéressante, cette vision panoramique lui a beaucoup servi pour situer le sujet lors de mouvements rapides, problème évoqué ci-dessus.

Enfin, l'expert et les sujets ont soulevé un problème en termes d'interprétation des flèches virtuelles. En effet, seules des flèches de la même couleur (rouge) ont été prévues à la conception de l'application T.A.C. Lorsque l'expert a voulu utiliser la désignation comme un moyen de présenter

une action (par exemple une flèche virtuelle vers le bas pour indiquer un mouvement à effectuer vers soi), ceci a souvent été mal compris. Nous pensons qu'un code couleur pour les flèches pourrait être plus pertinent en termes d'association de concept (action ou désignation).

17 SYNTHÈSE

Dans ce chapitre, nous avons présenté le système T.A.C. Ce système est une implémentation matérielle et logicielle de la proposition de fonctionnement d'un système d'assistance à distance que nous avons énoncée au chapitre précédent. Nous proposons ici une façon simple pour un opérateur de pouvoir se laisser guider visuellement par un expert distant.

Durant nos travaux, nous avons été en mesure d'étudier la pertinence de T.A.C. en tant qu'interface permettant une collaboration à distance sur une tâche de maintenance. Plus précisément, nous avons pu expérimenter le principe des références ostensives et leurs implications sur la relation de communication entre un expert et un opérateur.

La problématique était double. D'un côté, il fallait retranscrire l'environnement immédiat de l'opérateur vers l'expert afin que celui-ci se représente mentalement la scène (et puisse guider l'opérateur grâce à des indices virtuels). De l'autre côté, il ne fallait pas faire perdre à l'opérateur la conscience de sa réalité tout en intégrant le plus naturellement possible les indices virtuels apportés par l'expert. En cela, les lunettes MOVST du système T.A.C. contribuent à compenser l'une des asymétries sociales impliquées par l'éloignement entre deux personnes. L'utilisation de références ostensives est désormais possible dans les deux sens.

À travers des tests utilisateurs, nous avons constaté qu'employer le mécanisme des références ostensives pouvait faire gagner environ 10% de temps dans l'achèvement d'une tâche de maintenance distante sur des produits industriels. Nous devons cependant poursuivre nos expérimentations sur un échantillon plus important d'utilisateurs pour que le résultat soit statistiquement significatif. N'ayant plus à se perdre en explications descriptives, nous avons également pu observer une amélioration qualitative de la relation de communication entre les protagonistes, plus enclins à la collaboration. Qui plus est, nous confirmons les résultats de [BKS99] montrant que les erreurs commises sont immédiatement détectées par l'expert lorsque l'on améliore la perception de celui-ci vis-à-vis de l'environnement de l'opérateur. L'expert apporte ainsi sa caution au résultat, ce qui conforte la qualité de l'opération de maintenance.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Rappel du contexte

L'assistance revêt un caractère fondamental dans la maintenance, qu'elle prenne la forme d'un support formalisé ou d'une personne ayant les compétences adéquates. Grande variété de systèmes mécatroniques toujours plus complexes, renouvellement des gammes de plus en plus rapide, multiplicité des procédures, manque de « visibilité » du système à maintenir et incertitude des opérations à réaliser sont autant de nouvelles contraintes auxquelles doivent faire face les techniciens. Et pour ne rien arranger, les multiples formations voient leur durée compressée, ce qui n'est pas sans conséquence sur l'apprentissage et l'acquisition des connaissances. Les mécaniciens sont alors enclins à se former « sur le tas », ce qui peut à terme impliquer un accroissement des erreurs de procédures et par ce biais augmenter les coûts de maintenance, pour au final engendrer des pertes de temps considérables.

Au fil de l'évolution technologique, la recherche de l'efficacité et de l'efficacités ont amené les industriels à identifier, tester et intégrer de nouveaux outils dans les processus d'aide, de la documentation papier en passant par le téléphone jusqu'aux systèmes informatiques. Grâce à ladite évolution technologique, la décennie écoulée a vu se « matérialiser » un concept, datant pourtant des années 60, qui risque fort de changer le mode de travail des individus : la Réalité Augmentée.

Alors ? Nouvel outil miracle en perspective ? En quoi cela consiste-t-il ? Aurait-on enfin une aide efficace ? Quelle forme pourrait prendre un système d'aide intégrant ce nouveau concept ? Comment faire en sorte de limiter, voire supprimer certaines erreurs humaines ? C'est à ces questions, tout au long de ce mémoire, auxquelles nous avons apporté des éléments de réponse.

Synthèse du document

Nous avons commencé ce mémoire en définissant ce qu'est la réalité augmentée. Pour cela, nous avons retracé l'évolution de ses définitions et nous sommes rendus compte qu'elles sont toujours soumises à discussion et interprétation, bien que certaines fassent référence. Nous avons alors proposé notre propre définition, avec pour objectif d'établir une compréhension commune de ce concept lors de la lecture de ce document. Nous avons ensuite présenté les fondements technologiques qui constituent un système de réalité augmentée, cela afin d'en comprendre les rouages et donc les possibilités nous étant offertes dans la conception de notre propre système.

Nous avons consacré en partie le chapitre suivant à l'étude des mécanismes cognitifs qui amènent tout individu à réaliser des erreurs, et plus particulièrement dans la maintenance. Bien que les aides actuelles soient là pour y remédier, il est apparu que celles-ci sont jugées ambiguës, mal construites, peu pratiques, voire incomplètes, les opérateurs préférant donc la plupart du temps s'en remettre à eux-mêmes. Nous avons alors exposé les avantages des systèmes d'aide basés sur la réalité augmentée. Parmi les deux grands types que nous avons distingués, à savoir les « systèmes automatisés » et les « systèmes supervisés », nous nous sommes plus particulièrement intéressés à

cette dernière catégorie. En effet, lorsque survient une situation imprévue dans le formalisme de la tâche, les systèmes automatisés deviennent impuissants, et il faut alors avoir recours à une aide humaine extérieure. Les systèmes supervisés sont les représentants d'une telle solution, permettant à deux individus distants d'établir une relation de collaboration, mais les interactions possibles se révèlent insuffisantes. Les manques constatés nous ont alors permis d'établir notre problématique.

Par la suite, nous avons voulu montrer pourquoi, malgré ses capacités à faire des erreurs, l'humain est toujours un maillon indispensable en termes d'assistance, et donc pourquoi les systèmes supervisés sont à l'heure actuelle des outils incontournables. Ainsi, nous avons vu que la maîtrise des connaissances explicites et tacites est fondamentale à l'expertise, et qu'un expert humain possède des aptitudes naturelles à s'adapter au niveau de l'apprenant. De plus, une relation de collaboration entre les individus favorise l'émergence d'idées et de solutions tournées vers la résolution de problème. Nous avons alors décrypté les mécanismes d'une communication interpersonnelle afin d'en reporter les bénéfices au sein d'une collaboration distante. Il en est ressorti que les systèmes supervisés actuels ne font que peu de place à retranscrire correctement l'espace visuel de l'opérateur pour l'expert, et que les capacités d'interactions de celui-ci ne lui permettent pas d'user de références ostensives. En cela, nous nous sommes attachés à faire des propositions visant à « simuler » la coprésence de l'expert auprès de l'opérateur.

Enfin, nous avons présenté une implémentation de nos propositions sous la forme d'un système supervisé que nous avons nommé « **T.A.C.** ». Durant son développement, nous avons été en mesure de tester l'influence des références ostensives durant une phase de maintenance distante. Les premiers résultats montrent les avantages induits par la prise en compte de nos propositions.

Résumé de nos contributions et résultats

La question à laquelle nous avons voulu répondre durant nos travaux est partie intégrante de la formulation de la problématique que nous avons énoncée, à savoir :

« Pourquoi l'expert est-il encore un chaînon incontournable en termes d'assistance, et comment faire en sorte qu'il puisse à distance interagir avec un opérateur de façon presque aussi naturelle que s'il était coprésent ? ».

Pour cela, nous avons *identifié et mis en avant* les principaux mécanismes régissant une communication interpersonnelle lorsque des individus tentent de résoudre ensemble un problème posé :

- Les protagonistes partagent le même espace visuel. Dans la collaboration à distance, l'expert n'a pas forcément de relation spatiale avec les objets. Celui-ci doit donc pouvoir disposer d'un espace visuel périphérique afin de lui faciliter la prise de conscience de la situation. Ceci aura un effet direct sur la coordination des actions avec l'opérateur, et permettra de connaître en permanence le statut des travaux. Le manque de vision périphérique ferait donc perdre l'efficacité de la communication dans l'accomplissement d'une tâche. En termes d'espace visuel, il faut également tenir compte de la vision de précision qui est utilisée pour faciliter l'examen de détail.
- Les protagonistes ont la possibilité d'utiliser des références ostensives, c'est-à-dire des deixis (« ça ! », « là ! ») associées à des gestes de désignation. De nombreuses recherches suggèrent l'importance de la désignation d'objets dans un travail collaboratif. Ce type d'interaction est directement en rapport avec la notion d'espace visuel partagé cité ci-dessus.

Nous nous sommes alors en premier lieu attachés à proposer le **fonctionnement général d'un système supervisé** intégrant ces mécanismes : permettre à l'expert de voir ce que voit l'opérateur, comme s'il était « à sa place », et lui donner une possibilité d'augmenter une partie du champ visuel de celui-ci via un paradigme d'interaction adéquat.

Côté opérateur, pour respecter les exigences qu'impose son travail, nous avons traduit cela en contraintes :

1. Il doit avoir les mains libres pour exécuter sa tâche. L'utilisation du système d'aide ne doit pas entraver son activité.
2. Il doit être équipé d'un système de visualisation de réalité augmentée, mais celui-ci ne doit pas l'empêcher d'appréhender son environnement immédiat (condition de non-immersivité).
3. Le système de visualisation doit être orthoscopique, donnant l'impression d'une vision naturelle à son porteur. L'avantage est double en cas d'augmentation : ne pas soustraire le regard de la tâche et éviter le phénomène de transposition mentale entre réel et virtuel.
4. Il ne doit pas avoir un champ visuel surchargé d'éléments virtuels, toujours dans l'optique de ne pas lui occulter sa perception de la réalité.
5. Le système de visualisation doit être capable de retranscrire tant la vision orthoscopique que périphérique à l'expert.

Ces contraintes nous ont amené à concevoir notre propre système d'affichage de réalité augmentée, à savoir un HMD monoculaire orthoscopique See Through, système que nous avons nommé **MOVST**, qui est également capable d'envoyer via des flux vidéo la perception de son porteur.

Côté expert, les lunettes MOVST sont le lien lui permettant d'être virtuellement à la place de l'opérateur. Avoir accès à ce que voit exactement celui-ci le dispense du décalage de point de vue habituellement présent lorsque deux individus sont côte à côte. Il peut ainsi de façon précise guider l'opérateur dans l'accomplissement de son travail. De plus, la vision périphérique lui rend compte de la situation de la tâche, et donc d'une meilleure compréhension de celle-ci dans son contexte. Pour visualiser correctement ces deux flux d'information, nous avons suggéré une interface représentant la vidéo périphérique dans laquelle serait incrustée la vidéo orthoscopique (**visualisation PiP**).

Le partage de l'espace visuel nous a ouvert des possibilités d'interaction pour l'expert à l'égard de l'opérateur, et ce, grâce à la réalité augmentée. Nous avons alors proposé un nouveau paradigme d'interaction, nommé **P.O.A** pour « Picking Outlining Adding » dont les principes sont :

- **Picking** : Une interaction représentative du geste déictique.
- **Outlining** : Ajouter des éléments virtuels au regard de l'opérateur de manière à souligner les informations vocales données par l'expert, ou pour augmenter la compréhension d'une situation donnée.
- **Adding** : Calqué sur le principe des gestes kinétiques, l'Adding substitue ce geste en utilisant des animations 2D/3D adéquates.

En nous basant sur le principe de fonctionnement que nous avons énoncé, sur nos lunettes MOVST, sur notre visualisation PiP et sur notre paradigme P.O.A., nous avons conçu et implémenté notre propre système supervisé que nous avons baptisé **T.A.C.**, pour Télé-Assistance-Collaborative. Lors de l'élaboration de celui-ci, nous avons également présenté une méthode simple permettant un suivi

précis pour la désignation ainsi qu'un moyen de gérer la latence de communication, et donc les augmentations.

Durant nos travaux, nous avons été en mesure d'expérimenter le principe des références ostensives. À travers des tests utilisateurs, nous avons constaté qu'apporter ce mécanisme pouvait faire gagner environ 10% de temps dans l'achèvement d'une tâche de maintenance distante sur des produits industriels. Mais au-delà du temps de complétion d'une tâche, nous avons également pu observer de multiples bénéfices, comme une amélioration qualitative de la relation de communication entre les protagonistes, plus enclins à la collaboration, ou encore une détection immédiate des erreurs commises par l'opérateur. Dans ce dernier cas, l'avantage est double, l'expert apporte sa caution au résultat, l'opérateur est conforté quant à la qualité de l'opération de maintenance.

Ces premiers résultats tendent à montrer l'importance d'intégrer dans des systèmes supervisés les mécanismes de communications normalement réservés à des relations de coprésence, et nous encourageant donc à poursuivre dans cette voie.

Limitations et Perspectives

Durant le chapitre précédent, nous avons pu voir que le développement du système T.A.C. n'est pas entièrement finalisé. En effet, l'implémentation des principes de « l'Adding » et de « l'Oulining » nécessite d'explorer plus profondément le domaine de la réalité augmentée qui s'intéresse au suivi robuste temps réel d'objets 3D. La littérature sur le sujet offre deux principales catégories pour ce faire, les algorithmes basés primitives et les algorithmes basés sur une connaissance a priori du modèle (2D ou 3D). Comme nous l'avons indiqué, nous nous sommes orientés vers cette dernière catégorie, et pour cela nous avons commencé à développer notre propre méthode de suivi en nous inspirant des travaux de [PHMG06]. À l'issue de ce développement, il sera alors possible d'expérimenter en totalité les principes de notre paradigme d'interaction P.O.A.

Les premiers résultats obtenus durant notre expérimentation montrent que notre système pourrait supporter des processus de collaboration imprévisibles et indéterminés, et cela, grâce à une meilleure retranscription de l'espace visuel de l'opérateur à l'expert et à la possibilité pour celui-ci d'user de références ostensives. Cependant, bien que nos résultats expriment un potentiel favorable en termes de collaboration distante, nous devons poursuivre nos expérimentations sur un échantillon de tests plus important pour que les tendances, parfois fortes, soient statistiquement significatives.

Nous devons également nous intéresser à l'effet inattendu lorsque l'expert a voulu utiliser la désignation comme un moyen de présenter une action (par exemple une flèche virtuelle vers le bas pour indiquer un mouvement à effectuer vers soi). Ceci a souvent été mal interprété par les opérateurs. Nous connaissons l'importance des couleurs en termes d'association de concept pour les humains, cela mérite certainement d'être transposé à notre cas.

Enfin, l'essence même de T.A.C. en fait un système n'étant pas cantonné à la maintenance industrielle, nous souhaitons alors élargir son champ d'action en tant qu'outil pour la formation à distance (e-learning, validation d'acquis pour des e-portfolio...).

Références

A

- [ABB⁺01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Comput. Graph. Appl.*, 21(6) :34–47, 2001.
- [AHAS08] Yannick Allusse, Patrick Horain, Ankit Agarwal, and Cindula Saipriyadarshan. Gpucv : an opensource gpu-accelerated framework for image processing and computer vision. In *MM '08 : Proceeding of the 16th ACM international conference on Multimedia*, pages 1089–1092, New York, NY, USA, 2008. ACM.
- [AMR02] AMRA. Augmented reality assistance in train maintenance tasks. <http://evra.ibisc.univ-evry.fr/index.php/AMRA>, 2002.
- [Ana06] Margarita Anastassova. *L'analyse ergonomique des besoins en amont de la conception de technologies émergentes : le cas de la réalité augmentée pour la formation à la maintenance automobile*. PhD thesis, Université René Descartes, Paris 5, 2006.
- [And83] John Robert Anderson. *The architecture of cognition*. Lawrence Erlbaum Associates, 1983.
- [ARM07] ARMAR. Augmented Reality for Maintenance and Repair. <http://graphics.cs.columbia.edu/projects/armar/>, 2007.
- [ARV99] ARVIKA. Augmented reality for development, production, servicing. <http://www.arvika.de>, 1999.
- [Azu97] Ronald Azuma. A survey of augmented reality. *Presence : Teleoperators and Virtual Environments*, 6(4) :355–385, 1997.

B

- [Bar38] Chester I. Barnard. *The Functions of the Executive*. Harvard University Press, 1938.
- [BB03] Eric Brangier and Javier Barcenilla. *Concevoir un produit facile à utiliser : Adapter les technologies à l'homme*. Edition d'organisation, Paris, 2003.
- [BB09] Peter Bui and Jay Brockman. Performance analysis of accelerated image registration using gpgpu. In *GPGPU-2 : Proceedings of 2nd Workshop on General Purpose Processing on Graphics Processing Units*, pages 38–45, New York, NY, USA, 2009. ACM.
- [BBK⁺01] Martin Bauer, Bernd Bruegge, Gudrun Klinker, Asa MacWilliams, Thomas Reicher, Stefan Riß, Christian Sandor, and Martin Wagner. Design of a component-based augmented reality framework. In *ISAR '01 : Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, page 45, Washington, DC, USA, 2001. IEEE Computer Society.
- [Bér94] Francois Bérard. *Vision par ordinateur pour la réalité augmentée : Application au bureau numérique*. Master's thesis, Université Joseph Fourier, Grenoble, 1994.
- [BETVG08] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3) :346–359, 2008.

- [BG09] Patti Bao and Darren Gergle. What's "this" you say? : the use of local references on distant displays. In *CHI '09 : Proceedings of the 27th international conference on Human factors in computing systems*, pages 1029–1032, New York, NY, USA, 2009. ACM.
- [BHKS98] Martin Bauer, Timo Heiber, Gerd Kortuem, and Zary Segall. A collaborative wearable system with remote sensing. In *ISWC '98 : Proceedings of the 2nd IEEE International Symposium on Wearable Computers*, page 10, Washington, DC, USA, 1998. IEEE Computer Society.
- [BKBB99] M. Billinghamurst, H. Kato, S. Bee, and J. Bowskill. Asymmetries in collaborative wearable interfaces. *Wearable Computers, IEEE International Symposium*, 0 :133, 1999.
- [BKS99] Martin Bauer, Gerd Kortuem, and Zary Segall. "where are you pointing at?" a study of remote collaboration in a wearable videoconference system. In *ISWC '99 : Proceedings of the 3rd IEEE International Symposium on Wearable Computers*, page 151, Washington, DC, USA, 1999. IEEE Computer Society.
- [BOO95] Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *DIS '95 : Proceedings of the 1st conference on Designing interactive systems*, pages 157–166, New York, NY, USA, 1995. ACM.
- [Bou02] Jean-Yves Bouguet. Pyramidal implementation of the lucas kanade feature tracker : Description of the algorithm, 2002.
- [BR05] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality : Merging Real and Virtual Worlds*. A K Peters, Ltd., 2005.
- [Bra01] Eric Brangier. Approche ergonomique de l'assistance technique à l'homme : vers un modèle de la symbiose homme-technologie? In *IHM et assistance technique*, pages 48–56, 2001.
- [Bra07] Eric Brangier. Comment améliorer la performance de l'opérateur par des dispositifs d'aide au travail? In *Lévy-Leboyer, C, Louche, C., et Rolland, J-P (Eds)*, Management des organisations, pages 429–450. Edition d'organisation (Paris), 2007.
- [BRW98] Josianne Basque, Johanne Rocheleau, and Laura Winer. Une approche pédagogique pour l'école informatisée. Technical report, École informatisée Clés en main du Québec, 1998.
- [BSC04] Stanislaw Borkowski, Sherif Sabry, and James L. Crowley. Projector-camera pair : An universal io device for human machine interaction. In *Polish National Robotics Conference KKR VIII*, 2004.

C

- [Cae96] Jean Caelen. Définition et caractérisation des interfaces homme-machine. In *Nouvelles Interfaces Homme-Machine*, Arago 18, pages 31–43. Observatoire français des techniques avancées (OFTA), 1996.
- [Cak06] J. Cakmakci, O. Rolland. Head-worn displays : a review. *Display Technology*, 2 :199–216, 2006.

- [Can86] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6) :679–698, 1986.
- [CB91] Herbert H. Clark and Susan E. Brennan. Perspectives on socially shared cognition. In L. Resnick, J. Levine, and S. Teasley, editors, *American Psychological Association, ch. Grounding in Communication*, pages 127–149, Washington, 1991.
- [CGCS02] Alex Chaparro, Loren S. Groff, Barbara S. Chaparro, and Deborah Scarlett. Survey of aviation technical manuals phase 2 report : User evaluation of maintenance documents. Technical Report DOT/FAA/AR-02/34, U.S. Department of Transportation, Federal Aviation Administration, Washington, D.C. 20591, 2002.
- [Cha02] Virgile Charton. *Etude comparative de la perception d’ambiances lumineuses en milieu réel et en milieu virtuel*. PhD thesis, Institut National des Sciences Appliquées de Lyon, 2002.
- [CHR04] Ozan Cakmakci, Yonggang Ha, and Jannick P. Rolland. A compact optical see-through head-worn display with occlusion support. In *ISMAR ’04 : Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 16–25, Washington, DC, USA, 2004. IEEE Computer Society.
- [CK04] Herbert H. Clark and M. A. Krych. Speaking while monitoring addresses for understanding. In *Journal of Memory and Language*, volume 50, pages 62–81, 2004.
- [Cla96] Herbert H. Clark. *Using language*. Cambridge University Press, 1996.
- [CM92] T.P. Caudell and D.W. Mizell. Augmented reality : an application of heads-up display technology to manual manufacturing processes. *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, 2 :659–669, 1992.
- [Coh88] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Taylor & Francis, Inc., 2nd edition edition, 1988.
- [Coo95] Alan Cooper. The myth of metaphor. *Visual Basic Programmer’s Journal*, 1995.
- [Cou06] P. Couedelo. CAMKA system. <http://www.camka.com>, 2006.
- [CS89] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13(2) :259 – 294, 1989.
- [Cut03] J.E. Cutting. Reconceiving perceptual space. *Looking into pictures : An interdisciplinary approach to pictorial space*, pages 215–238, 2003.

D

- [dB09] Nelly de Bonnefoy. *Système Informationnel Mixte, Ubiquitaire et Intuitif - Support aux opérateurs nomades de maintenance aéronautique*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France, 2009.
- [DBBO95] Pierre Dillenbourg, Michael J. Baker, Agnès Blaye, and Claire O’Malley. The evolution of research on collaborative learning. In E. Spada and Reiman P., editors, *Learning in Humans and Machine : Towards an interdisciplinary learning science*, pages 189–211, 1995.
- [DD80] Stuart Dreyfus and Hubert Dreyfus. A five-stage model of the mental activities involved in directed skill acquisition. Technical report, United States Air Force Office of Scientific Research, University of California, Berkeley, 1980.

- [DD95] Daniel F. Dementhon and Larry S. Davis. Model-based object pose in 25 lines of code. *Int. J. Comput. Vision*, 15(1-2) :123–141, 1995.
- [Dés97] Mario Désilets. Connaissances déclaratives et procédurales : des confusions à dissiper. *Revue des sciences de l'éducation*, 23(2) :289–308, 1997.
- [DGB08] Andreas Dünser, Raphaël Grasset, and Mark Billinghurst. A survey of evaluation techniques used in augmented reality studies. In *SIGGRAPH Asia '08 : ACM SIGGRAPH ASIA 2008 courses*, pages 1–27, New York, NY, USA, 2008. ACM.
- [Did05] Jean-Yves Didier. *Contributions à la dextérité d'un système de réalité augmentée mobile appliqué à la maintenance industrielle*. PhD thesis, Université d'Evry Val d'Essonne, Evry, 2005.
- [DN00] Emmanuel Dubois and Laurence Nigay. Augmented reality : which augmentation for which reality? In *DARE '00 : Proceedings of DARE 2000 on Designing augmented reality environments*, pages 165–166, New York, NY, USA, 2000. ACM.
- [DRM⁺05] Jean-Yves Didier, David Roussel, Malik Mallem, Samir Otmane, Sylvie Naudet, Quoc-Cuong Pham, Steve Bourgeois, Christine Mégard, Christophe Leroux, and Arnaud Hocquard. AMRA : Augmented Reality Assistance for Train Maintenance Tasks. In *Workshop Industrial Augmented Reality, 4th ACM/IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2005)*, Vienna Autriche, 2005.
- [DSAO⁺97] G. Doherty-Sneddon, A. Anderson, C. O'Malley, S. Langton, S. Garrod, and V. Bruce. Face-to-face and video mediated communication : A comparison of dialogue structure and task performance. In *Journal of Experimental Psychology*, volume 3, pages 105–125, 1997.
- [Dup97] Gordon Dupont. The dirty dozen errors in maintenance. In *Human Factors Issues in Aircraft Maintenance and Inspection Meeting 11 Proceedings*, Washington, DC : FAA, 1997.
- [DW99] Desmond F. D'Souza and Alan Cameron Wills. *Objects, components, and frameworks with UML : the catalysis approach*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

E

- [EF69] Paul Ekman and Wallace V. Friesen. The repertoire of nonverbal behavioral categories : Origins, usage, and coding. In *Semiotica*, volume 1, pages 49–98, 1969.
- [Efr41] David Efron. *Gesture and environment*. King's Crown Press., New York, 1941.
- [EGR91] C. A. Ellis, S. J. Gibbs, and G.L. Rein. Groupware : Some issues and experiences. In *Communications of the ACM*, pages 38–58, 1991.
- [Emb00] David Embrey. Preventing human error : Developing a consensus led safety culture based on best practice. *Human Reliability Associates Ltd*, 2000.
- [EN93] Peggy A. Ertmer and Timothy J. Newby. Behaviorism, cognitivism, constructivism : Comparing critical features from a design perspective. In *Performance Improvement Quarterly*, volume 6, pages 50–72, 1993.

F

- [FMP01] Philippe Fuchs, Guillaume Moreau, and J.P. Papin. *Le traité de la réalité virtuelle, première édition*. les Presses de l'Ecole des Mines de Paris, 2001.
- [FMS93] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Commun. ACM*, 36(7) :53–62, 1993.
- [Fra00] Mike C. Fraser. *Working with Objects in Collaborative Virtual Environments*. PhD thesis, University of Nottingham, 2000.
- [FSK03] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *CHI '03 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 513–520, New York, NY, USA, 2003. ACM.
- [FSY⁺04] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human Computer Interaction*, 19(3) :273–309, 2004.
- [FV02] Morten Fjeld and Benedikt M. Voegtli. Augmented chemistry : An interactive educational workbench. In *ISMAR '02 : Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 259, Washington, DC, USA, 2002. IEEE Computer Society.

G

- [Gao92] Zhu Ying Gao. *Etude psychophysique de la vision en relief humaine en télévision stéréo*. PhD thesis, INSA Lyon, 1992.
- [Gib77] James Jerome Gibson. The theory of affordances. In R. Shaw and J. Bransford, editors, *Perceiving, acting, and knowing : Toward an ecological psychology*, pages 67–82, Hillsdale, NJ : Erlbaum, 1977.
- [Gib86] James Gibson. *The Ecological Approach To Visual Perception*. Lawrence Erlbaum Associates, new edition edition, 1986.
- [GKF04] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. Action as language in a shared visual space. In *CSCW '04 : Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 487–496. ACM Press, 2004.
- [Gou03] Claude Goulet. PI@nète Psy. Méthodologies des sciences humaines. http://www.collegeahuntsic.qc.ca/Pagesdept/Sc_Sociales/psy/methosite/, 2003.
- [Gra89] Barbara Gray. *Collaborating : Finding Common Ground for Multiparty Problems*. Jossey-Bass, 1989.
- [Gro98] Vincent Grosjean. *Assistance à la conduite dans les situations dynamiques : influence de la construction d'une perspective temporelle sur la performance experte*. PhD thesis, Université de Liège, 1998.
- [Gru94] J. Grudin. Computer-supported cooperative work : History and focus. In *Computer*, volume 27(5), pages 19–26, 1994.
- [GSH⁺07] Joseph L. Gabbard, J. Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. Active text drawing styles for outdoor augmented reality : A user-based study and design implications. *IEEE VIRTUAL REALITY CONFERENCE*, page 42, 2007.

H

- [HB03] David J. Haniff and Chris Baber. User evaluation of augmented reality systems. In *IV '03 : Proceedings of the Seventh International Conference on Information Visualization*, page 505, Washington, DC, USA, 2003. IEEE Computer Society.
- [HF07] Steve Henderson and Steve Feiner. Augmented reality for maintenance and repair (armar). *United States Air Force Research Lab*, Technical Report AFRL-RH-WP-TR-2007-0112, 2007.
- [HF08] Steven J. Henderson and Steven Feiner. Opportunistic controls : leveraging natural affordances as tangible user interfaces for augmented reality. In *VRST '08 : Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 211–218, New York, NY, USA, 2008. ACM.
- [HF09] Steven J. Henderson and Steven Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *ISMAR '09 : Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 135–144, Washington, DC, USA, 2009. IEEE Computer Society.
- [HF10] Steven Henderson and Steven Feiner. Opportunistic tangible user interfaces for augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 16(1) :4–16, 2010.
- [HGAB08] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3) :1–30, 2008.
- [Hit08] HitLab. BuildAR, <http://www.hitlabnz.org/wiki/buildar>, 2008.
- [HL91] Christian Heath and Paul Luff. Disembodied conduct : communication through video in a multi-media office environment. In *CHI '91 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 99–103, New York, NY, USA, 1991. ACM.
- [HL96] Christian Heath and Paul Luff. Convergent activities : Collaborative work and multimedia technology in london underground line control rooms. In *Cognition and Communication at Work : Distributed Cognition in the Workplace*, pages 96–130. In D. Middleton and Y. Engestrom (eds.), 1996.
- [Hol92] Douglas E. Holmgren. Design and construction of a 30-degree see-through head-mounted display. Technical report, Chapel Hill, NC, USA, 1992.
- [Hol95] Richard Lee Holloway. *Registration errors in augmented reality systems*. PhD thesis, Chapel Hill, NC, USA, 1995.
- [Hol97] Richard L. Holloway. Registration error analysis for augmented reality. *Presence : Teleoperators and Virtual Environments*, 6(4) :413–432, 1997.
- [HS81] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *ARTIFICIAL INTELLIGENCE*, 17 :185–203, 1981.
- [HS88] C Harris and M Stephens. A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

- [HV96] Beverly L. Harrison and Kim J. Vicente. An experimental evaluation of transparent menu usage. In *CHI '96 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 391–398, New York, NY, USA, 1996. ACM.
- [HZ95] Yan Huang and Xinhua Zhuang. Motion-partitioned adaptive block matching for video compression. In *ICIP '95 : Proceedings of the 1995 International Conference on Image Processing (Vol. 1)-Volume 1*, page 554, Washington, DC, USA, 1995. IEEE Computer Society.

J

- [Jed94] Didier Jedliczka. *Compétences et alternances*. Éditions Liaisons, 1994.
- [JLS97] Marco C. Jacobs, Mark A. Livingston, and Andrei State. Managing latency in complex augmented reality systems. In *I3D '97 : Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 49–ff., New York, NY, USA, 1997. ACM.
- [JMMP08] Francis Jambon, Nadine Mandran, Brigitte Meillon, and Christian Perrot. Évaluation des systèmes mobiles et ubiquitaires : proposition de méthodologie et retours d'expérience. *Ergo-IA'08 "L'humain au coeur des systèmes et de leur développement"*, pages 107–116., 2008.
- [Joh88] R. Johansen. *Groupware : Computer Support for Business Teams*. The Free Press, 1988.
- [Jon91] David H. Jonassen. Objectivism versus constructivism : Do we need a new philosophical paradigm ? In *Educational Technology Research and Development*, volume 39, pages 5–14, 1991.
- [Jul60] Bella Julesz. Binocular depth perception of computer-generated patterns. In *Bell System Technical Journal*, number 39, pages 1125–62, 1960.

K

- [Kal60] RE Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, (82 (Series D)) :35–45, 1960.
- [KB99] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *IWAR '99 : Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, page 85, Washington, DC, USA, 1999. IEEE Computer Society.
- [KC98] Kurtis Keller and D'nardo Colucci. Perception in HMDs : what is it in head-mounted displays (HMDs) that really make them all so terrible ? *Helmet-and Head-Mounted Displays III*, 3362(1) :46–53, 1998.
- [KFS03] Robert E. Kraut, Susan R. Fussell, and Jane Siegel. Visual information as a conversational resource in collaborative physical tasks. *Hum.-Comput. Interact.*, 18(1) :13–49, 2003.
- [KGF02] Robert E. Kraut, Darren Gergle, and Susan R. Fussell. The use of visual information in shared visual spaces : informing the development of virtual co-presence. In *CSCW '02 : Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 31–40, New York, NY, USA, 2002. ACM.

- [KL04] Joshua M. Knapp and Jack M. Loomis. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence : Teleoper. Virtual Environ.*, 13(5) :572–577, 2004.
- [KOF06] Adam D. I. Kramer, Lui Min Oh, and Susan R. Fussell. Using linguistic features to measure presence in computer-mediated communication. In *CHI '06 : Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 913–916, New York, NY, USA, 2006. ACM.
- [KS39] M. G. Kendall and B. Babington Smith. The problem of m rankings. In *Annals of mathematical statistics*, volume 10, pages 275–287, 1939.
- [Kva00] T. Kvan. Collaborative design : what is it ? In *Automation in Construction*, volume 9, pages 409–415, 2000.

L

- [LA08] Mark A. Livingston and Zhuming Ai. The effect of registration error on tracking distant augmented objects. In *ISMAR '08 : Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 77–86, Washington, DC, USA, 2008. IEEE Computer Society.
- [Lan05] Frédéric Landragin. Une caractérisation de la référence ostensive indirecte. In *Revue de Sémantique et de Pragmatique (RSP)*, volume 18. Presses Universitaires d'Orléans, 2005.
- [LB98] Guy Le Boterf. *L'Ingénierie Des Compétences*. édition d'Organisation, Paris, 1998.
- [LB00] Vincent Lepetit and Marie-Odile Berger. Handling Occlusion in Augmented Reality Systems : A Semi-Automatic Method. In *IEEE and ACM International Symposium on Augmented Reality*, pages 137 – 146, 2000.
- [Leg04] Pierre Legendre. Species Associations : the Kendall coefficient of concordance revisited. In *Journal of Agricultural, Biological, and Environmental Statistics*, volume 10, pages 226–245, 2004.
- [LF84] Richard S. Lazarus and Susan Folkman. *Stress, appraisal, and coping*. Springer, 1984.
- [LK81] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81 : Proceedings of the 7th international joint conference on Artificial intelligence*, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.
- [LMF04] Jeremy Legardeur, Christophe Merlo, and Isabelle Franchistéguy. *Methods and Tools for Cooperative and Integrated Design*, chapter Empirical Studies in Engineering Design and Health Institutions, pages pp. 385–396. KLUWER Academic Publishers, 2004.
- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60 :91–110, 2004.
- [LW02] Robert S. Laramee and Colin Ware. Rivalry and interference with a head-mounted display. *ACM Trans. Comput.-Hum. Interact.*, 9(3) :238–251, 2002.
- [LWAS07] Jane Li, Anja Wessels, Leila Alem, and Cara Stitzlein. Exploring interface with representation of gesture for remote collaboration. In *OZCHI '07 : Proceedings of the*

19th Australasian conference on Computer-Human Interaction, pages 179–182, New York, NY, USA, 2007. ACM.

M

- [Mac96] Wendy Mackay. *Réalité augmentée : le meilleur des deux mondes. La recherche, Special issue on L'ordinateur au doigt et à l'oeil.*, 284, 1996.
- [Mai07] Madjid Maldi. *Suivi Hybride en présence d'Occultations pour la Réalité Augmentée*. PhD thesis, Université d'Évry Val d'Essonne, Évry, 2007.
- [Mas54] Abraham Maslow. *Motivation and personality*. Harper, New York, 1954.
- [McN92] David McNeill. *Hand and mind : What gestures reveal about thought*. University of Chicago Press., 1992.
- [Mer03] Christophe Merlo. *Modélisation des connaissances en conduite de l'ingénierie – Mise en œuvre d'un environnement d'assistance aux acteurs de la conception*. PhD thesis, Université de Bordeaux 1., 2003.
- [MF96] Blair MacIntyre and Steven Feiner. Language-level support for exploratory programming of distributed virtual environments. In *UIST '96 : Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 83–94, New York, NY, USA, 1996. ACM.
- [MG93] D.A. Marx and R.C. Graeber. Human error in aircraft maintenance. Technical report, Boeing Commercial Airplane Group, Seattle, Washington, 1993.
- [MG08] Jacques Marsot and Virginie Govaere. Prévention des risques en maintenance industrielle : apports et limites des techniques de réalité augmentée. In *IHM '08 : Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine*, pages 51–54, New York, NY, USA, 2008. ACM.
- [MGDB05] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. Dart : a toolkit for rapid design exploration of augmented reality experiences. In *SIGGRAPH '05 : ACM SIGGRAPH 2005 Papers*, pages 932–932, New York, NY, USA, 2005. ACM.
- [MIMS05] Andrew P. Martin, Melody Y. Ivory, Rodrick Megraw, and Beverly Slabosky. Exploring the persistent problem of user assistance. Technical Report IS-TR-2005-08-01, Information School, University of Washington, 2005.
- [Mit93] G. Mitchell. What do users really want from computer documentation? In *International professional communication conference (IPCC)*, pages 27–31, Philadelphia PA, 1993.
- [MMC09] Pranav Mistry, Pattie Maes, and Liyan Chang. Wuw - wear ur world : a wearable gestural interface. In *CHI '09 : Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4111–4116, New York, NY, USA, 2009. ACM.
- [MMT⁺94] Paul Milgram, Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio" Kishino. Augmented reality : A class of displays on the reality-virtuality continuum. *SPIE : Telemanipulator and Telepresence Technologies*, 2351 :282–292, 1994.
- [Mol88] Abraham Moles. *Théorie structurale de la communication et société*. Technique et Scientifique des Télécommunications. 1988.

- [MR08] Malik Malleem and David Roussel. *Réalité augmentée : Principes, technologies et applications*. Techniques de l'ingénieur, 2008.
- [MS09] Erick Mendez and Dieter Schmalstieg. Importance masks for revealing occluded objects in augmented reality. In *Proceedings of the 2009 ACM Symposium on Virtual Reality Software and Technology (VRST 09)*, pages 247–248. ACM, 2009.

N

- [NF02] Leonid Naimark and Eric Foxlin. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In *ISMAR '02 : Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 27, Washington, DC, USA, 2002. IEEE Computer Society.
- [NM98] U. Neumann and A. Majoros. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *VRAIS '98 : Proceedings of the Virtual Reality Annual International Symposium*, page 4, Washington, DC, USA, 1998. IEEE Computer Society.
- [Nor88] Donald A. Norman. *The Psychology of Everyday Things*. ISBN : 0-465-06709-3. Basic Books, 1988.
- [NT95] Ikujiro Nonaka and Hirotaka Takeuchi. *The Knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, 1995.
- [NW06a] David G. Novick and Karen Ward. What users say they want in documentation. In *SIGDOC '06 : Proceedings of the 24th annual ACM international conference on Design of communication*, pages 84–91, New York, NY, USA, 2006. ACM.
- [NW06b] David G. Novick and Karen Ward. Why don't people read the manual? In *SIGDOC '06 : Proceedings of the 24th annual ACM international conference on Design of communication*, pages 11–18, New York, NY, USA, 2006. ACM.

O

- [OCMB95] Marc Olano, Jon Cohen, Mark Mine, and Gary Bishop. Combatting rendering latency. In *I3D '95 : Proceedings of the 1995 symposium on Interactive 3D graphics*, pages 19–ff., New York, NY, USA, 1995. ACM.
- [OFc+03] Jiazhi Ou, Susan R. Fussell, X. chen, Leslie D. Setlock, and Jie Yang. Gesture communication over video stream : Supporting multimodal interaction for remote collaborative physical tasks. In *ICMI*, 2003.
- [Oli10] Jérôme Olive. *Annotations gestuelles pour capitaliser les interactions en environnement virtuel informé : Application à la maintenance industrielle*. PhD thesis, Université de Technologie de Compiègne, 2010.

P

- [Pat02] Manoj Patankar. Root cause analysis of rule violations by aviation maintenance technicians. Technical Report Grant No. 2001-G-001 FAA Office of Aviation Medicine, Washington, DC, 2002.

- [PHMG06] Juri Platonov, Hauke Heibel, Peter Meier, and Bert Grollmann. A mobile markerless ar system for maintenance and repair. In *ISMAR '06 : Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 105–108, Washington, DC, USA, 2006. IEEE Computer Society.
- [Pin01] Claudio Pinhanez. Augmenting reality with projected interactive displays. In *Proceeding of International Symposium on Virtual and Augmented Architecture (VAA '01)*, 2001.
- [Pol58] Michael Polanyi. *Personal Knowledge : Towards a Post-Critical Philosophy*. The University of Chicago Press, 1958.
- [Pol66] Michael Polanyi. *The Tacit Dimension*. The University of Chicago Press, 1966.
- [PT03] Wayne Piekarski and Bruce H. Thomas. An object-oriented software architecture for 3d mixed reality applications. In *ISMAR '03 : Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 247, Washington, DC, USA, 2003. IEEE Computer Society.
- [Pue03] M. Puelo. it's not just summer reading anymore : building bridges through the year between school and public librarians. NMRLS Introduction to Dynamics Youth Library Services Workshop, 2003.

R

- [RA00] Jun Rekimoto and Yuji Ayatsuka. Cybercode : designing augmented reality environments with visual tags. In *DARE '00 : Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10, New York, NY, USA, 2000. ACM.
- [Ras86] Jens Rasmussen. *Information Processing and Human-Machine Interaction : An Approach to Cognitive Engineering*. Elsevier Science Inc., New York, NY, USA, 1986.
- [RBHL⁺05] Jannick P. Rolland, Frank Biocca, Felix Hamza-Lup, Yanggang Ha, and Ricardo Martins. Development of head-mounted projection displays for distributed, collaborative, augmented reality applications. *Presence : Teleoperators and Virtual Environments*, 14(5) :528–549, 2005.
- [RBLM97] Matthias Rauterberg, Martin Bichsel, Ulf Leonhardt, and M. Meier. Build-it : a computer vision-based interaction technique of a planning tool for construction and design. In *INTERACT '97 : Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction*, pages 587–588, London, UK, UK, 1997. Chapman & Hall, Ltd.
- [RD06] Gerhard Reitmayr and Tom Drummond. Going out : robust model-based tracking for outdoor augmented reality. In *ISMAR '06 : Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 109–118, Washington, DC, USA, 2006. IEEE Computer Society.
- [Rea93] James Reason. *L'erreur humaine*. Presses Universitaires de France, 1993.
- [Rea97] James Reason. *Managing the Risks of Organizational Accidents*. Ashgate Publishing Company, 1997.
- [Rek98] J. Rekimoto. Matrix : A realtime object identification and registration method for augmented reality. In *APCHI '98 : Proceedings of the Third Asian Pacific Computer and Human Interaction*, page 63, Washington, DC, USA, 1998. IEEE Computer Society.

- [Rey02] Claude Reyterou. *Une méthodologie pour l'intégration des Nouvelles Technologies de l'Information et de la Communication dans les processus de Maintenance Aéronautique*. PhD thesis, EADS, 2002.
- [RF00] Jannick P. Rolland and Henry Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence : Teleoper. Virtual Environ.*, 9(3) :287–309, 2000.
- [Ros04] Bertrand Rose. *Proposition d'un référentiel support à la conception collaborative : COMED (Collaborative CONflict Management in Engineering Design), Prototype logiciel dans le cadre du projet IPPOP*. PhD thesis, Université Henri Poincaré, Nancy France, 2004.
- [RVDWB00] Corrine Ribert-Van De Weerdts and Eric Brangier. L'usage et l'efficacité des aides à la maintenance en télédiffusion. In *Le travail humain*, volume 63, pages 331–352. Presses Universitaires de France, 2000.
- [RWC99] Ramesh Raskar, Greg Welch, and Wei-Chao Chen. Table-top spatially-augmented reality : Bringing physical models to life with projected imagery. In *IWAR '99 : Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, page 64, Washington, DC, USA, 1999. IEEE Computer Society.

S

- [SAH⁺01] Andrei State, Jeremy Ackerman, Gentaro Hirota, Joohee Lee, and Henry Fuchs. Dynamic virtual convergence for video see-through head-mounted displays : Maintaining maximum stereo overlap throughout a close-range work space. *ISAR*, 00 :137, 2001.
- [SD06] Gilles Simon and Julien Decollogne. *Intégrer images réelles et images 3D*. Dunod, 2006.
- [SdL03] Bernd Schwald and Blandine de Laval. An augmented reality system for training and assistance to maintenance in the industrial context. *11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2003 (WSCG)*, 11(1) :1213–6972, 2003.
- [SFC⁺01] B. Schwald, J. Figue, E. Chauvineau, F. Vu-Hong, A. Robert, M. Arbolino, M. Schneider, B. Laval, F. Dumas de Raully, F.G. Anez, O. Baldo, and J. Santos. STARMATE : Using augmented reality technology for computer guided maintenance of complex mechanical elements. *eBusiness and eWork Conference (e2001), Venice*, 2001.
- [SFH⁺02] Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavári, L. Miguel Encarnação, Michael Gervautz, and Werner Purgathofer. The studierstube augmented reality project. *Presence : Teleoperators and Virtual Environments*, 11(1) :33–54, 2002.
- [SFPG06] Sudipta N. Sinha, Jan-Michael Frahm, Marc Pollefeys, and Yakup Genc. Gpu-based video feature tracking and matching. *Workshop on Edge Computing Using New Commodity Architectures*, 2006.
- [SG92] Richard HY So and Michael J. Griffint. Compensating lags in head-coupled displays using head position prediction and image deflection. *Journal of Aircraft*, 9(6), 1992.
- [SG97] Richard HY So and Michael J. Griffint. Lag compensation by image deflection and prediction : a review on the potential benefits to virtual training applications for

manufacturing industry. *International conference on human-computer interactions*, 21B :997–1000, 1997.

- [SH02] Brett E. Shelton and Nicholas R. Hedley. Using augmented reality for teaching earth-sun relationships to undergraduate geography students. In *First IEEE International Augmented Reality Toolkit Workshop*, Darmstadt, Germany, 2002.
- [SH10] Orit Shaer and Eva Hornecker. Tangible user interfaces : Past, present, and future directions. *Found. Trends Hum.-Comput. Interact.*, 3(1–2) :1–137, 2010.
- [SKF05] Andrei State, Kurtis P. Keller, and Henry Fuchs. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *ISMAR '05 : Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 28–31, Washington, DC, USA, 2005. IEEE Computer Society.
- [SKK⁺03] Nobuchika Sakata, Takeshi Kurata, Takekazu Kato, Masakatsu Kourogi, and Hideaki Kuzuoka. WACL : Supporting telecommunications using wearable active camera with laser pointer. In *ISWC '03 : Proceedings of the 7th IEEE International Symposium on Wearable Computers*, page 53, Washington, DC, USA, 2003. IEEE Computer Society.
- [SKK06] Nobuchika Sakata, Takeshi Kurata, and Hideaki Kuzuoka. Visual assist with a laser pointer and wearable display for remote collaboration. *CollabTech06*, pages 66–71, 2006.
- [SLB08] Hartmut Seichter, Julian Looser, and Mark Billinghurst. Composar : An intuitive tool for authoring ar applications. In *ISMAR '08 : Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 177–178, Washington, DC, USA, 2008. IEEE Computer Society.
- [SO58] Fred C. Sabin and Kenneth N. Ogle. Accommodation-convergence association. In *AMA Arch. Ophthalmol.*, volume 59, pages 324–332, 1958.
- [SS01] Alex W. Stedmon and Robert J. Stone. Re-viewing reality : human factors of synthetic training environments. *International Journal of Human-Computer Studies*, 55(4) :675–698, 2001.
- [ST94] Jianbo Shi and Carlo Tomasi. Good features to track. In *Computer Vision and Pattern Recognition*, pages 593 – 600, Ithaca, NY, USA, 1994. IEEE Computer Society Conference, Cornell University.
- [Sut68] Ivan E. Sutherland. A head-mounted three dimensional display. In *AFIPS '68 (Fall, part I) : Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764, New York, NY, USA, 1968. ACM.
- [SWC93] J. Short, E. Williams, and B. Christie. Visual communication and social interaction. pages 153–164, 1993.

T

-
- [Tan91] John C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34(2) :143–160, 1991.
 - [Tar92] Jacques Tardif. *Pour un enseignement stratégique, l'apport de la psychologie*. Les Editions Logiques, Montréal, 1992.

- [Tho09] Indira Thouvenin. Interaction et connaissance : construction d'une expérience dans le monde virtuel. Mémoire d'HDR, Université de Technologie Compiègne, 2009.
- [TK91] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical Report Technical Report CMU-CS-91-132, International Journal of Computer Vision, 1991.
- [TOBM03] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. Comparative effectiveness of augmented reality in object assembly. In *CHI '03 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80, New York, NY, USA, 2003. ACM.

U

- [UKC92] UKCAA. Maintenance error. Technical report, Asia Pacific Air Safety (United Kingdom Civil Aviation Authority), 1992.
- [UKC09] UKCAA. Aircraft maintenance incident analysis. Reference : CAA paper 2007/04, (United Kingdom Civil Aviation Authority), July 2009.

W

- [Wel93] Pierre Wellner. Interacting with paper on the digitaldesk. *Communication. ACM*, 36(7) :87–96, 1993.
- [Wil91] P. Wilson. *Computer Supported Cooperative Work : An Introduction*. KLUWER Academic Publishers, 1991.
- [WKS03] Jens Weidenhausen, Christian Knoepfle, and Didier Stricker. Lessons learned on the way to industrial augmented reality applications, a retrospective on arvika. In *Computers and Graphics*, volume 27, pages 887–891, 2003.
- [Wlo95] Matthias M. Wloka. Lag in multiprocessor virtual reality. *Presence*, 4 :50–63, 1995.
- [WN03] Karen Ward and David G. Novick. Hands-free documentation. In *SIGDOC '03 : Proceedings of the 21st annual international conference on Documentation*, pages 147–154, New York, NY, USA, 2003. ACM.
- [WOSL03] Stefan Wiedenmaier, Olaf Oehme, Ludger Schmidt, and Holger Luczak. Augmented reality (ar) for assembly processes design and experimental evaluation. *International Journal of Human-Computer Interaction*, 16(3) :497 – 514, 2003.
- [WRSO07] Robin Wolff, Dave J. Roberts, Anthony Steed, and Oliver Otto. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology*, 29(1) :11–26, 2007.
- [WVS05] Harald Wuest, Florent Vial, and Didier Stricker. Adaptive line tracking with multiple hypotheses for augmented reality. In *ISMAR '05 : Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 62–69, Washington, DC, USA, 2005. IEEE Computer Society.

Z

- [ZAK05] Carolina Zingale, Vicki Ahlstrom, and Bonnie Kudrick. Human factors guidance for the use of handheld, portable, and wearable computing devices. Technical report, Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405, 2005.
- [ZBG02] Xiao Wei Zhong, Pierre Boulanger, and Nicolas D. Georganas. Collaborative augmented reality : A prototype for industrial training. *21th Biennial Symposium on Communication*, 2002.

TITLE

T.A.C. System : Collaborative Remote Assistance (**T**élé-**A**ssistance **C**ollaborative).

Augmented Reality and NICT for operators and experts in the context of a supervised industrial maintenance task.

ABSTRACT

T.A.C. system was built in order to propose a new mode of remote communication. When an operator needs to be assisted on a manual task, classical technical support methods as phone or visioconference rapidly show their limits in terms of human interaction. By allowing a distant expert to see exactly what an operator is seeing, he could interact with him in interactive time thanks to augmented reality via an adequate interaction paradigm we named « Picking Outlining Adding ». In guiding the operator through visual guidance information, expert removes the ambiguities of language and communications difficulties are avoided. Operations are then easily performed. The experiments we have conducted go in this direction, showing the effectiveness of this new mode of communication. Moreover, interactions principles are sufficiently general to be applied to other use cases that go beyond the originally planned industrial maintenance.

Key-Words : Tele-Assistance ; Collaboration ; Augmented Reality ; Maintenance.

AUTEUR : Sébastien BOTTECCHIA.

TITRE : Système T.A.C. : **T**élé-**A**ssistance **C**ollaborative. Réalité augmentée et NTIC au service des opérateurs et des experts dans le cadre d'une tâche de maintenance industrielle supervisée.

DIRECTEUR DE THÈSE : Jean-Pierre JESSEL.

CO-DIRECTEUR DE THÈSE : Jean-Marc CIEUTAT.

LIEU DE SOUTENANCE : ESTIA-RECHERCHE, Technopôle Izarbel, 64210 Bidart, FRANCE.

DATE DE SOUTENANCE : Le 20 septembre 2010.

RÉSUMÉ :

Le système T.A.C. a été conçu avec l'objectif de proposer un nouveau mode de communication distant. Lorsqu'un opérateur a besoin d'être assisté sur une tâche manuelle, les supports techniques classiques comme le téléphone ou la visioconférence montrent rapidement leurs limites en termes d'interaction humaine. En permettant à un expert distant de voir exactement ce que l'opérateur voit, il peut interagir avec lui en temps réel grâce à la réalité augmentée via un paradigme d'interaction adéquat que nous avons nommé « Picking Outlining Adding ». En guidant l'opérateur grâce à des indices virtuels, l'expert lève les ambiguïtés de langage et les difficultés de communication sont évitées. Les opérations sont ainsi plus facilement exécutées. Les expérimentations que nous avons menées vont dans ce sens, montrant l'efficacité de ce nouveau mode de communication. Qui plus est, les principes d'interactions sont suffisamment généraux pour dépasser le cadre original de la maintenance industrielle.

MOTS CLÉS : Télé-Assistance ; Collaboration ; Réalité Augmentée ; Maintenance.

DISCIPLINE : Informatique.

INTITULÉ ET ADRESSE DES LABORATOIRES :

Institut de Recherche en Informatique de Toulouse (IRIT), Université Paul Sabatier, 118 Route de Narbonne, 31062 TOULOUSE CEDEX 9, FRANCE.

École Supérieure des Technologies Industrielles Avancées (ESTIA), Technopôle Izarbel, 64210 Bidart, FRANCE.