



**HAL**  
open science

# Sharing big display : développement des technologies et métaphores d'interactions nouvelles pour le partage collaboratif d'affichage en groupe ouvert

Damien Marion

## ► To cite this version:

Damien Marion. Sharing big display : développement des technologies et métaphores d'interactions nouvelles pour le partage collaboratif d'affichage en groupe ouvert. Interface homme-machine [cs.HC]. Université de Bordeaux, 2017. Français. NNT : 2017BORD0924 . tel-01738673

**HAL Id: tel-01738673**

**<https://theses.hal.science/tel-01738673>**

Submitted on 20 Mar 2018

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

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

THÈSE PRÉSENTÉE  
POUR OBTENIR LE GRADE DE  
**DOCTEUR DE**  
**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR  
SPÉCIALITÉ COGNITIVE

Par Damien MARION

**Sharing Big Display : Développement des technologies et  
métaphores d'interactions nouvelles pour le partage collaboratif  
d'affichage en groupe ouvert.**

Sous la direction de : Jean-Marc ANDRE  
Responsable scientifique : Pierre-Alexandre FAVIER

Soutenance le 21 Décembre 2017

Membres du jury :

M. CLAVERIE, Bernard Professeur – Bordeaux INP, ENSC	Président
M. ESTRAILLIER, Pascal Professeur – Université de La Rochelle	Rapporteur
M. PARAPONARIS, Claude Professeur – CNRS	Rapporteur
Mme PINEDE, Nathalie Maître de conférences – Université Bordeaux Montaigne	Examineur
M. ANDRE, Jean-Marc Professeur – Bordeaux INP, ENSC	Directeur
M. FAVIER, Pierre-Alexandre Maître de conférences – Bordeaux INP, ENSC	Co-directeur
M. GARDINETTI, Emmanuel Responsable domaine scientifique Hommes-Systèmes – DGA	Invité



A mon grand-père parti trop tôt mais demeurant un exemple à suivre

A ma grand-mère pour sa bienveillance

A mes parents sans qui cela n'aurait jamais été possible

A Sophie

A Elodie



# SHARING BIG DISPLAY : DEVELOPPEMENT DES TECHNOLOGIES ET METAPHORES D'INTERACTIONS NOUVELLES POUR LE PARTAGE COLLABORATIF D'AFFICHAGE EN GROUPE OUVERT

## Résumé

Aujourd'hui, les écrans ont envahi nos quotidiens. Cette colonisation a engendré l'avènement de nouveaux écrans, les écrans de grand format. Or, ces grands écrans sont souvent utilisés comme de simples affichages statiques n'autorisant que très peu d'interaction aux utilisateurs en présence. Mais alors comment permettre des interactions multi-utilisateur et surtout quels modes et moyens d'interactions proposer afin d'interagir avec ces grands écrans ? Il faut évidemment définir des règles de collaboration : que faire si plusieurs personnes consultent la même information en même temps ? Comment permettre à un nouvel arrivant de prendre conscience de ce qu'il y a à voir alors que des utilisateurs ont déjà "transformé" l'affichage par leur utilisation courante ?

Face à ces enjeux, notre travail a consisté, dans un premier temps, à tester des interactions multi-utilisateurs sur grand écran à l'aide d'un dispositif de détection et de suivi de la position et de l'orientation de la tête (*head tracking*) instrumenté. Nous avons ainsi mis en évidence l'importance du concept de « vision flottante » dans la recherche collaborative d'informations sur grand écran. Dans un second temps, nos recherches se sont concentrées sur le développement d'un dispositif d'*head tracking* permettant des interactions multi-utilisateur (jusqu'à 6 personnes), sans artefacts et sans calibrage avec un écran de grand format. Enfin, nous présentons l'étude comparative d'un même affichage tantôt statique et tantôt rendu dynamique à l'aide de notre dispositif. Cette étude est basée sur l'évaluation de l'expérience utilisateur (UX). Ainsi, nous avons pu dégager les premières recommandations concernant les métaphores d'interactions à mettre en place pour interagir avec un écran de grand format dans un contexte d'usage en groupe ouvert.

Ce travail de doctorat a été consacré à la mise en avant de la faisabilité, de l'intérêt et de la proposition de premières orientations concernant la conception d'interactions multi-utilisateurs avec un écran de grand format basées sur une technique d'*head tracking*.

**Mots clés :** Interaction Hommes-Machines ; Interface Hommes-Systèmes ; Head Tracking ; Ecran de grand format ; Lieux Publics ; Expérience Utilisateur ; Facteur Humain.

---

# SHARING BIG DISPLAY: TECHNOLOGY DEVELOPMENT AND NEW INTERACTION METAPHORS FOR COLLABORATIVE SHARING DISPLAY IN OPEN GROUP.

## Abstract

Screens have invaded our daily life. Amongst them, large displays are becoming increasingly present in public places. Sometimes, a few interaction are proposed to the users, but most of the time they are simply used as static displays. So, how to interact with these large displays and especially how to allow multi-user interactions? Of course, we have to define rules for collaboration: what we have to do if several people consult the same information at the same time? How to allow a newcomer to become aware of what there is to see when users have already "transformed" the display by their current use?

First, our work consisted in testing instrumented multi-user interactions, based on a head tracking system, during a collaborative information seeking task, on a large display. In this part, we have highlighted the importance of the concept of "floating vision". Then, our research focused on the development of a head tracking system allowing intuitive interactions, without needing neither a special equipment nor individual calibration. Our system supports several users (up to six) simultaneously interacting with individualized information on a large display. Finally, we present the study of performance gain within a context of multi-user competitive consultation of information. We compare the benefit of an adaptive display (information move in front of users who are focusing on them), with a standard display. This study is based on user experience (UX) analysis. Thus, we were able to identify the first recommendations about interaction metaphors allowing intuitive interactions with a large display in an open group context.

This research was devoted to highlighting the feasibility, interest and proposition of initial orientations concerning the design of multi-user interactions with large public displays based on a head tracking technique.

**Keywords:** Human-Computer Interaction ; Human-System Interface ; Head Tracking ; Large Publics Displays ; User Experience ; Knowledge Exchange ; Human Factor.

**Equipe Cognitive & Ingénierie Humaine**

Laboratoire de l'Intégration du Matériau au Système  
IMS – UMR CNRS 5218  
Université de Bordeaux  
109 Avenue Roul, 33405 Talence.

# Remerciements

Maintenant que ce travail de longue haleine touche à sa fin, il est temps de remercier toutes les personnes qui y ont contribué. Je m'excuse d'avance de ne pouvoir citer tout le monde, mais je connais ce que je dois à chacun.

En premier lieu, je tiens à exprimer ma reconnaissance aux professeurs Pascal ESTRAILLIER et Claude PARAPONARIS pour l'honneur qu'ils m'ont fait en acceptant d'être rapporteur de mon travail et en participant à mon jury de thèse.

Je tiens également à remercier le professeur Bernard CLAVERIE d'avoir accepté la présidence du jury.

Je remercie sincèrement Nathalie PINEDE d'avoir accepté d'être examinatrice de mes travaux.

Je remercie la Direction Générale de l'Armement pour le financement accordé. Merci à Didier BAZALGETTE, sans qui cette thèse n'aurait pas vu le jour. Je tiens également à remercier Emmanuel GARDINETTI d'avoir accepté d'être présent pour ma soutenance, pour sa disponibilité et la relation privilégiée qu'il a établie sur la fin de mon doctorat.

Je souhaite exprimer une profonde gratitude à Jean-Marc ANDRE pour tous ces conseils avisés, ses relectures à la virgule, sa gentillesse et sa disponibilité infaillible même dans les périodes difficiles.

Une pensée particulière pour Coralie EYRAUD-DUBOIS, Véronique LESPINET-NAJIB, Catherine SEMAL et Jérôme SARACCO pour leurs conseils, leur aide et surtout leur patience.

Je remercie Edwige CLERMONT de m'avoir accordé sa confiance pour donner des cours dans son module.

Je ne remercierai jamais assez Quentin CHIBAUDEL, Théo MOULIERE-SEBAN et Baptiste PREBOT pour leur soutien sans faille, leur aide plus que précieuse et surtout pour les frontons endiablés de cette fin de rédaction qui m'ont permis de tenir et d'endurer cette épreuve.

Un grand merci à Amélie ROCHE, Audrey BONA, Christophe BEY et Sami LINI pour leur aide et les grandes discussions souvent scientifiques et parfois moins.

Je remercie l'ensemble de l'équipe Cognitive & Ingénierie Humaine pour leurs différentes contributions à mes travaux.

Je remercie également l'ensemble du personnel de l'Ecole Nationale Supérieure de Cognitive, notamment Isabelle SESE, Jean-Paul CAYERE, Nadège RODRIGUEZ, Denis PLOTTON et Nathalie PELLOQUIN pour l'aide apportée et pour leur gentillesse tout au long de mes travaux. Merci à Benoît LEBLANC de s'être occupé même à distance de mon travail de doctorat.

Je remercie l'ensemble du personnel du Laboratoire Intégration du Matériau au Système pour leur disponibilité et leur gentillesse.

Je remercie aussi l'ensemble des élèves de l'ENSC pour leur contribution à mes travaux et plus particulièrement Marwin HENRI, Théodore LETHOUZE, Paul MARTI et Lydie VAN HOEVELAKEN pour avoir accepté de me prêter leurs têtes à chaque fois que j'en avais besoin.

Enfin, je remercie également toutes les personnes ayant contribué à mes travaux en venant passer les expériences.

Dans un second temps, je souhaite présenter mes remerciements plus personnels. Ces remerciements vont aux personnes qui me sont très chères et qui ont participé, parfois même sans l'imaginer, à passer les épreuves souvent difficiles liées à ce travail de doctorat, à garder confiance et surtout à continuer d'avancer. Je ne peux exprimer mon entière gratitude avec les mots qui vont suivre, mais rien de plus normal que d'essayer :

Merci à mes parents d'avoir toujours été là, de m'avoir toujours permis de faire ce que j'aime, parfois à leurs dépens, et de sans cesse veiller à notre bonheur à ma sœur et moi.

Merci à ma sœur, je serai toujours là pour veiller sur toi.

Merci à toute ma famille, ma grand-mère, mes oncles et tantes, cousins et cousines, Zizine et papy, la famille HIXE, pour leur soutien indéfectible et leur compréhension sans bornes durant ces dernières années où je n'ai pu être présent autant que je l'aurais souhaité.

Merci à tous mes amis, les plus anciens comme les plus récents, pour tous les moments partagés où j'ai pu, grâce à vous, m'évader et profiter de la vie. Je ne peux tous vous citer tant j'ai la chance que vous soyez aussi nombreux, mais une pensée toute particulière va à chacun d'entre vous.

Merci à toi, Elodie, pour tout. Sans ton soutien et ton amour, je n'aurais certainement pas mené à bien ce doctorat. Ce travail est aussi le fruit de ton implication.

Enfin, je terminerai ces remerciements par Pierre-Alexandre FAVIER. Plus qu'un directeur de thèse, tu es devenu un véritable ami. Je te serai éternellement reconnaissant de m'avoir choisi et pris sous ton aile. Je ne te remercierai jamais assez pour toute l'aide apportée, la confiance que tu m'as accordée et pour tous les échanges scientifiques et personnels que nous avons partagés.

# Table des matières

Résumé.....	5
Abstract .....	6
Remerciements .....	7
Table des matières .....	9
Liste des figures .....	15
Table des Tableaux.....	17
Glossaire.....	19
Organisation de la thèse .....	21
PARTIE I - CONTEXTE THEORIQUE .....	25
Chapitre 1 : Introduction .....	27
Contexte .....	27
Cadre théorique .....	28
Objectif et contribution .....	36
Publications .....	37
Soumis :.....	37
Intérêt d'une démarche anthropocentrée .....	38
Chapitre 2 : Etat de l'art .....	41
L'attention .....	41
L'attention sélective .....	41
L'attention partagée.....	43
L'attention contrôlée & automatique .....	44
L'attention dans cette thèse .....	44
Etat de l'art sur les écrans de grand format.....	45
L'omniprésence des écrans .....	45
Notion de contexte : privé, semi-privé et public .....	45
Différents types d'écrans interactifs.....	45
Les grands écrans horizontaux .....	46
Les grands écrans verticaux .....	46
Type d'interaction avec les grands écrans : directes <i>versus</i> non directes .....	47
Avantages des écrans de grand format .....	47
L'augmentation de la taille de l'écran : quels impacts sur les utilisateurs ? .....	47
Les grands écrans catalyseurs des interactions sociales.....	49
Faiblesses des écrans de grand format .....	49
Absence de considération des écrans de grand format : Display Blindness .....	49
La surcharge informationnelle .....	49
Les préjugés du contexte .....	49
L'audience capturée .....	50
Fugacité de l'attention envers les grands écrans .....	50
Panneaux d'affichage physiques <i>versus</i> panneaux d'affichage digitaux .....	51
Embarras social .....	51
Espace personnel .....	51
Appréhension de l'inconnu .....	51
Trop d'espace .....	53
Recommandations générales pour la mise en œuvre des écrans de grand format .....	53
Positionnement des écrans .....	53
L'importance du contenu .....	53

Le format du contenu affiché : un atout majeur ? .....	53
Contenu dynamique en fonction du contexte .....	54
Mode attractif .....	55
L'importance de la navigation différenciée .....	55
L'acteur unique .....	55
Fiabilité et robustesse des systèmes techniques .....	55
Privilégier les habitudes et la spécialisation fonctionnelle pour la collaboration ....	56
Lignes directrices pour la conception d'affichage de grand format.....	56
Introduction aux méthodes de conception d'expérience utilisateur .....	56
Analyse heuristique .....	57
Les heuristiques spécialisées aux écrans de grand format .....	58
Modes et moyens d'interaction avec les écrans de grand format.....	59
Proposition de catégorisation des modes d'interactions avec les écrans de grand format .....	59
Le suivi du regard comme mode d'interaction.....	63
Suivi du regard : Interactions et paradigmes d'interface.....	65
Head Tracking versus Eye tracking .....	67
Paradigmes d'interface et écrans de grand format .....	68
Problèmes du Multi-utilisateur .....	68
Conserver la zone de confort des utilisateurs .....	68
Transférabilité des interfaces WIMP ?.....	68
Adaptations des interfaces WIMP aux grands écrans .....	69
Conclusion de l'état de l'art .....	74
PARTIE II - APPROCHE EMPIRIQUE .....	77
Chapitre 3 : Recherche collaborative d'informations sur écran de grand format .....	79
Introduction .....	79
Contexte .....	79
La vision flottante.....	80
Principe de l'expérience .....	81
Hypothèses .....	82
Protocole expérimental.....	84
Description Technique .....	84
Head tracking .....	84
Point d'attention .....	85
Programmes informatiques pour l'expérience .....	86
Déroulement de l'expérience .....	86
Recrutement .....	87
Indépendance et dépendance à l'égard du champ .....	87
Configuration physique .....	88
Consignes .....	89
L'expérience.....	89
Matériel .....	90
Participants .....	90
Résultats .....	91
Analyses des données.....	91
Modalité PA <i>versus</i> Modalité S .....	91
H1.1 : Nombre total d'objets trouvés (NTOT).....	91
H1.2 : Temps pour trouver le 1 <sup>er</sup> objet (T1) .....	92
H1.3 : Le nombre moyen d'objets trouvés par image (NbMO).....	93
H1.4 : Le nombre de mauvais clics (NbMC) .....	94

H1.5 : Le nombre de « tout juste » réalisés (NbTJ) .....	94
H1.6 : Le nombre de « zéro trouvé » réalisés (NbZT) .....	95
Synthèse : modalité PA vs modalité S .....	96
Ordre PA-S <i>versus</i> Ordre S-PA.....	97
Synthèse : ordre PA-S vs ordre S-PA.....	97
Discussion .....	102
L'effet des Séries d'images : Série1 <i>versus</i> Série2 .....	102
L'effet des modalités .....	103
L'effet d'ordre de la passation .....	103
Conclusions .....	104
Résumé.....	105
Chapitre 4 : Vers une interaction à multiples utilisateurs non appareillés et sans calibrage avec les écrans de grand format. ....	107
Introduction .....	107
Contexte .....	107
Les yeux et la tête : un système masse-ressort .....	108
Données brutes mesurées par le Kinect V2.....	110
Principe de l'expérience .....	110
Hypothèses .....	112
Protocole expérimental.....	113
Matériel .....	113
Méthode.....	114
Participants .....	117
Résultats .....	118
Analyse des données .....	118
Différence entre les angles de rotation théoriques calculés et les angles de rotation effectifs mesurés par le Kinect V2 .....	118
Précision du Kinect V2 suffisante pour distinguer les cas de consultation.....	120
Comparaison des groupes selon chaque cas de consultation pour la rotation en lacet en fonction des trois distances.....	120
Distance la plus proche de l'écran LS1 .....	120
Distance intermédiaire LS2.....	121
Distance la plus loin de l'écran LS3.....	122
Comparaison des groupes selon chaque cas de consultation pour la rotation en tangage en fonction des trois distances .....	123
Caractérisation des mouvements de la tête via la Kinect V2 .....	124
Comparaison des groupes cas par cas de consultation pour la rotation en lacet en fonction de chaque distance .....	124
Cas 0 : les groupes regardent en face d'eux .....	124
Cas 1 : les groupes regardent à plus une colonne écran par rapport à leur position .....	124
Cas 2 : les groupes regardent à plus deux colonnes écran par rapport à leur position.....	125
Cas 3 : les groupes regardent à plus trois colonnes écran par rapport à leur position.....	125
Cas 4 : les groupes regardent à plus quatre colonnes écran par rapport à leur position.....	125
Cas 5 : les groupes regardent à plus cinq colonnes écran par rapport à leur position.....	125

Comparaison des groupes cas par cas de consultation pour la rotation en tangage en fonction de chaque distance .....	125
Cas 1 : les groupes regardent la ligne supérieure de l'écran .....	126
Cas 2 : les groupes regardent la ligne intermédiaire de l'écran .....	126
Cas 3 : les groupes regardent la ligne inférieure de l'écran .....	126
Discussion .....	127
Le capteur Kinect V2 .....	127
Caractérisation des mouvements de la tête .....	127
Observation de différences significative entre les angles de rotation de la tête pour des cas de consultation différents .....	127
Absence de différence significative entre les angles de rotation de la tête pour les même cas de consultation .....	127
Conclusions .....	129
Résumé .....	129
Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via head tracking avec les écrans de grand format. ....	131
Introduction .....	131
Contexte .....	131
Identification des zones d'attention sur l'écran .....	132
Calcul de la zone d'attention sur l'écran .....	132
A l'aide d'un réseau de neurones artificiels .....	136
Neurone biologique .....	136
Neurone formel/artificiel .....	137
Perceptron multicouche avec rétro-propagation du gradient d'erreur .....	138
A l'aide de moyennes mobiles et des écarts moyens mesurés .....	139
Moyennes Mobiles .....	139
Procédure d'application des écarts moyens mesurés .....	140
En résumé .....	144
Interactions .....	146
Paradigme d'Interface n°1 (PI-1) : Affichage de l'image de la tête des utilisateurs dans leurs Zones Utilisateurs respectives .....	146
Paradigme d'Interface n°2 (PI-2) : Affichage des Zones d'Attention des utilisateurs détectés .....	146
Cas Interactif n°0 (CI-0) : Intégralité de la Zone d'Attention comprise dans une affiche .....	147
CI-0 et Zone utilisateur avec affiche : CI-0-0 .....	148
CI-0 et Zone utilisateur sans affiche : CI-0-1 .....	148
CI-0-2-X : Cas interactifs se déroulant après réalisation de CI-0-1-X .....	150
Principe de l'expérience .....	150
Le SUS et l'AttrakDiff .....	151
Hypothèses .....	152
Protocole expérimental .....	154
Matériel .....	154
Configuration physique .....	154
Méthode .....	154
Détails computationnels .....	154
Milieu d'implantation .....	154
Description détaillée des deux modalités d'affichage .....	154
Déroulement de l'expérience .....	155
Enregistrement des données .....	157

Participants .....	157
Résultats .....	158
Analyse des données .....	158
Comparaisons des modalités en termes de performance.....	158
Comparaison du nombre d'erreurs commises (NbErr) .....	158
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	158
Comparaison Modalité S versus Modalité D : nombre d'erreurs commises.....	158
Comparaison du nombre de bonnes réponses données (NbBonRep) .....	159
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	159
Comparaison Modalité S versus Modalité D : nombre de bonnes réponses données.....	159
Comparaison du temps écoulé pour réaliser les missions dans chaque modalité (TmpMiss).....	159
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	159
Comparaison Modalité S versus Modalité D des groupes pairs : temps écoulé pour réaliser les missions dans chaque modalité.....	160
Comparaison Modalité S versus Modalité D des groupes impairs : temps écoulé pour réaliser les missions dans chaque modalité.....	160
Comparaisons des modalités à l'aide du ressenti des utilisateurs .....	161
Comparaisons de la préférence d'utilisation d'une modalité par rapport à l'autre	161
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	161
Comparaison Modalité S versus Modalité D : préférences des utilisateurs.....	162
Comparaisons du ressenti de rapidité d'une modalité par rapport à l'autre.....	163
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	163
Comparaison Modalité S versus Modalité D : ressenti de rapidité .....	164
Comparaisons du sentiment d'efficacité des utilisateurs d'une modalité par rapport à l'autre .....	165
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	165
Comparaison Modalité S versus Modalité D pour les groupes Pairs : sentiment d'efficacité.....	166
Comparaison Modalité S versus Modalité D pour les groupes Impairs : sentiment d'efficacité.....	166
Comparaison de l'envie d'utilisation de la modalité D dans les lieux publics. ....	166
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	166
Comparaison pour ou contre l'utilisation de la Modalité D dans les lieux publics .....	167
Comparaison des notes attribuées à l'aide du questionnaire AttrakDiff.....	167
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	168
Comparaison des scores AttrakDiff obtenus par chaque modalité pour l'ordre de passation D-S.....	169
Comparaison des scores AttrakDiff obtenus par chaque modalité pour l'ordre de passation S-D.....	170
Comparaison des notes attribuées à l'aide du questionnaire SUS .....	171
Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs.....	171
Comparaison des scores SUS obtenus par chaque modalité pour l'ordre de passation D-S.....	172
Comparaison des scores SUS obtenus par chaque modalité pour l'ordre de passation S-D.....	173
Comparaisons des scores du SUS et de l'AttrakDiff .....	173
Comparaison des scores SUS et AttrakDiff de la modalité D.....	173

Pour l'ordre de passation D-S .....	173
Pour l'ordre de passation S-D .....	174
Comparaison des scores SUS et AttrakDiff de la modalité S .....	174
Pour l'ordre de passation D-S .....	174
Pour l'ordre de passation S-D .....	175
Discussion .....	176
Effet des modalités sur la performance de consultation.....	176
Ressenti des utilisateurs envers les modalités .....	176
Résultats de l'évaluation de la modalité D et de la modalité S .....	177
Comparaison SUS et AttrakDiff .....	178
Conclusions .....	179
Résumé.....	179
<b>PARTIE III – SYNTHÈSE &amp; PERSPECTIVES .....</b>	<b>181</b>
Chapitre 6 : Synthèse générale, limites et perspectives. ....	183
Synthèse générale.....	183
La zone d'attention un paradigme d'interface au service de la vision flottante.....	183
Vers une interaction multi-utilisateur, sans appareillage et sans calibrage avec les écrans de grand format. ....	184
Evaluation d'une interaction multi-utilisateur, sans appareillage et sans calibrage via <i>head tracking</i> avec les écrans de grand format. ....	185
Bibliographie.....	187
Annexes.....	203
Annexe I : Formulaire de Consentement libre, éclairé et exprès (première expérimentation).....	204
Annexe II : Spécification fonctionnelle du logiciel de passation (première expérience) .....	205
Ecran N°1 : .....	205
Ecran N°2A : .....	206
Ecran N°2B : .....	207
Ecran N°3 : .....	208
Ecran N°4 : .....	209
Ecran N°5 : .....	210
Ecran N°6 : .....	211
Ecran N°7 : .....	212
Ecran N°8 : .....	213
Ecran N°9 : .....	214
Ecran N°10 : .....	215
Ecran N°11 : .....	216
Autres Fonctionnalités.....	217
Le score du groupe : .....	217
Annexe III : Consignes affichées à l'écran (première expérience) .....	218
Annexe IV : Questionnaire <i>System Usability Scale (SUS)</i> .....	219
Annexe V : Questionnaire <i>AttrakDiff</i> .....	220
Annexe VI : Formulaire de Consentement Libre, Eclairé et Exprès (deuxième et troisième expérience) .....	221
Annexe VII : Fiche mission .....	222
Annexe VIII : Questionnaire Global .....	223

# Liste des figures

FIGURE 1 : SCHEMA GENERAL DE LA DEMARCHE MISE EN PLACE ET DE L'ORGANISATION DE LA THESE. ....	23
FIGURE 2 : CHAMP VISUEL HUMAIN, ISSUE DE (STANDARD 1999). ....	30
FIGURE 3 : LIMITE INDIQUANT LA POSSIBILITE DE METTRE EN PLACE DES INTERACTIONS VIA HEAD TRACKING EN FONCTION DE LA TAILLE DE L'ECRAN (DIAGONALE, RATIO D'AFFICHAGE EGAL A 1) ET DE LA DISTANCE D'INTERACTION EN CENTIMETRES. ....	31
FIGURE 4 : SEGMENTATION DE L'ECRAN EN FONCTION DU NOMBRE D'UTILISATEURS, JEU VIDEO MARIO KART DE NINTENDO. ....	32
FIGURE 5 : MUTUALISATION DE L'AFFICHAGE EN FONCTION DU NOMBRE D'UTILISATEURS, FIFA 15 D'EA SPORTS. LE ROND BLEU ET LE ROND ROUGE ONT ETE AJOUTES SUR L'IMAGE POUR AIDER A LA COMPREHENSION.....	32
FIGURE 6 : GOULOT D'ETRANGLEMENT DU CANAL UNIQUE, ADAPTE DE (BROADBENT 2013). ....	42
FIGURE 7 : MODELE DE L'ATTENTION DE BROADBENT, FILTRE PRECOCE, ADAPTE DE (BROADBENT 2013).....	42
FIGURE 8 : MODELE DE L'ATTENTION DE TREISMAN, FILTRE DISTRIBUE, ADAPTE DE (TREISMAN 1964). ....	43
FIGURE 9 : CONFIGURATION PHYSIQUE DE L'EXPERIENCE DE TAN ET AL, ISSUE DE (TAN, GERGLE ET AL. 2003). ....	48
FIGURE 10 : MODELE P3 PROPOSE PAR DILLON ET MORRIS (1996, 1999) .....	52
FIGURE 11 : CATEGORISATION DES MODES D'INTERACTION EN FONCTION DES TYPES D'INTERACTIONS. ....	60
FIGURE 12 : CATEGORISATION AFFINEE DES MODES D'INTERACTION EN FONCTION DES TYPES D'INTERACTIONS.....	61
FIGURE 13 : (A) ET (B) REPRESENTE LA REDUCTION DU NOMBRE DE PHASES D'INTERACTION VU PAR FITZMAURICE (FITZMAURICE ET BUXTON 1997). (C) REPRESENTE L'EXTENSION DU CONCEPT DE FITZMAURICE AUX INTERFACES TACTILES PAR TERRENGHI (TERRENGHI 2007). LA DERNIERE PARTIE (D) REPRESENTE NOTRE VISION DU CONCEPT DE REDUCTION DU NOMBRE DE PHASE D'INTERACTION APPLIQUE AUX INTERACTIONS ATTENTIONNELLES. ....	62
FIGURE 14 : STRUCTURE DE L'OEIL HUMAIN, ISSUE DE (MORIMOTO ET MIMICA 2005).....	65
FIGURE 15 : PROPOSITION DE DIFFERENTS COMPORTEMENTS DU CURSEUR LORS D'UN DEPLACEMENT, ISSUE DE (BAUDISCH, CUTRELL ET AL. 2003).....	69
FIGURE 16 : REPRESENTATION DES TECHNIQUES DRAG-AND-POP ET DRAG-AND-PICK, ISSUE DE (BAUDISCH, CUTRELL ET AL. 2003). ....	70
FIGURE 17 : REPRESENTATION DU PARADIGME D'INTERFACE DRAG-AND-THROW, ISSUE DE (COLLOMB, HASCOËT ET AL. 2005).....	71
FIGURE 18 : REPRESENTATION DU PARADIGME D'INTERACTION PUSH-AND-POP, ISSUE DE (COLLOMB, HASCOËT ET AL. 2005).....	71
FIGURE 19 : TECHNIQUES FLOWMENU, ISSUE DE (GUIMBRETIERE ET WINOGRAD 2000). (A) POUR ZOOMER L'UTILISATEUR DEPLACE LE STYLET DE LA ZONE CENTRALE VERS « OBJET... ». (B) UN SOUS-MENU (SURLIGNER, BOUGER, ZOOM) APPARAÏT ET LA PREMIERE CATEGORIE DU MENU (TEXTE...) EST GRISEE. L'UTILISATEUR CHOISIT « ZOOM » ET REVIENT VERS LA ZONE CENTRALE CE QUI FAIT APPARAÏTRE LA SELECTION DES NIVEAUX DE ZOOM, LE ZOOM ACTUEL EST DE 75%. (C) L'UTILISATEUR MONTE ET SELECTIONNE LA VALEUR 100%. (D) LE RETOUR DANS LA ZONE CENTRALE APPLIQUE LE ZOOM ET LA VALEUR 100% EST AFFICHEE AU CENTRE. ....	72
FIGURE 20 : ZOOMSCAPE TECHNIQUE, ISSUE DE (GUIMBRETIERE, STONE ET AL. 2001). DE (A) VERS (D) L'UTILISATEUR DEPLACE UNE GRILLE (3X3) CARRÉS VERS LE HAUT. LORSQU'UNE PARTIE DE LA GRILLE ARRIVE DANS LA PARTIE SUPERIEURE DE L'ECRAN, ELLE EST RETRECIE A 25%. ....	72
FIGURE 21 : TECHNIQUE SIGMALENS UTILISANT LE CONCEPT DE FOCUS PLUS CONTEXT, ISSUE DE (PIETRIGA ET APPERT 2008).....	73
FIGURE 22 : RELATIONS ENTRE LES PRINCIPAUX DETERMINANTS DE L'UX AVEC LES GRANDS ECRANS.....	74
FIGURE 23 : SCHEMA GENERAL DE CONCLUSION DE L'ETAT DE L'ART .....	76
FIGURE 24 : PROJECTION DU CHAMP VISUEL DANS LE PLAN ECRAN ET AFFICHAGE DE LA ZONE D'ATTENTION.....	81
FIGURE 25 : SCHEMA DES DEUX MODALITES COMPAREES DANS NOTRE PREMIERE EXPERIENCE : MODALITES PA ET S. ....	82
FIGURE 26 : OPTITRACK – V120:TRIO : TRAQUEUR OPTIQUE SIX DEGRES DE LIBERTE.....	84
FIGURE 27 : ENSEMBLE DE MARQUEURS CONSTITUANT UN RIGID BODY. ....	84
FIGURE 28 : EXEMPLE D'UN RIGID BODY COMPOSE DE 4 MARQUEURS DANS MOTIVE:TRACKER. ....	85
FIGURE 29 : CASQUE MUNI DE MARQUEURS. ....	85
FIGURE 30 : SCHEMA DECRIVANT LE DISPOSITIF D'HEAD TRACKING A L'AIDE DE LA TECHNOLOGIE OPTITRACK. ....	86
FIGURE 31 : EXEMPLE DE TACHES A COMPLETER ISSUES DU TEST EXTENDED CSA-WA (PETERSON ET DEARY 2006). (A) LA REPONSE EST « NON » CAR LES DEUX FIGURES NE SONT PAS LES MEMES. (B) LA REPONSE EST « OUI » CAR LA FIGURE DE GAUCHE EST PRESENTE DANS LA FIGURE DE DROITE. ....	88
FIGURE 32 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE NTOT. ....	98
FIGURE 33 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE T1. ....	98
FIGURE 34 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE NbMO. ....	99

FIGURE 35 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE NbMC.....	100
FIGURE 36 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE NbTJ.....	100
FIGURE 37 : BOX PLOTS – COMPARAISON ORDRE DE PASSATION PA-S VERSUS S-PA POUR LE CRITERE NbZT.....	101
FIGURE 38 : LES TROIS TYPES DE ROTATION DE LA TETE, D’APRES (BENSON ET JOBSON 1973). .....	109
FIGURE 39 : PHOTO DE L’INSTALLATION DE L’EXPERIENCE CONCERNANT LA CARACTERISATION DES MOUVEMENTS DE LA TETE LORS D’UNE TACHE DE CONSULTATION D’UN AFFICHAGE DE GRAND FORMAT A L’AIDE D’UN DISPOSITIF D’HEAD TRACKING.....	111
FIGURE 40 : MODELISATION DE L’INSTALLATION AVEC LES COTES EN METRE. ....	113
FIGURE 41 : BODY JOINTS ISSUS DU SDK DE LA KINECT V2. ....	114
FIGURE 42 : IMAGE AFFICHEE A L’ECRAN DECOUPEE EN 18 AFFICHES NUMEROTEES DE TAILLE EQUIVALENTES PRESENTANT UN TITRE ET SA DEFINITION.....	115
FIGURE 43 : REPARTITION DE LA POSITION DES GROUPES PAR RAPPORT A L’ECRAN DURANT L’EXPERIENCE. ....	116
FIGURE 44 : PLAN DE REFERENCE DES DONNEES FOURNIES PAR LA KINECT : PLAN KINECT. ....	133
FIGURE 45 : SCHEMA DU CHANGEMENT DU PLAN DE REFERENCE, DU PLAN KINECT VERS LE PLAN ECRAN. ....	133
FIGURE 46 : CALCUL DE LA ZONE D’ATTENTION SUR L’ECRAN VIA LES DONNEES KINECT V2 APRES ETALONNAGE. ....	134
FIGURE 47 : DESCRIPTION DES ANGLES DE ROTATION DE LA TETE. ....	135
FIGURE 48 : SCHEMA D’UN NEURONE BIOLOGIQUE.....	137
FIGURE 49 : SCHEMA D’UN NEURONE FORMEL .....	138
FIGURE 50 : DESCRIPTION DE LA CONFIGURATION PHYSIQUE DE L’EXPERIENCE. ....	140
FIGURE 51 : SCHEMA DE LA PROCEDURE D’APPLICATION DES ECARTS MOYENS MESURES.....	144
FIGURE 52 : PROCEDURE COMPLETE DU CALCUL DE LA ZONE D’ATTENTION PERMETTANT UNE INTERACTION AVEC LES ECRANS DE GRAND FORMAT. ....	145
FIGURE 53 : DESCRIPTION DES ZONES UTILISATEUR A L’ECRAN. ....	146
FIGURE 54 : PHOTO ZONES D’ATTENTION SUR L’ECRAN.....	147
FIGURE 55 : ATTENUATION DE L’OPACITE DE LA ZONE D’ATTENTION. (A) 30% D’OPACITE, (B) 20% D’OPACITE, (C) 10% D’OPACITE ET (D) 0% D’OPACITE. ....	148
FIGURE 56 : CI-0-1-1 : ANIMATION DE DEPLACEMENT D’UNE AFFICHE AVEC LA ZONE D’ATTENTION. ....	149
FIGURE 57 : CI-0-1-2 : COPIE INSTANTANEE DE L’AFFICHE PRESENTE DANS LA ZONE UTILISATEUR D’UN AUTRE, AVEC SA ZONE D’ATTENTION. ....	149
FIGURE 58 : CI-0-2-1 : APPARITION DE L’AFFICHE AVEC DANS LA ZONE UTILISATEUR A L’AIDE DE LA ZONE D’ATTENTION. ....	150
FIGURE 59 : COMPOSITION DE L’AFFICHAGE.....	151
FIGURE 60 : TABLEAUX RECAPITULATIFS DES MISSIONS POUR CHAQUE UTILISATEUR DANS CHACUNE DES MODALITES. ....	156
FIGURE 61 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES DE PREFERENCES DES UTILISATEURS ENVERS UNE MODALITE (D, S OU N) EN FONCTION DE L’ORDRE DE PASSATION. ....	162
FIGURE 62 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES DE PREFERENCES ENVERS UNE MODALITE (D, S OU N) DE L’ENSEMBLE DES UTILISATEURS. ....	163
FIGURE 63 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES DE RESSENTI DE RAPIDITE DES UTILISATEURS VIS-A-VIS D’UNE MODALITE (D, S OU N) EN FONCTION DE L’ORDRE DE PASSATION. ....	164
FIGURE 64 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES POUR LE RESSENTI DE RAPIDITE VIS-A-VIS UNE MODALITE (D, S OU N) DE L’ENSEMBLE DES UTILISATEURS.....	164
FIGURE 65 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES DU SENTIMENT D’EFFICACITE DES UTILISATEURS VIS-A-VIS D’UNE MODALITE (D, S OU N) EN FONCTION DE L’ORDRE DE PASSATION. ....	165
FIGURE 66 : HISTOGRAMME PRESENTANT LA REPARTITION DES NOTES INDIQUANT L’ENVIE D’UTILISATION DE LA MODALITE D EN FONCTION DE L’ORDRE DE PASSATION. ....	167
FIGURE 67 : EXEMPLE D’ITEM (PAIRE DE MOTS) A REMPLIR UTILISE DANS L’ATTRAKDIFF. ....	168
FIGURE 68 : HISTOGRAMME DES VALEURS MOYENNES OBTENUES POUR CHAQUE SOUS-ECHELLE PAR CHAQUE MODALITE .....	178

# Table des Tableaux

TABLEAU 1 : ECHELLE DE DIAGONALE D'ÉCRAN EN FONCTION DU CHAMP DE VISUEL DE CONFORT ET DE LA DISTANCE, ISSUE DE (TERRENGHI, QUIGLEY ET AL. 2009).	28
TABLEAU 2 : DESCRIPTION DE LA POPULATION ET DE LA REPARTITION SELON L'ORDRE DE PASSATION DES MODALITES.	90
TABLEAU 3 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE NTOT.	91
TABLEAU 4 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE NTOT.	91
TABLEAU 5 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE T1.	92
TABLEAU 6 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE T1.	92
TABLEAU 7 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE NbMO.	93
TABLEAU 8 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE NbMO.	93
TABLEAU 9 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE NbMC.	94
TABLEAU 10 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE NbMC.	94
TABLEAU 11: STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE NbTJ.	95
TABLEAU 12 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE NbTJ.	95
TABLEAU 13 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER SERIE1 VS SERIE2 VIS-A-VIS DU CRITERE NbZT.	95
TABLEAU 14 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER PA VS S VIS-A-VIS DU CRITERE NbZT.	96
TABLEAU 15 : SYNTHESE DES RESULTATS DE COMPARAISON ENTRE LES SERIES D'IMAGES (SERIE1 VS SERIE2) ET LES MODALITES (PA VS S).	96
TABLEAU 16 : SYNTHESE DES RESULTATS DE COMPARAISON ENTRE LES ORDRES DE PASSATION DE L'EXPERIENCE (PA-S VS S-PA).	97
TABLEAU 17 : DESCRIPTION DE LA POPULATION ET DE LA REPARTITION DES GROUPES.	117
TABLEAU 18 : COMPARAISON DES ANGLES DE ROTATION EN LACET THEORIQUES ET DES ANGLES DE ROTATION EN LACET EFFECTIFS MESURES PAR LE KINECT V2 EN DEGRES.	119
TABLEAU 19 : COMPARAISON DES ANGLES DE ROTATION EN TANGAGE THEORIQUES ET DES ANGLES DE ROTATION EN TANGAGE EFFECTIFS MESURES PAR LE KINECT V2 EN DEGRES.	119
TABLEAU 20 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE LA PLUS PROCHE DE L'ÉCRAN (LS1).	120
TABLEAU 21 : P-VALUES APRES PROCEDURE DE STEEL-DWASS-CRITCHLOW-FLIGNER POUR LA COMPARAISON DES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE INTERMEDIAIRE (LS1).	121
TABLEAU 22 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE INTERMEDIAIRE (LS2).	121
TABLEAU 23 : P-VALUES APRES PROCEDURE DE STEEL-DWASS-CRITCHLOW-FLIGNER POUR LA COMPARAISON DES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE INTERMEDIAIRE (LS2).	122
TABLEAU 24 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE LA PLUS LOIN DE L'ÉCRAN (LS3).	122
TABLEAU 25 : P-VALUES APRES PROCEDURE DE STEEL-DWASS-CRITCHLOW-FLIGNER POUR LA COMPARAISON DES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN LACET (YAW) POUR LA DISTANCE LA PLUS LOIN DE L'ÉCRAN (LS3).	123
TABLEAU 26 : TABLEAU RECAPITULATIF DES P-VALUES APRES UN TEST DE TUKEY POUR LA COMPARAISON DES CAS DE CONSULTATION PAR RAPPORT A LA ROTATION EN TANGAGE (PITCH) ET SELON LES 3 DISTANCES (LS1, LS2 ET LS3).	123
TABLEAU 27 : RESULTATS APRES TEST DE TUKEY COMPARANT LES GROUPES REGARDANT A PLUS UNE COLONNE ECRAN (CE) VERS LA DROITE.	124
TABLEAU 28 : RESULTATS APRES TEST DE TUKEY COMPARANT LES GROUPES REGARDANT A PLUS QUATRE COLONNES ECRAN (CE) VERS LA DROITE.	125
TABLEAU 29 : MOYENNES DE LA ROTATION EN LACET (YAW), EXPRIMEES EN DEGRES, EN FONCTION DE LA DISTANCE ET DU CAS DE CONSULTATION.	141
TABLEAU 30 : VALEURS MOYENNES DES SEUILS (S <sub>1</sub> -S <sub>2</sub> ET S <sub>2</sub> -S <sub>3</sub> ) ENTRE LES LIGNES ECRAN (LE), EXPRIMEES EN PIXELS, EN FONCTION DE LA DISTANCE A L'ÉCRAN ET DU CAS DE CONSULTATION.	142
TABLEAU 31 : REPARTITION DES TITRES D'AFFICHE EN MODALITE D.	155
TABLEAU 32 : REPARTITION DES TITRES D'AFFICHE EN MODALITE S.	155
TABLEAU 33 : DESCRIPTION DE LA POPULATION ET DE LA REPARTITION DES GROUPES.	157

TABLEAU 34 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE NOMBRE D'ERREURS COMMISES EN FONCTION DE L'ORDRE DE PASSATION DE L'EXPERIENCE.....	158
TABLEAU 35 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE NOMBRE D'ERREURS COMMISES SELON LA MODALITE.....	158
TABLEAU 36 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE NOMBRE DE BONNES REPONSES EN FONCTION DE L'ORDRE DE PASSATION DE L'EXPERIENCE.....	159
TABLEAU 37 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE NOMBRE DE BONNES REPONSES DONNEES SELON LA MODALITE.....	159
TABLEAU 38 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE TEMPS ECOULE (EN MINUTES) POUR REALISER LES MISSIONS DANS CHAQUE MODALITE EN FONCTION DE L'ORDRE DE PASSATION DE L'EXPERIENCE.....	160
TABLEAU 39 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE TEMPS ECOULE (EN MINUTES) POUR REALISER LES MISSIONS DANS CHAQUE MODALITE EN FONCTION DES MODALITES POUR L'ORDRE DE PASSATION S-D.....	160
TABLEAU 40 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE TEMPS ECOULE (EN MINUTES) POUR REALISER LES MISSIONS DANS CHAQUE MODALITE EN FONCTION DES MODALITES POUR L'ORDRE DE PASSATION D-S.....	160
TABLEAU 41 : STATISTIQUES DESCRIPTIVES DE LA REPARTITION DE LA PREFERENCE DES UTILISATEURS POUR CHAQUE MODALITE.....	161
TABLEAU 42 : STATISTIQUES DESCRIPTIVES DE LA REPARTITION DU RESSENTI DE RAPIDITE DES UTILISATEURS POUR CHAQUE MODALITE.....	163
TABLEAU 43 : STATISTIQUES DESCRIPTIVES DE LA REPARTITION DU SENTIMENT D'EFFICACITE DES UTILISATEURS POUR CHAQUE MODALITE.....	165
TABLEAU 44 : STATISTIQUES DESCRIPTIVES DE LA REPARTITION DE L'ENVIE D'UTILISATION DE LA MODALITE D DANS LES LIEUX PUBLICS EN FONCTION DE L'ORDRE DE PASSATION.....	166
TABLEAU 45 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES DES SOUS-ECHELLES DE L'ATTRAKDIFF POUR LA MODALITE D EN FONCTION DE L'ORDRE DE PASSATION (P ET I).....	168
TABLEAU 46 : TABLEAU RECAPITULATIF DES P-VALUES ASSOCIEES AUX COMPARAISONS DE CHAQUE SCORE DES SOUS-ECHELLES POUR LA MODALITE D EN FONCTION DE L'ORDRE DE PASSATION.....	168
TABLEAU 47 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES DES SOUS-ECHELLES DE L'ATTRAKDIFF POUR LA MODALITE S EN FONCTION DE L'ORDRE DE PASSATION (P ET I).....	169
TABLEAU 48 : TABLEAU RECAPITULATIF DES P-VALUES ASSOCIEES AUX COMPARAISONS DE CHAQUE SCORE DES SOUS-ECHELLES POUR LA MODALITE S EN FONCTION DE L'ORDRE DE PASSATION.....	169
TABLEAU 49 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES DES SOUS-ECHELLES DE L'ATTRAKDIFF ENTRE LA MODALITE D ET LA MODALITE S POUR L'ORDRE DE PASSATION D-S.....	170
TABLEAU 50 : TABLEAU RECAPITULATIF DES P-VALUES ASSOCIEES AUX COMPARAISONS DE CHAQUE SCORE DES SOUS-ECHELLES ENTRE LA MODALITE S ET LA MODALITE D POUR L'ORDRE DE PASSATION D-S.....	170
TABLEAU 51 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES DES SOUS-ECHELLES DE L'ATTRAKDIFF ENTRE LA MODALITE D ET LA MODALITE S POUR L'ORDRE DE PASSATION S-D.....	171
TABLEAU 52 : TABLEAU RECAPITULATIF DES P-VALUES ASSOCIEES AUX COMPARAISONS DE CHAQUE SCORE DES SOUS-ECHELLES ENTRE LA MODALITE S ET LA MODALITE D POUR L'ORDRE DE PASSATION S-D.....	171
TABLEAU 53 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE SCORE GLOBAL DU SUS POUR LA MODALITE D EN FONCTION DE L'ORDRE DE PASSATION (P ET I).....	172
TABLEAU 54 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LE SCORE GLOBAL DU SUS POUR LA MODALITE S EN FONCTION DE L'ORDRE DE PASSATION (P ET I).....	172
TABLEAU 55 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES GLOBAUX DU SUS ENTRE LA MODALITE D ET LA MODALITE S POUR L'ORDRE DE PASSATION D-S.....	172
TABLEAU 56 : STATISTIQUES DESCRIPTIVES DES DONNEES ETUDIEES POUR COMPARER LES SCORES GLOBAUX DU SUS ENTRE LA MODALITE D ET LA MODALITE S POUR L'ORDRE DE PASSATION S-D.....	173
TABLEAU 57 : STATISTIQUES DESCRIPTIVES DE LA COMPARAISON DES SCORES SUS ET ATTRAKDIFF OBTENUS POUR LA MODALITE D POUR L'ORDRE DE PASSATION D-S.....	174
TABLEAU 58 : STATISTIQUES DESCRIPTIVES DE LA COMPARAISON DES SCORES SUS ET ATTRAKDIFF OBTENUS POUR LA MODALITE D POUR L'ORDRE DE PASSATION S-D.....	174
TABLEAU 59 : STATISTIQUES DESCRIPTIVES DE LA COMPARAISON DES SCORES SUS ET ATTRAKDIFF OBTENUS POUR LA MODALITE S POUR L'ORDRE DE PASSATION D-S.....	174
TABLEAU 60 : STATISTIQUES DESCRIPTIVES DE LA COMPARAISON DES SCORES SUS ET ATTRAKDIFF OBTENUS POUR LA MODALITE S POUR L'ORDRE DE PASSATION S-D.....	175

# Glossaire

<b>CCU</b>	Conception Centrée Utilisateur
<b>ENSC</b>	Ecole Nationale Supérieure de Cognitique
<b>GUI</b>	Graphical User Interface : interface graphique
<b>HT</b>	Head Tracking : détection de la position et de l'orientation de la tête
<b>IHM</b>	Interface Homme-Machine
<b>IHS</b>	Interface Homme-Système
<b>SDK</b>	Software Development Kit : kit de développement logiciel
<b>UX</b>	User eXperience : Expérience Utilisateur
<b>WIMP</b>	Windows, Icons, Menus, Pointing device : fenêtres, icônes, menus, dispositif de pointage



# Organisation de la thèse

Nous proposons de préciser le contenu de chaque chapitre afin d'aider le lecteur à comprendre l'organisation du document.

## Chapitre 1 : Introduction

Au cours de ce chapitre nous aborderons dans un premier temps la problématique de l'interaction en groupe ouvert avec les écrans de grand format dans une société submergée par les nouvelles technologies. Puis, nous présentons une classification des écrans en fonction de leur taille ce qui permet de déterminer, pour une distance donnée, le nombre d'utilisateurs potentiels et la possibilité d'utiliser une interaction via *head tracking*. Nous décrivons ensuite les premiers questionnements concernant le partage d'affichage unique en groupe ouvert et les objectifs et contributions qui en ont découlé. Enfin, nous présentons la nécessité d'appliquer une démarche anthropocentrée dans un contexte d'interactions Hommes-Systèmes et d'analyse des futurs usages.

## Chapitre 2 : Etat de l'art

Ce chapitre est divisé en deux parties. La première partie est consacrée au concept de l'attention. La deuxième partie de cet état de l'art porte sur les interactions avec les écrans de grand format. Elle nous permet d'en pointer les avantages et inconvénients et de présenter notre choix des modes d'interaction et des paradigmes d'interface viables pour nos cas d'usages. Nous présentons également les grandes lignes directrices en matière de design d'expérience utilisateurs.

## Chapitre 3 : Recherche collaborative d'information sur écran de grand format

Dans ce chapitre nous présenterons en détail notre première expérience. Cette expérience réalisée avec une technique d'*head tracking* nécessitant un appareillage des utilisateurs nous a permis de confirmer la mise en application du concept de « vision flottante » proposé. Elle a également participé à la révélation de l'intérêt d'afficher à l'écran la « zone regardée » de chaque utilisateur. Ainsi l'expérience montre que la matérialisation de la « zone regardée » améliore la performance du groupe d'utilisateurs dans un contexte de recherches coopératives d'informations.

## Chapitre 4 : Vers une interaction multi-utilisateur, non appareillée et sans calibrage avec les écrans de grand format.

Au cours de ce chapitre, nous réalisons la description de la deuxième expérience. Elle vise à caractériser et qualifier les mouvements de la tête des utilisateurs lors de la consultation d'un écran de grand format à l'aide d'une Kinect V2. Les données récoltées nous ont permis de développer un prototype autorisant des interactions multi-utilisateurs, sans appareillage invasif et ne nécessitant aucun calibrage préalable à son utilisation.

## Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via *head tracking* avec les écrans de grand format.

Dans ce chapitre, nous exposons la dernière expérimentation qui a consisté en l'étude et l'analyse de l'expérience utilisateur du prototype découlant de la deuxième expérience. Elle

montre l'intérêt de la dynamisation de l'affichage basée sur la détection de la position et de l'orientation de la tête. Elle présente également la comparaison des résultats d'évaluation d'interface obtenus à l'aide des questionnaires *System Usability Scale* et *AttrakDiff* dans leur version française. Enfin, cette expérimentation appuie l'intérêt de notre démarche en soulignant l'accueil positif des utilisateurs pour notre dispositif.

*Chapitre 6 : Synthèse générale, limites et perspectives.*

A travers ce chapitre 6, nous concluons ce mémoire de thèse par une discussion générale autour des différents résultats obtenus. Nous poursuivrons par la présentation d'une synthèse des résultats visant à conclure, dans une dernière partie, sur les perspectives envisageables et évolutions futures de ce travail.

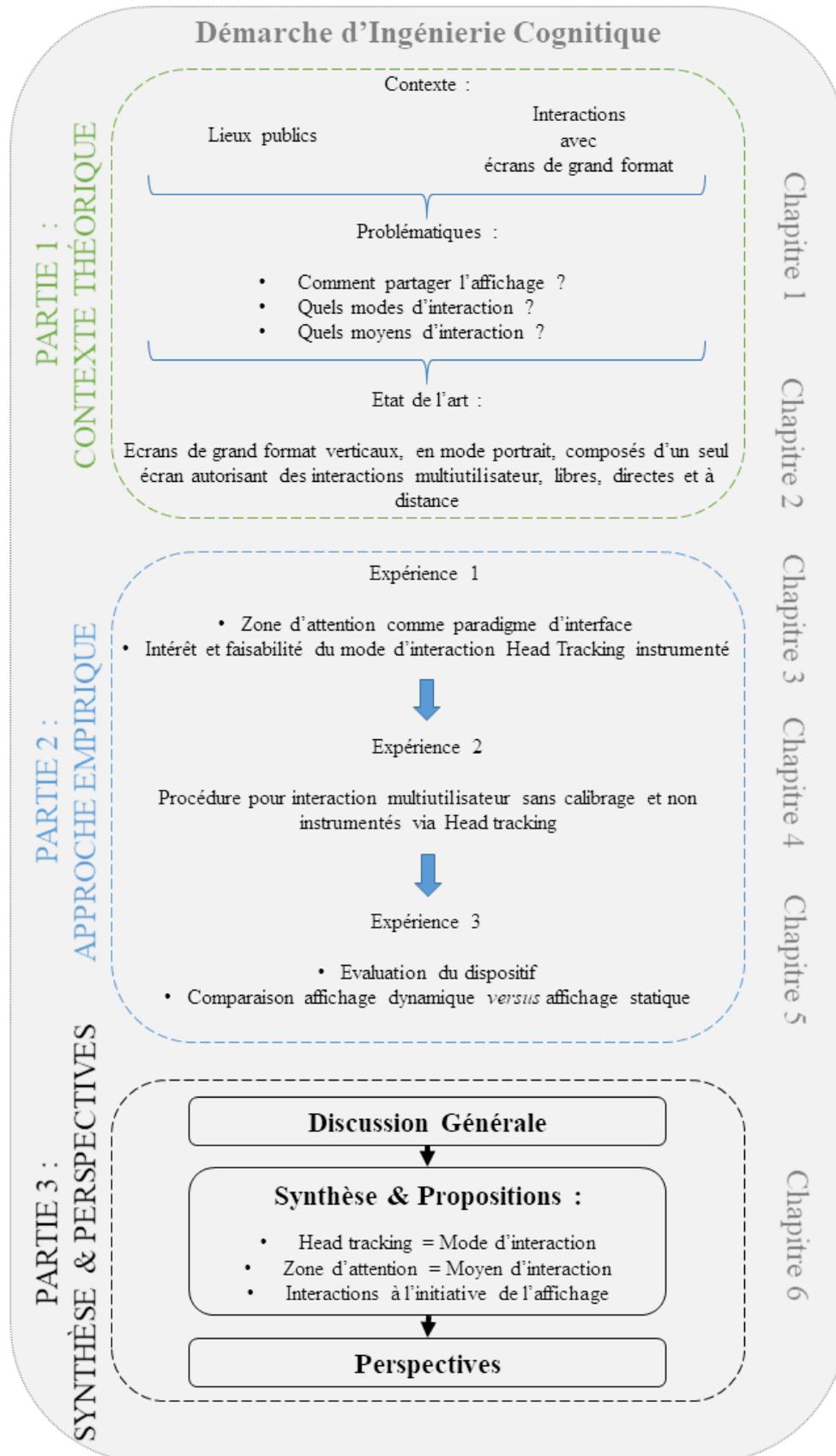


Figure 1 : Schéma général de la démarche mise en place et de l'organisation de la thèse.



# PARTIE I - CONTEXTE

## THEORIQUE

---

*« La plus avancée des technologies ne change pas l'être humain. Par conséquent, c'est dans l'analyse de son comportement face aux technologies et de l'usage qu'il en fait que réside les innovations de demain. »*

---



# Chapitre 1 : Introduction

## Contexte

Nous vivons l'époque de la technologie dite "pervasive" : les dispositifs informatiques, même très simples et très réduits, sont omniprésents dans nos activités quotidiennes. Ces nouvelles technologies sont à l'origine du développement d'applications toujours plus sophistiquées (avatars, réalité virtuelle et augmentée immersive, *serious games*) et dotées d'un énorme potentiel visant la facilitation et l'amélioration de la qualité et l'efficacité des interactions sociales. Après avoir mené une course à la puissance, à l'intégration, à la miniaturisation et à la basse consommation, depuis quelques années les industriels et les chercheurs ont pris de nouvelles orientations. Le nouveau combat est celui de l'utilisabilité et de l'acceptabilité, autrement dit celui des interfaces simples et intuitives. Le défi majeur dans la conception de ces interfaces réside dans l'implémentation de programmes capables de créer des situations interactionnelles adaptées à différents contextes (interactions au travail, en situation d'apprentissage ou encore applications à la clinique). Les inventions les plus récentes et les plus marquantes pour le grand public répondent à ces exigences, bien qu'elles ne soient pas nécessairement les plus complexes technologiquement. Tout le monde retient par exemple l'émergence du gestuel peu instrumenté (*WiiMote* de Nintendo) ou non instrumenté (*Kinect* de Microsoft), ou encore l'avènement du tactile et des capteurs d'orientation/accélération (*iPhone*, *iPod* et *iPad* de Apple).

Ces succès commerciaux ont un revers : les utilisateurs deviennent de plus en plus réfractaires à l'utilisation de moyens "classiques" d'interaction. Il est frustrant et rebutant d'apprendre à se servir d'une interface complexe quand on a eu l'expérience d'une telle intuitivité.

Dans cette évolution bien connue de tous, l'affichage est resté pendant longtemps le parent pauvre de cette course au progrès. Pourtant, les dispositifs d'affichage sont l'apanage de l'être humain. En effet, ils nous entourent depuis toujours, que ce soient les dessins sur les murs des grottes préhistoriques, les tableaux d'art, les photographies exposées dans nos maisons, les tableaux à l'école, la signalisation sur les routes, les portes de nos réfrigérateurs, les panneaux d'affichage publicitaires, les télévisions, etc., l'humain s'est toujours servi des surfaces qui l'entourent comme moyen d'affichage. Ces moyens d'affichage sont désormais de plus en plus digitalisés (Funkhouser et Li 2000). Néanmoins, les définitions et les dimensions des surfaces d'affichage digitales ont mis de nombreuses années à progresser, surtout si nous comparons cette progression à celle de la puissance des microprocesseurs par exemple.

La démocratisation récente d'écrans de très grands formats et de la "haute définition" change également le rapport de l'utilisateur à l'écran. Combinée aux dispositifs d'interactions nouveaux (tactile multipoints, capteur spatial capacitif, etc.), l'interaction la plus directe possible avec l'image affichée est au centre des attentes de l'utilisateur (Ishii et Ullmer 1997). Mais les modes d'interactions et les paradigmes d'interface utilisés jusque-là avec les ordinateurs ne s'exportent pas aussi bien qu'il n'y paraît sur les surfaces d'affichage de grand format (Robertson, Czerwinski et al. 2005).

# Cadre théorique

Nos travaux s'intéressent au partage d'une surface d'affichage unique en groupe ouvert d'utilisateurs non instrumentés dans un contexte de relations sociales hétérogènes entre les individus. Nous nous plaçons donc dans le cas d'une surface d'affichage assez importante pour être consultée par plusieurs personnes simultanément et ayant des relations variées entre elles (groupe d'amis, inconnus, départs et arrivées d'utilisateurs, *etc.*). Le nombre de personnes consultant la surface d'affichage fluctue en fonction des départs et arrivées des individus. La surface d'affichage doit réorganiser son contenu en fonction de l'intérêt et de la position des utilisateurs en présence. Les utilisateurs ne doivent pas être contraints de porter ou d'utiliser un artefact<sup>1</sup> pour interagir avec la surface d'affichage. Les interactions doivent être simples et naturelles.

De ce cas d'utilisation émergent de nombreux questionnements :

- Comment organiser, déplacer, enrichir, appauvrir, *etc.* les informations affichées sur l'écran sans impacter les conditions de consultation de chaque individu ?
- Comment permettre des interactions entre les utilisateurs et l'écran ?
- Quelles métaphores d'interactions mettre en œuvre pour faciliter l'eXpérience Utilisateur (UX) ?
- Quelles techniques mettre en œuvre pour éviter d'instrumenter les utilisateurs ?

De manière générale, comment rendre accessible l'information à l'écran dans les meilleures conditions de confort et de performance (temps de consultation, qualité de consultation, *etc.*) possibles pour l'ensemble des utilisateurs présents ?

Commençons par préciser la notion d'écran de grand format. Dans leur article, Terrenghi et al. (Terrenghi, Quigley et al. 2009) proposent une extension de la classification des écrans en fonction de leur taille (diagonale : pouce/pied/verge anglaise) de Weiser (Weiser 1991), basée sur des unités de mesure archaïques (Perche anglaise = 5,5 verges anglaises et Chaîne anglaise = 22 verges anglaises). Le tableau ci-dessous présente la relation entre la taille de l'écran (diagonale) et le champ visuel de confort à une distance donnée :

*Tableau 1 : Echelle de diagonale d'écran en fonction du champ de visuel de confort et de la distance, issue de (Terrenghi, Quigley et al. 2009).*

Echelle	Exemple	Taille d'écran	Distance	Champ visuel
<b>Pouce</b>	Smartphone	3 cm	40 cm	4°
<b>Pied</b>	Ordinateur portable	35 cm	70 cm	28°
<b>Verge</b>	Télévision	1 m	3 m	19°
<b>Verge</b>	Table interactive	1 m	1 m	53°
<b>Perche</b>	Grand écran	5 m	10 m	28°
<b>Chaîne</b>	Très grand écran	20 m	50 m	23°

<sup>1</sup> Définition Larousse 2017 : En anthropologie, produit ayant subi une transformation, même minime, par l'homme, et qui se distingue ainsi d'un autre provoqué par un phénomène naturel.

Ils poursuivent leur analyse en présentant des situations d'usage en relation avec la taille des écrans (échelle). Pour eux, un écran de la catégorie d'échelle « Pied » autorise une à deux personnes à consulter l'écran, un écran de la catégorie d'échelle « Verge » autorise entre 3 et 9 personnes à le consulter et un écran de la catégorie d'échelle « Perche » autorise entre 10 et 100 personnes à le consulter. Le contexte d'usage retenu pour notre étude étant celui des lieux publics, nous considérons en accord avec la nomenclature de Terrenghi et al. (2009) qu'un écran et de grand format s'il entre dans la catégorie d'échelle « Perche » et nous nous plaçons dans cette catégorie pour nos travaux.

Le champ visuel est une caractéristique importante du système visuel humain. Spector (1990) le définit ainsi : « le champ visuel est la portion de l'espace dans laquelle les objets sont visibles au même moment durant une fixation régulière du regard dans une direction. » (Spector 1990). Karwowski et Peacock (Karwowski et Peacock 1993) déterminent que dans le plan horizontal la limite de rotation acceptable des yeux est de 30° de part et d'autre de la ligne du regard et dans le plan vertical cette limite est de 15°. Soit un champ visuel binoculaire sans rotation de la tête de 60° dans le plan horizontal et de 30° dans le plan vertical (Karwowski et Peacock 1993, Tilley 2002, Van der Zanden 2016). Le département de la défense américaine (Standard 1999) décrit ces limites du champ visuel à 70° dans le plan horizontal, 40° vers le haut et 20° vers le bas dans le plan vertical. Il considère cependant une zone optimale de 30° pour le plan horizontal et le plan vertical comme le montre la *Figure 2*.

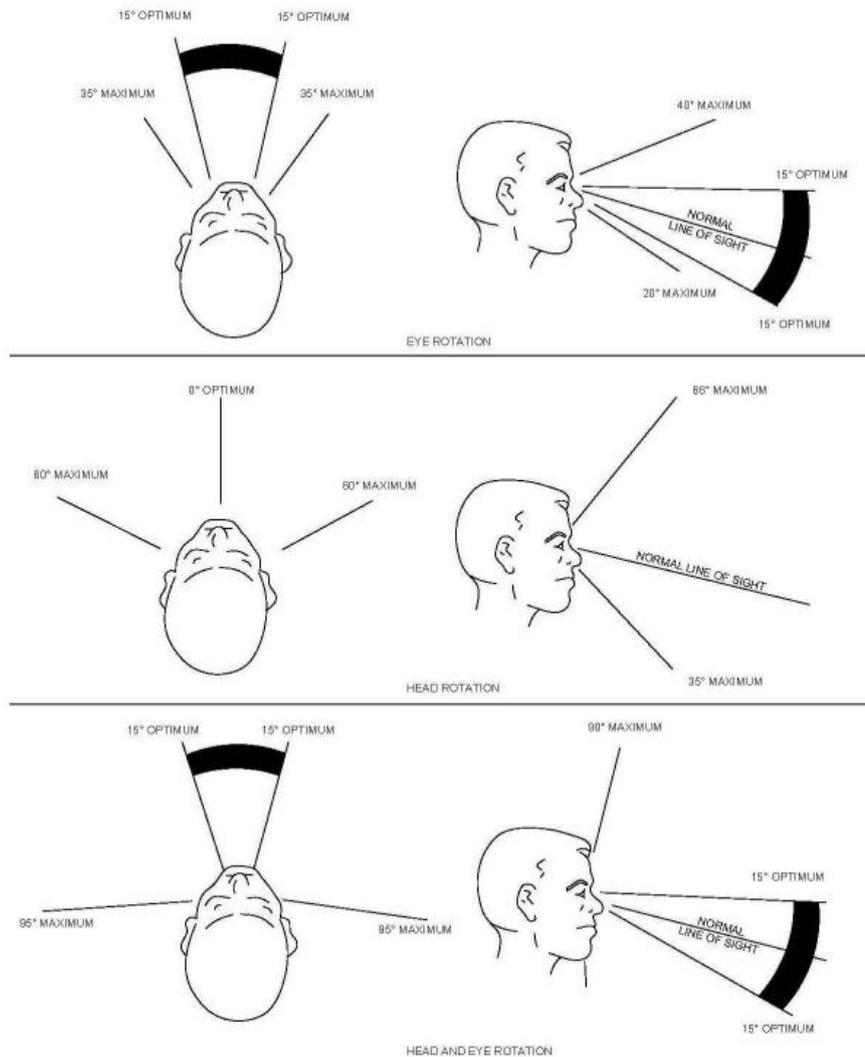
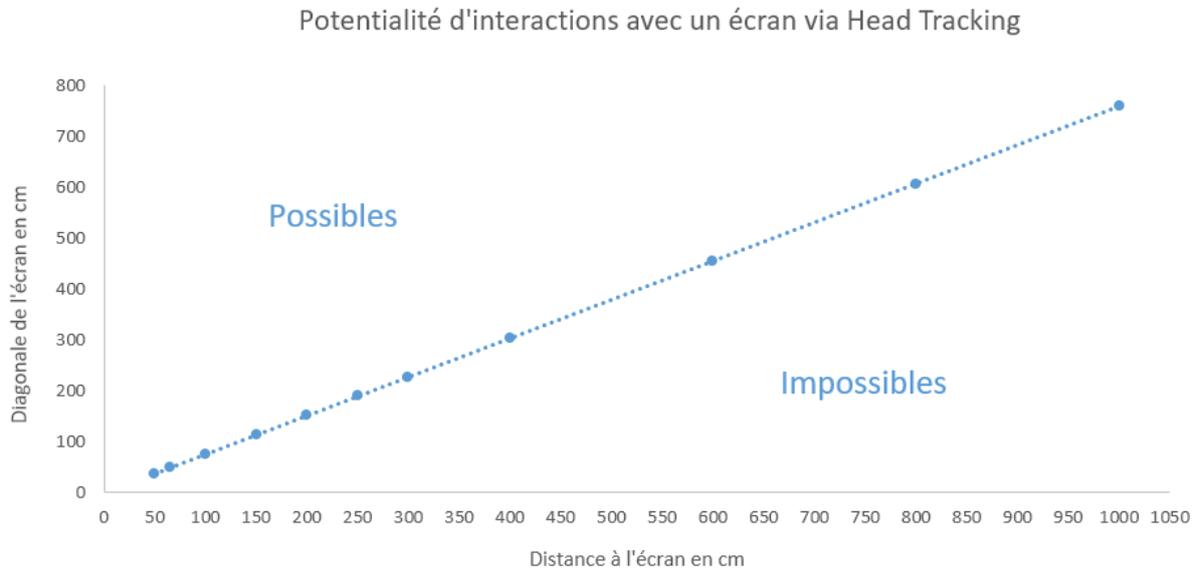


Figure 2 : Champ visuel humain, issue de (Standard 1999).

Plusieurs auteurs mettent en évidence que la distance préférée des utilisateurs lors de l'utilisation d'un écran pour une tâche de travail est comprise dans un intervalle allant de 40 à 90 centimètres avec une moyenne de 65 centimètres (Brown et Schaum 1980, Grandjean, Nishiyama et al. 1982). Ces recommandations prennent en compte, à l'aide de calculs trigonométriques, pour un écran disposé à 65 centimètres d'un individu et pour des valeurs d'angles de  $30^\circ$  de champ visuel (valeurs optimales) dans les deux plans, le champ visuel de « confort » sans mouvements de la tête est compris dans un écran de 49,3 centimètres de diagonale, soit un écran d'un peu plus de 19 pouces.

Nos recherches s'articulent autour d'une technique d'interaction basée sur la détection et le suivi de la position et de l'orientation de la tête (*head tracking*) afin d'identifier la zone visée par la tête sur l'écran. Nous nommons la zone visée par la tête sur l'écran, la zone d'attention. C'est cette zone d'attention qui va alors servir pour interagir avec l'écran de grand format. Du fait de la taille importante des écrans de grand format, nous effectuons mécaniquement des mouvements de tête pour consulter l'ensemble des informations qu'ils affichent. Cette mobilité nous permet de consulter les informations affichées dans des conditions de confort optimales. C'est pourquoi, nous nous sommes particulièrement intéressés à l'*head tracking* comme support de l'interaction entre des utilisateurs et un écran de grand format. Nous justifions l'intérêt de ce choix dans le *Chapitre 2 : Etat de l'art*

Nous considérons donc un écran comme grand à partir du moment où cet écran est assez important pour nécessiter un déplacement de la tête de l'utilisateur quand il consulte l'intégralité du contenu affiché. Comme vu précédemment, en considérant qu'un humain bouge sa tête pour consulter une information au-delà de 30° de son champ de vision (axe vertical et horizontal), la *Figure 3* ci-dessous présente la taille minimale de diagonale d'écran nécessaire en fonction de la distance à l'écran pour lesquelles un utilisateur doit bouger la tête pour consulter de manière confortable l'ensemble de l'écran.



*Figure 3 : Limite indiquant la possibilité de mettre en place des interactions via head tracking en fonction de la taille de l'écran (diagonale, ratio d'affichage égal à 1) et de la distance d'interaction en centimètres.*

Si l'écran est assez grand pour susciter des mouvements de la tête de l'utilisateur pour consulter l'intégralité de son contenu, nous pouvons mettre en place une technique permettant d'interagir avec celui-ci à base d'une détection de ces mouvements. Dans nos travaux, un écran de grand format est donc un écran imposant des mouvements de la tête (permettant la mise en place d'interaction via *head tracking*) afin de consulter l'intégralité de l'affichage à une distance donnée.

La notion de « partage » de la surface d'affichage unique peut être abordée selon deux angles différents. Selon le dictionnaire français Larousse, partager c'est :

- « Diviser quelque chose en plusieurs éléments distincts : Partager le gâteau en six parts. » ;

ou

- « Posséder quelque chose avec une ou plusieurs personnes : Partager le pouvoir. ».

La première définition considère le partage comme une division de l'objet possédé entre les individus tandis que la deuxième le considère comme une unité « possédée » par plusieurs individus. Cette problématique est abordée par le domaine des jeux vidéo en voici deux exemples. La *Figure 4*, montre une surface d'affichage divisée en autant de zones que d'individus en présence (ici quatre) nous parlerons alors de « partage compétitif ».



Figure 4 : Segmentation de l'écran en fonction du nombre d'utilisateurs, jeu vidéo Mario Kart de Nintendo.

La Figure 5 propose une surface d'affichage mutualisée pour l'ensemble des individus que l'on nommera « partage collaboratif » d'affichage.



Figure 5 : Mutualisation de l'affichage en fonction du nombre d'utilisateurs, FIFA 15 d'EA Sports. Le rond bleu et le rond rouge ont été ajoutés sur l'image pour aider à la compréhension.

La littérature montre que cette problématique est très souvent adressée en partageant la surface de manière “autoritaire” : l'affichage est segmenté en autant de zones qu'il y a d'utilisateurs (Wilmore 2004), comme dans la Figure 4. Ceci pose problème pour la gestion d'un groupe dynamique d'utilisateurs : comment gérer le départ ou l'arrivée d'un utilisateur pour redistribuer la ressource ? Cette approche borne aussi les interactions possibles entre les utilisateurs. Le partage explicite d'informations passe par l'utilisation de commandes prévues et donc figées (Kennedy 2008). De plus cette segmentation de la surface d'affichage en une

multitude de zones correspondant au nombre d'utilisateurs en présence ne devient plus envisageable si un grand nombre d'utilisateurs sont présents. La mutualisation de l'affichage devient alors incontournable, mais comment redistribuer l'affichage si un utilisateur vient à changer brusquement de point d'intérêt ? La dynamisation de l'affichage à l'aide d'interactions est donc indispensable.

Les solutions existantes consistent généralement en une préemption de la totalité de l'affichage par l'utilisateur "actif", celui-ci étant déterminé par un dispositif de contrôle individuel (terminal, téléphone portable, *etc.*)(Buckley et Duhig 2005, Scheible et Ojala 2005). Ces différentes solutions requièrent d'instrumenter les utilisateurs pour autoriser des interactions avec l'affichage.

Dans le cas de partage collaboratif d'affichage, les actions d'un utilisateur peuvent venir perturber la « vue » d'un autre. Ce type d'affichage est utilisé dans le cas où un objet est le point commun de l'attention des utilisateurs, dans notre exemple le ballon. Néanmoins, si un des utilisateurs souhaite se détacher de cette vue centrée sur le ballon pour voir son gardien de but par exemple, il ne le peut pas. Ce mode de partage d'affichage implique donc que les utilisateurs aient un but commun.

Etymologiquement, collaborer est emprunté au latin *collaborare*, qui se décompose en *con-* (« avec ») et *laborare* (« travailler »), littéralement « travailler avec ». Selon Grosz, une activité collaborative se définit par des actions menées par plusieurs personnes mais régies par un but commun (Grosz 1996). Dans nos travaux, nous aborderons donc la notion de partage collaboratif de l'affichage lorsque les utilisateurs devront réaliser des actions sur l'écran (parfois différentes) régies par un but commun. Nous parlerons de partage collaboratif et compétitif comme des modes d'affichages. Ces deux modes de partage d'affichage peuvent être agrégés dans un seul comme nous le montrerons dans le *Chapitre 5* de ce document.

Une autre approche rencontrée dans la communauté scientifique consiste en un partage de l'écran, mais pas de l'affichage : des dispositifs permettant d'afficher plusieurs images, discriminées selon l'angle sous lequel on regarde l'écran sont en plein développement (Schwerdtner 2008, Urey, Chellappan et al. 2011). De nombreux brevets, assez « génériques » dans leurs définitions, déposés par des industriels majeurs (Microsoft, Google, *etc.*) laissent deviner le grand enjeu de ce questionnement.

Au-delà des problèmes de partage de l'affichage se posent les modes et moyens d'interactions avec les écrans de grand format. Le terme « interaction » est composé du préfixe latin *inter-*, entre, et de action, du latin *actio*, faculté d'agir, activité, action, acte, fait, accomplissement, dérivé du verbe *agere*, agir, faire. L'interaction est donc l'action ou l'influence réciproque qui peut s'exercer entre deux ou plusieurs objets, corps, phénomènes ou systèmes physiques. Selon Marc et Picard (Marc et Picard 2002) « une interaction est un échange d'information, d'affects ou d'énergie entre deux agents au sein d'un système. C'est une action réciproque qui suppose l'entrée en contact de sujets ». Ainsi, les interactions définissent les modes et moyens mis en œuvre afin que les utilisateurs puissent contrôler et communiquer avec l'écran. Les moyens d'interaction sont les outils permettant d'interagir (la souris, le clavier, la tête, la main, la voix, *etc.*). Les modes d'interaction sont la manière dont s'opère l'interaction (gestuelle, par pointage, tactile, analyse vocale, *etc.*). Nous parlerons ici des interactions Homme-Machine qui constituent un domaine à part entière de la Recherche.

Différents types d'interaction existent dans la littérature, allant de l'utilisation d'un ou de plusieurs ordinateurs déportés, du tactile (stylets et tactiles multipoints)(Guimbretière, Stone et al. 2001, Chouvardas, Miliou et al. 2008), de pointeurs laser (Olsen Jr et Nielsen 2001, Davis et Chen 2002, Myers, Bhatnagar et al. 2002), de la voix (Hannun, Case et al. 2014), ou encore d'interactions gestuelles (mouvements de la tête, des yeux, de la main, du corps, *etc.*)(Jacob et Karn 2003, Vogel et Balakrishnan 2005, Duchowski 2007, Lee et Hollerer

2007, Lee, Green et al. 2008, Mistry et Maes 2009, Hilliges, Kim et al. 2012, Schiavo, Mencarini et al. 2013, Markussen, Jakobsen et al. 2014, Piumsomboon, Altimira et al. 2014). Lorsqu'un système interactif implique plusieurs moyens et modes d'interaction, l'interaction est dite multimodale car elle met en jeu plusieurs modalités sensorielles et motrices humaines. Dans notre contexte d'usage que sont les lieux publics, nous cherchons à nous abstraire d'une instrumentation des utilisateurs. Par exemple, les technologies tactiles multipoints permettent de s'abstraire de cette contrainte d'instrumentation des utilisateurs et sont ancrées dans les habitudes et les usages du grand public. Les utilisateurs peuvent alors interagir en même temps et directement avec leurs doigts avec la surface d'affichage. Le principal frein est la limitation du nombre d'utilisateurs pouvant interagir. En effet, cette solution implique que les utilisateurs soient proches de la surface d'affichage ce qui entraîne l'obstruction de celle-ci pour les utilisateurs plus éloignés (Vogel et Balakrishnan 2004). Dans le cas où les utilisateurs peuvent utiliser des terminaux déportés pour interagir avec la surface d'affichage, évitant ainsi de masquer celle-ci, le nombre de terminaux déployés détermine le nombre d'utilisateurs pouvant interagir avec l'affichage (Shen, Everitt et al. 2003, Shen, Vernier et al. 2004, Kennedy 2008). D'autres techniques autorisent des interactions non appareillées des utilisateurs comme les technologies à base de commandes vocales. Néanmoins, celles-ci ne nous semblent pas adaptées à une utilisation dans un milieu bruyant comme les lieux publics. C'est pourquoi, dans nos travaux, nous nous sommes plus particulièrement intéressés à la famille des interactions gestuelles.

A partir de ces interactions, il convient ensuite de concevoir les paradigmes d'interface. Thomas Kuhn (Kuhn 1972), qui introduit le mot paradigme dans son livre intitulé *La structure des révolutions scientifiques* le définit ainsi : « nous entendons par paradigme une sorte de métathéorie, un cadre de pensée, à l'intérieur duquel un consensus est réuni pour définir les questions pertinentes qui orientent les expériences à faire, et qui définissent la science « normale », jusqu'à ce qu'un changement intervienne qui, plus qu'une théorie, est un changement total de perspective ». Un paradigme est donc une représentation du monde, une manière de voir les choses, un modèle cohérent de vision du monde qui repose sur une base définie. L'interface quant à elle est définie par le dictionnaire Larousse comme : « Plan ou surface de discontinuité formant une frontière commune à deux domaines aux propriétés différentes et unis par des rapports d'échanges et d'interaction réciproques. ». Par extension, les paradigmes d'interface sont donc les modèles cohérents définissant les rapports d'échanges entre les utilisateurs et l'écran. Alan Cooper (Cooper 1995) distingue trois grands paradigmes d'interface :

- Le paradigme technologique : l'interface exprime les termes de sa propre construction, la manière dont elle a été construite. L'utilisateur doit comprendre comment fonctionne le système dans le but de l'utiliser avec succès. Exemple : la robotique industrielle
- Le paradigme métaphorique : le comportement de l'interface mime celui d'un objet de la vie courante et déjà maîtrisé par l'utilisateur. Exemple la notion de bureau en informatique. C'est par analogie que l'utilisateur comprend comment utiliser le système.
- Le paradigme idiomatique : le comportement de l'interface est stéréotypé, cohérent et donc simple à apprendre mais pas nécessairement calqué sur des objets de la vie. L'utilisateur doit donc apprendre (en imitant, en écoutant, en observant ou en reproduisant des séquences) des comportements simples. Par exemple le symbole « \_ » situé en haut à droite d'une fenêtre sur un ordinateur permet de réduire cette fenêtre.

Le terme « métaphores d'interaction » utilisé dans l'intitulé de ce travail de doctorat fait donc référence au paradigme métaphorique de Cooper. Néanmoins, Cooper ajoute : « Dans le domaine du design d'interface-utilisateur, le futur sera idiomatique, suite logique de la capacité normale des êtres humains à apprendre facilement et rapidement, tant que nous ne les forçons pas à comprendre le comment et le pourquoi. Il y a une infinité d'idiomes à inventer, plutôt qu'un éventail de métaphores à exploiter. Les métaphores semblent d'abord être un gain pour les utilisateurs débutants mais elles se montrent lourdes de conséquences lorsque l'on progresse dans l'utilisation approfondie d'un logiciel. Il est donc préférable de concevoir les choses de manière idiomatique, en utilisant la métaphore de manière occasionnelle, lorsque l'une d'elles nous tombe sous la main. ». En accord avec la vision de l'auteur, nous avons concentré notre effort sur l'utilisation mixte de ces deux paradigmes. Nous emploierons alors le terme de conception de familles d'interaction dans le reste de ce document pour désigner le panachage des paradigmes utilisés dans la conception d'interface.

## Objectif et contribution

L'objectif de ce projet est d'utiliser un dispositif de type « *head tracking* » (suivi automatique de la position et de l'orientation de la tête) pour estimer la zone de focalisation des utilisateurs sur un écran de grande taille. Le propos n'est pas d'obtenir une information de la zone regardée aussi précise que par un dispositif de type « *eye tracking* » (suivi automatique de la position et de l'orientation des yeux), mais d'être capable d'identifier la zone d'intérêt d'un utilisateur (Ashdown, Oka et al. 2005).

En fonction des zones d'intérêt identifiées, le but est de permettre une réorganisation dynamique des informations affichées pour amener au maximum ces zones d'intérêt dans les zones de confort visuel des utilisateurs. C'est donc les informations qui se réorganisent en fonction des utilisateurs et non l'inverse, l'initiative du déplacement est bien du côté de l'information. L'avantage est alors de pouvoir utiliser un dispositif non intrusif (caméra vidéo), si possible sans calibrage, pour une exploitation collaborative de l'affichage.

Un exemple d'application d'une telle démarche dans notre quotidien serait la proposition d'un affichage interactif des trains au départ dans une gare. Si plusieurs utilisateurs le consultent simultanément, il peut être intéressant que les informations qui concernent chacun (heure de départ, numéro du quai, *etc.*) "migrent" spontanément en face de lui. Il faut évidemment définir des règles de collaboration : que faire si plusieurs personnes consultent la même information ? Comment permettre à un nouvel arrivant de prendre conscience de ce qu'il y a à voir alors que des utilisateurs ont déjà "transformé" l'affichage par leur utilisation courante ?

Notre but n'est pas de développer une telle technologie. Nous nous intéressons aux usages qui peuvent en être faits avec une approche d'ingénierie des facteurs humains, afin d'être force de proposition pour 1) aider à spécifier et développer de telles technologies et 2) ouvrir les champs d'applications et proposer une réflexion avancée sur les usages avant même que ces technologies ne soient matures.

Les principales contributions de notre travail peuvent être résumées comme suit :

- étude de l'art des différentes études existantes sur les écrans de grand format ;
- une expérimentation confirmant l'intérêt du concept proposé de « vision flottante » dans la recherche collaborative d'informations sur grand écran ;
- la caractérisation des mouvements de la tête des individus lors de la consultation d'un écran de grand format à l'aide du capteur Kinect V2 de Microsoft ;
- la proposition et l'étude de métaphores d'interactions multi-utilisateurs non instrumentées lors de la consultation d'informations sur un écran de grand format ;
- la comparaison des résultats d'évaluation d'interface des questionnaires *System Usability Scale* et *AttrakDiff* dans leur version française ;
- la mise en lumière de l'intérêt d'utiliser des interactions multi-utilisateur, non instrumentées et sans calibrage, basées sur une technique de type « *head tracking* » ;
- la révélation du bon accueil réservé par les utilisateurs à ces moyens, modes et paradigmes d'interaction ;
- Le potentiel d'application considérable qu'offre cette technique d'interaction.

## Publications

Marion, D., Favier, P. A., Bey, C., Lini, S., & Andre, J. M. (2015, 2015-07-08). *Etude outillée de l'attention visuelle flottante dans la recherche collaborative d'informations sur grands écrans*. Paper presented at the EPIQUE 2015, Aix-en-Provence, France.

Marion, D., Favier, P. A., & André, J. M. (2017, August). *Caractérisation des mouvements de la tête lors d'une tâche de consultation d'un affichage de grand format à l'aide d'un dispositif de head tracking*. In *29ème conférence francophone sur l'Interaction Homme-Machine* (117-124). ACM.

## Soumis :

Favier, P. A., Marion, D., & André, J. M. (2017, September). *Use of the Kinect v2 to Characterize Usual Head Movements during Information Consultation on a Large Display*. CHI 2018.

## Intérêt d'une démarche anthropocentrée

Ce travail de doctorat sur l'étude des interactions entre un écran de grand format et un groupe d'utilisateurs s'inscrit dans le domaine de la Recherche des interactions Homme-Machine (IHM). Selon Michel Beaudouin-Lafon (Beaudouin-Lafon 1992), l'interaction homme-machine est « un domaine pluridisciplinaire consacré à l'étude, la conception, le développement, l'évaluation de systèmes informatiques interactifs mais aussi, plus largement, à l'impact organisationnel, sociétal et même éthique des technologies de l'information. » L'évolution permanente des nouvelles technologies et l'émergence continue de nouveaux usages nous conduisent à nous intéresser de plus en plus aux interactions entre l'Homme et son environnement.

*« Il serait sot de nier l'importance de la communication efficace entre l'homme et la machine, aussi bien que l'inverse. Ma prévision est toutefois que la vraie révolution des prochaines décennies viendra davantage encore de ce que les hommes ont à se dire par l'intermédiaire des machines »*

— James Cannavino, *The Next Generation of Interactive Technologies* (Frenkel 1989).

Cette approche visant à concevoir les systèmes techniques et à en offrir l'interaction aux utilisateurs est mise en œuvre à l'aide de deux démarches bien distinctes. La démarche anthropocentrée qui consiste à placer les futurs utilisateurs au centre du processus de conception (Mayhew 1999), afin de développer des systèmes techniques complexes en adéquation avec leurs besoins et leurs usages, est en opposition avec la démarche technocentrée. Cette démarche anthropocentrée est de plus en plus en vogue dans l'industrie mais aussi dans la Recherche. En analysant les mécanismes de la pensée humaine face aux systèmes techniques dans le but de rendre les interactions plus naturelles et intuitives, elle nécessite la mise en application de concert des sciences techniques et des sciences humaines.

La Cognitique est définie par Jean-Michel Truong comme « la discipline scientifique et pratique technique, branche de l'informatique, qui a pour objet l'acquisition et la représentation formelle des connaissances et des modes de raisonnement, en vue de leur simulation à l'aide d'ordinateurs » (Claverie 2005). Le terme a été formé grâce à l'association de deux mots, « Connaissances » et « Automatique ». Cette discipline est enseignée à l'Ecole Nationale Supérieure de Cognitique, seule école publique française, formant des ingénieurs cogniticiens spécialistes de la prise en compte des Facteurs Humains dans les Interfaces Hommes-Systèmes (IHS). En effet, le domaine d'expertise d'un ingénieur cogniticien ne s'arrête pas aux IHM, mais se prolonge dans l'analyse et l'étude des systèmes complexes comme les organisations humaines par exemple. Pour un ingénieur cogniticien, la Cognitique est avant tout un état d'esprit, une manière particulière d'appréhender les problèmes. C'est pourquoi la définition de la Cognitique citée par Carrilho comme : « la science du traitement de la connaissance ; ensemble des méthodes et techniques visant à adapter la technologie aux capacités, limites, et préférences humaines » (Roventa-Frumusani 2015) caractérise selon moi qu'une partie de la démarche d'un ingénieur cogniticien car celui-ci limite sa définition à l'adaptation de la technologie aux préférences humaines. Or un ingénieur cogniticien est capable d'intervenir sur des systèmes non technologiques. Ainsi, il est capable d'adapter la ou les technologies aux besoins (Lespinet-Najib 2013) et préférences humaines (Rasmussen 1983). Nous parlons bien d'un utilisateur face à un système technique ou de plusieurs utilisateurs face à un ou des systèmes techniques mais également d'analyser et d'adapter des

systèmes complexes. Cette discipline est donc à l'intersection du monde de l'ingénierie technique et de l'ingénierie humaine.

L'ingénierie cognitive (Boy 2003) ou Cognitique est la discipline centrale de ce travail de doctorat. Elle s'appuie sur la démarche de Conception Centrée Utilisateur (CCU). Amélie Roche dans sa thèse de doctorat (Roche 2015) s'en est également inspirée pour proposer une méthode de conception innovante adaptée aux personnes en situation de handicap, appelée Approche Méthodologique Innovante de Conception Adaptée Systémique (AMICAS). La CCU consiste à prendre en compte les utilisateurs et leurs besoins tout au long du processus de conception d'un système complexe (Norman et Draper 1986, Nielsen 1994, Norman 2002) et permet de s'assurer un « bon accueil » des futurs utilisateurs (Rogers, Sharp et al. 2011). Cette démarche est à l'origine d'une norme internationale, la norme ISO 9241-210 (DIS 2009). Cette norme fournit des exigences et des recommandations relatives aux principes et aux activités de conception centrée sur l'opérateur humain, intervenant tout au long du cycle de vie des systèmes informatiques interactifs. Elle est destinée à être utilisée par les responsables de la gestion des processus de conception, et traite des manières dont les composants matériels et les logiciels des systèmes interactifs permettent d'améliorer l'interaction homme-système. La norme ISO replace l'intérêt de cette démarche : « Cette approche favorise l'efficacité et l'efficience<sup>2</sup>, améliore le bien-être de l'homme, ainsi que la satisfaction des utilisateurs, l'accessibilité et la durabilité, et réduit les effets nuisibles potentiels de leur utilisation sur la santé, la sécurité et les performances ».

Il apparaît alors primordial dans le cadre de notre travail de s'appuyer sur l'ingénierie cognitive. En effet, notre but est de proposer de nouvelles familles d'interactions intuitives, naturelles et affordantes<sup>3</sup> (Gibson 1977, Gaver 1991, Norman 1999, Gibson 2014, Bona 2016). Ces familles d'interactions doivent amener un confort d'utilisation et non une difficulté ou une gêne. Les cas d'application principaux se situant le plus souvent dans les lieux publics, les interactions doivent permettre à une population hétérogène d'un point de vue physique, cognitif, social et générationnel d'interagir le plus naturellement et simplement possible avec le système. Le système proposé doit être utilisable par tous.

Nous empruntons le concept d'utilisabilité d'un système à Nielsen qui le définit au sein d'un concept plus large, l'acceptabilité (Nielsen 1994). Selon lui, l'acceptabilité décrit si un système satisfait tous les besoins et exigences des utilisateurs. Ainsi, l'acceptabilité d'un système est une combinaison d'une dimension sociale et d'une dimension pratique : l'acceptabilité sociale et l'acceptabilité pratique. L'acceptabilité sociale est l'adhésion des utilisateurs envers le système. L'acceptabilité pratique est divisée en deux catégories :

- l'utilité pratique, qui indique si un système peut être utilisé pour atteindre un objectif fixé, elle-même divisée en deux sous catégories :
  - l'utilité théorique qui explique l'adéquation entre l'outil et l'activité à réaliser ;
  - l'utilisabilité qui indique la façon dont les utilisateurs vont se servir du système ;
- le coût, la compatibilité, la faisabilité.

---

<sup>2</sup> Définition Larousse 2017 : Capacité d'un individu ou d'un système de travail d'obtenir de bonnes performances dans un type de tâche donné ; efficacité : Donner la preuve de son efficience. L'efficience c'est l'efficacité à coût maîtrisé.

<sup>3</sup> Capacité d'un élément à suggérer sa propre utilisation.

Il définit l'utilisabilité comme la facilité d'apprentissage et de mémorisation, l'efficacité et la fiabilité du système (faible taux d'erreurs et corrections faciles) et la satisfaction avec laquelle un utilisateur réalise une tâche dans un environnement donné (plaisir d'utilisation).

La notion d'expérience utilisateur est présentée pour la première fois par Norman et al., dans le but de couvrir l'ensemble de l'expérience d'un utilisateur avec un outil de service (Norman, Miller et al. 1995). Mais au-delà de l'aspect fonctionnel de l'utilité et de l'utilisabilité, l'UX aborde le côté subjectif de l'expérience vécue par un utilisateur avec un système interactif. Les dimensions émotionnelles, hédoniques et esthétique viennent s'ajouter aux dimensions fonctionnelles abordée par Norman (Lallemand, Gronier et al. 2013). L'UX englobe donc des aspects fonctionnels (utilisabilité et utilité de Norman) et des aspects émotionnels, hédoniques et esthétiques comme le présente le modèle de Drewitz et al., et celui de Thüring et al. (Thüring et Mahlke 2007, Drewitz et Brandenburg 2010). Néanmoins, il serait imprudent de ne considérer l'expérience utilisateur que du point de vue de l'utilisateur et non du point de vue du contexte et de l'environnement dans lequel il se trouve. C'est de ce constat qu'émerge le courant situé de l'action et de la cognition associé aux travaux de Lucy Suchman (Suchman 1987, Suchman 2007). Cette approche considère la construction de l'action et de la cognition dans la situation liée à l'engagement de l'individu à un moment donnée avec le monde extérieur et ne considère plus l'action comme une simple réponse à des stimuli (Relieu, Salembier et al. 2004, Villame 2004). La cognition située est donc influencée par de nombreux facteurs comme les caractéristiques matérielles, sociales et culturelles. Ainsi, l'individu agira en fonction de la signification qu'il accorde aux événements qu'il rencontre et à la situation qui l'entoure. Il est donc primordial d'intégrer les facteurs environnementaux et contextuels et de développer des connaissances fines sur l'activité et la diversité des situations rencontrées par les utilisateurs afin de concevoir des systèmes qui soient réellement en adéquation avec leurs besoins.

Nous nous attacherons à prendre en compte la globalité de l'expérience utilisateur (utilisabilité + utilité + aspects émotionnels, hédoniques et esthétiques) et la cognition située (situation environnementale et contexte social) dans l'étude et le développement des interactions entre les utilisateurs et l'écran.

Cependant, cette démarche visant à mettre exclusivement l'utilisateur au centre de la boucle de conception est souvent pointée du doigt par l'ingénierie technique et le management car elle conduirait à des solutions « idéales », mais difficilement réalisables techniquement et engendrerait des coûts financiers, humains et temporels souvent imaginés comme trop importants (Abrams, Maloney-Krichmar et al. 2004) malgré une réelle valeur ajoutée et un très bon retour sur investissement. Néanmoins, l'histoire montre que les évolutions dans les usages sont apparues grâce à des innovations dites « de ruptures ». Ces ruptures dans les usages ne sont pas le fruit de nouvelles créations mais bien du mode d'utilisation qu'en font les utilisateurs dans un contexte et un environnement donnés. C'est pourquoi, appliquer une démarche de conception centrée utilisateur dans sa globalité est essentiel pour développer les innovations de demain.

# Chapitre 2 : Etat de l'art

Ce chapitre est constitué de deux parties. La première partie concerne les recherches menées sur le concept d'attention et la définition de ce concept dans nos travaux de recherche. La deuxième partie s'attache à présenter un état de l'art sur les écrans de grand format. Cette deuxième partie permet d'expliquer le cheminement de notre démarche et aboutit à la présentation des premières lignes directrices en termes de conception d'interactions avec les écrans de grand format.

## L'attention

Dans cette première partie, nous utilisons souvent le concept d'attention très controversé en psychologie et psychologie cognitive. Il convient donc d'en donner une définition. Le dictionnaire Larousse définit le terme comme : « Activité ou état par lesquels un sujet augmente son efficacité à l'égard de certains contenus psychologiques (perceptifs, intellectuels, mnésiques, etc.), le plus souvent en sélectionnant certaines parties ou certains aspects et en inhibant ou négligeant les autres ».

William James, psychologue américain écrit : « L'attention est la prise de possession par l'esprit, sous une forme claire et vive, d'un objet ou d'une suite de pensées parmi plusieurs qui semblent possibles [...] Elle implique le retrait de certains objets afin de traiter plus efficacement les autres » (James 1890).

Ces deux premières définitions laissent transparaître les notions de sélection et de contrôle de l'attention.

## L'attention sélective

Dans sa théorie de 1958 (Broadbent 1958), Donald Broadbent expose le fonctionnement de l'attention à travers un filtre, nommé le filtre attentionnel. Son modèle de l'attention est basé sur les travaux expérimentaux sur l'écoute dichotique de Cherry (Cherry 1953). En effet, dans son expérience Cherry fait écouter à l'aide d'un casque audio un message (principal) dans une oreille et un autre (secondaire) dans l'autre oreille. Les sujets doivent répéter le message principal et ignorer le secondaire. Les résultats montrent que les caractéristiques sémantiques des stimuli secondaires (changement de langue) ne sont pas détectées, seules les caractéristiques physiques élémentaires (changement de voix : féminine → masculine) sont traitées. Ainsi, pour Broadbent (Broadbent 2013), le traitement de l'information se fait dans un canal attentionnel unique avec sélection des informations sur la base des caractéristiques physiques élémentaires des stimuli. L'information non sélectionnée n'accède pas aux étapes ultérieures du traitement cognitif, c'est l'image d'un canal unique avec un goulot d'étranglement.

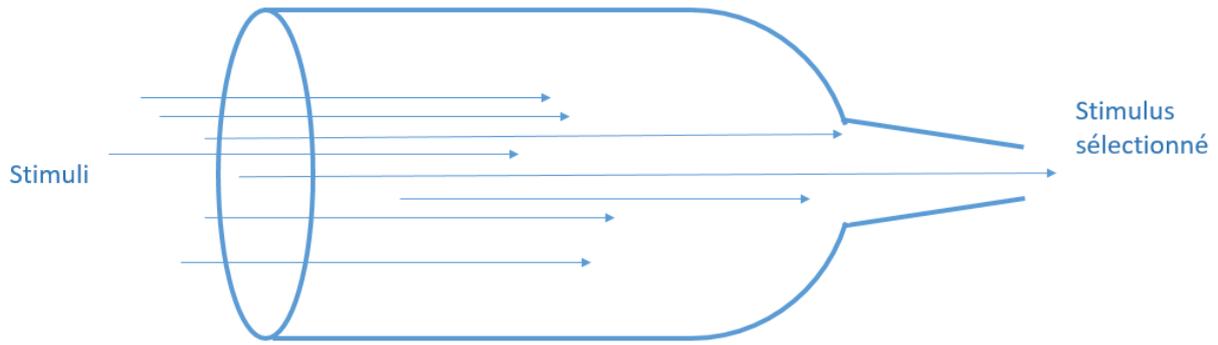


Figure 6 : Goulot d'étranglement du canal unique, adapté de (Broadbent 2013).

Le modèle de Broadbent se compose donc d'un filtre des stimuli dit précoce.

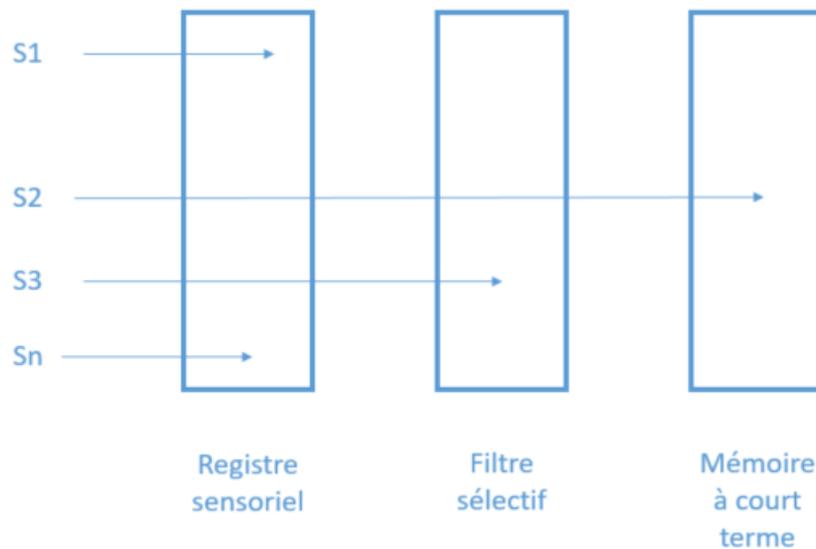


Figure 7 : Modèle de l'attention de Broadbent, filtre précoce, adapté de (Broadbent 2013).

Cependant, Moray met en évidence l'effet « cocktail party » (Moray 1959). Moray constate que lorsqu'un sujet focalise son attention sur un interlocuteur dans une situation où d'autres conversations se tiennent en parallèle (« soirée cocktail »), il ignore effectivement le contenu de ces conversations (messages secondaires), jusqu'à ce qu'une information importante lui surgisse (exemple : son prénom, celui de sa compagne, etc.). Or selon le modèle de Broadbent la « capture » attentionnelle d'une information contenue dans un message secondaire n'est pas possible. C'est ainsi que Treisman propose un modèle alternatif définissant un filtre distribué hiérarchiquement à différent niveau de traitement de l'information. Un niveau précoce opérant une atténuation ou une accentuation de l'information sur la base des caractéristiques physiques élémentaires du message puis un niveau tardif dans la mémoire à long terme dépendant d'un seuil d'excitation variable en fonction de l'importance de l'information (Treisman 1964).

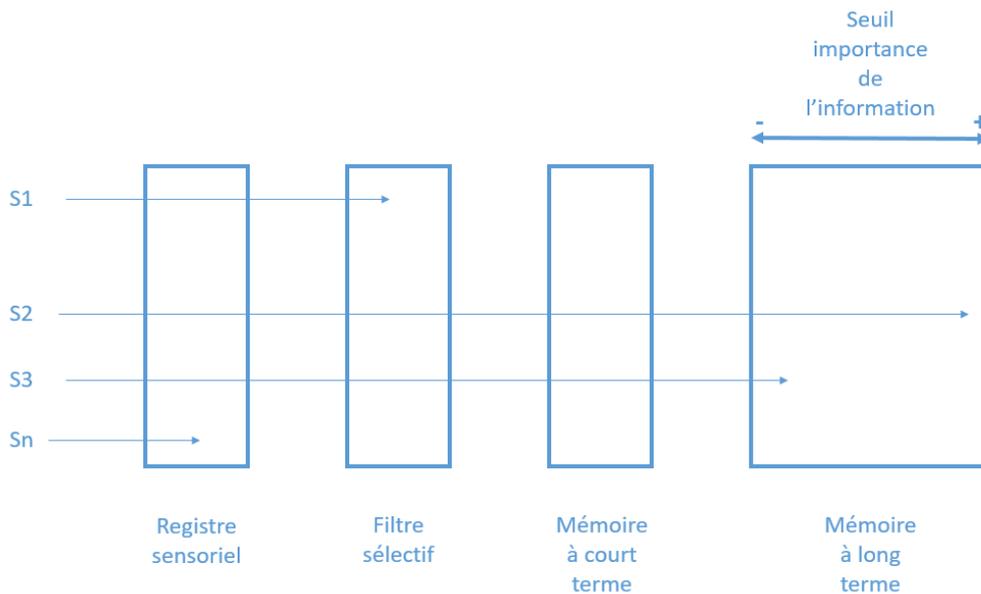


Figure 8 : Modèle de l'attention de Treisman, filtre distribué, adapté de (Treisman 1964).

Néanmoins pour Deutsch et Deutsch (Deutsch et Deutsch 1963) et Norman (Norman 1968), il n'y a pas de filtrage précoce de l'attention, tous les stimuli (principaux et secondaires) atteignent la mémoire à long terme, la sélection des informations se fait après analyse sémantique reposant sur la pertinence du stimulus, c'est le filtre tardif.

Les différentes visions de l'attention présentées jusqu'ici sont basées sur l'idée d'une capacité limitée de traitement de l'information avec la présence et la place d'un filtre attentionnel à différents niveaux. D'autres approches de l'attention basées, non plus sur des contraintes structurelles mais sur l'existence de contraintes de traitement, existent pour expliquer les baisses de performance. Elles considèrent que les ressources attentionnelles / mentales / cognitives sont limitées et qu'elles se répartissent en fonction des tâches à accomplir, c'est l'attention partagée.

## L'attention partagée

Kahneman s'inscrit dans cette approche (Kahneman 1973). Pour Kahneman, l'attention est une capacité limitée pouvant être déplacée de manière flexible en fonction des décisions de l'opérateur humain à chaque instant. Celui-ci peut décider de focaliser son attention sur une seule activité ou de la distribuer mais ses ressources attentionnelles sont limitées. Certes, les ressources sont limitées mais Navon et Gopher, avec leur théorie des ressources multiples, montrent que l'humain possède une structure de traitement permettant une utilisation en parallèle de ressources différenciées (Navon et Gopher 1979). Ainsi, Wickens dans son modèle 4D établit que deux tâches interféreront d'autant moins qu'elles auront des demandes de ressources séparées (Wickens 1991). Par exemple, une tâche auditivo-verbale et une tâche visuo-spatiale seront moins dégradées que deux tâches sollicitant les mêmes ressources attentionnelles (deux tâches visuo-spatiales).

L'analyse et l'évaluation de la gestion des ressources cognitives est au cœur de la psychologie cognitive (Yerkes et Dodson 1908). Dans leurs thèses de doctorat respectives, Sami Lini et Christophe Bey présentent l'état de l'art des différents modèles de la gestion de ces ressources

cognitives (Lini 2013, Bey 2016). Le terme de « ressources » apparaît dans (Norman et Bobrow 1975).

## L'attention contrôlée & automatique

Shiffrin & Schneider proposent deux systèmes attentionnels pouvant s'exécuter en parallèle (Schneider et Shiffrin 1977). Les processus attentionnels automatiques utilisés pour des tâches simples et sur-apprises traitant des stimuli familiers. Les processus attentionnels de contrôles mis en jeu pour des tâches complexes moins familières. Les processus attentionnels automatiques sont rapides, fonctionnent en parallèle, ne consomment pas de ressources attentionnelles, sont difficiles à interrompre et peuvent être exécutés simultanément avec les processus de contrôles attentionnels. Les processus de contrôles sont lents, fonctionnent en séquentiel, consomment beaucoup de ressources attentionnelles, s'interrompent facilement et ne peuvent pas être exécutés simultanément avec une autre activité contrôlée.

Posner (Posner, Nissen et al. 1978) centre ses travaux sur l'étude de l'orientation visuo-spatiale et met en avant l'existence de trois réseaux neuraux attentionnels : un réseau antérieur (régions préfrontales) responsable de l'orientation délibérée de l'attention, un réseau postérieur (cortex post-pariétal) responsable de l'orientation automatique de l'attention et un réseau d'éveil (latéralisé à droite) qui contribue à une alerte.

Laberge distingue lui deux formes d'attention en termes temporels (Laberge 1995). La première permet une focalisation très fine et labile (50 à 100 millisecondes) correspondant à de l'attention sélective et la deuxième qui a une fonction préparatoire pouvant durer plusieurs secondes, l'attention soutenue.

## L'attention dans cette thèse

Nous pouvons donc définir l'attention selon 3 composantes principales :

- l'intensité : qui correspond à la possibilité d'allouer des ressources attentionnelles plus ou moins importantes (attention soutenue, alerte, vigilance) pour répondre à des stimuli de manière plus ou moins efficiente ;
- le contrôle : permettant d'allouer son attention de manière volontaire ou non et d'en choisir son intensité ;
- la sélectivité : assurant une sélection des éléments à traiter, indispensable à cause de la limitation des ressources attentionnelles malgré des canaux multiples.

Ainsi, nous définissons les termes suivant de la sorte :

- attention sélective : la faculté de se concentrer sur une seule source d'information et de ne pas traiter les autres. Ce processus peut être contrôlé mais est souvent automatique ;
- attention partagée ou divisée : la faculté de traiter plusieurs sources d'informations à la fois ou de traiter plusieurs tâches en même temps grâce notamment aux multiples canaux attentionnels restreints par leur ressources attentionnelles propres ;
- attention commutée : la faculté de basculer son attention d'une information ou une tâche à l'autre. Pas de division de l'attention mais une focalisation de celle-ci sur une information ou une tâche particulière ;
- attention soutenue : la faculté de rester concentré sur une tâche ou une information durant une longue période.

# Etat de l'art sur les écrans de grand format

## L'omniprésence des écrans

L'immersion digitale est de plus en plus présente dans les lieux publics. Les écrans interactifs et les grands écrans sont omniprésents que ce soit dans la rue, les galeries de supermarchés, dans les vitrines de magasins, les stades, les gares, les aéroports, les écoles et les musées, etc. Ils sont partout. En quelques années, les panneaux d'affichages traditionnels sont devenus des écrans numériques très performants permettant d'afficher des contenus multimédias toujours plus variés. Cette intensification des écrans est en partie due à l'évolution considérable de leur performance mais également à leurs coûts qui deviennent attractifs. Cependant, en explorant l'impact sur les usages de tels dispositifs, la recherche scientifique tend à mettre en évidence des freins à leur utilisation et des précautions à prendre pour leur conception et leur implantation.

## Notion de contexte : privé, semi-privé et public

Les différentes recherches réalisées sur les écrans de grand format sont divisées, selon nous, en trois contextes d'utilisation :

- le contexte privé : les actions de l'utilisateur ne sont pas exposées à la vue d'autres utilisateurs ;
- le contexte semi-privé : les actions de l'utilisateur sont exposées à la vue seulement d'autres utilisateurs identifiés/connus ;
- le contexte public : les actions de l'utilisateur sont exposées à la vue de tous.

Il est important de noter que ces contextes d'utilisation ne sont pas exclusifs les uns des autres. En effet, certains systèmes peuvent favoriser un usage plutôt qu'un autre selon la fonctionnalité utilisée. Par exemple, dans l'étude de (Brignull et Rogers 2003), les utilisateurs sont à une soirée cocktail, tous les utilisateurs ne se connaissent pas et ils peuvent, s'ils le souhaitent, publier un message qui sera visible par tous sur un grand écran à l'aide d'un terminal déporté (écran + souris + clavier). Lorsqu'un utilisateur écrit un message celui-ci n'est pas visible sur le grand écran tant qu'il n'a pas validé l'envoi du message. Ainsi, dans un premier temps le système permet à l'utilisateur d'être dans un contexte privé voire semi-privé (regard sur l'écran des autres utilisateurs présents à côté du terminal) lors de la rédaction du message puis, dans un second temps, dans un contexte public lors de la publication de celui-ci sur le grand écran à la vue de tous. Par contre, il est important de noter que cette différenciation du contexte d'usage s'opère sur des affichages différents.

## Différents types d'écrans interactifs

De nombreux domaines de recherche s'intéressent aux écrans de grand format interactifs. Ces écrans interactifs peuvent se distinguer selon leur orientation (portrait *versus* paysage), leur disposition (horizontale *versus* verticale), les interactions proposées (directes *versus* indirectes, libres *versus* instrumentées), leur forme (un seul écran *versus* un mur composé de

plusieurs écrans), le contenu proposé (images, textes, vidéos, etc.) et le type d’affichage possible (statiques *versus* interactifs).

Durant de nombreuses années, la conception d’un grand écran nécessitait l’association de plusieurs écrans ou projecteurs et de plusieurs ordinateurs en réseaux formant ainsi un mur d’écrans. Ces murs d’écrans imposent de fortes contraintes liées aux différentes bordures physiques (Tan et Czerwinski 2003, Mackinlay et Heer 2004) formant un quadrillage visible à l’œil nu et brisant ainsi la continuité de l’image affichée (pouvant être gênant pour l’affichage de textes). Néanmoins, ils permettent un affichage à moindre coût avec une meilleure résolution et un meilleur contraste de couleur dans des conditions de luminosité élevée notamment en comparaison à un rétroprojecteur. De plus, des auteurs montrent que les bordures à la jonction de deux écrans peuvent devenir un moyen rapide de catégorisation spatiale de l’information (Grudin 2001, Ball et North 2005).

Les avancées dans le domaine des composants matériels et de la puissance de calcul ont largement contribué à l’amélioration de la résolution<sup>4</sup> et de la taille des écrans de grand format permettant ainsi d’accroître celles-ci sans avoir forcément recours à un mur d’écran. Nous ne nous attarderons pas sur les problèmes techniques et de configurations que génèrent le déploiement et la conception des grands écrans. Ni et al. et Guimbretière et al., présentent un état de l’art de ces problèmes dans leurs articles (Guimbretière, Stone et al. 2001, Ni, Schmidt et al. 2006).

Nous avons choisi d’étudier plus particulièrement les grands affichages composés d’un seul écran. Parmi les grands affichages composés d’un seul écran, nous pouvons faire une première distinction basée sur leur disposition par rapport au sol, à savoir verticale ou horizontale.

## Les grands écrans horizontaux

L’étude des techniques de bureaux virtuels tels que *DigitaDESK* par exemple, autorise la fusion de documents physiques et digitaux à l’aide d’une caméra et d’un vidéoprojecteur placés au-dessus du bureau. Ces techniques permettent de réaliser un copier/coller de l’information contenue dans un document physique vers un document digital (Wellner 1993). Ces études sur les bureaux virtuels ont débouché sur l’avènement des tables interactives dans les lieux publics. Les principaux problèmes de ces tables sont liés à leur disposition. En effet, leur disposition à l’horizontale entraîne une limitation physique très rapide en termes de nombre d’utilisateurs et une gestion de l’orientation du texte vite problématique lorsque deux utilisateurs sont situés face à face et souhaitent lire un même contenu (Shen, Everitt et al. 2003, Shen, Vernier et al. 2004, Krumbholz, Leigh et al. 2005). Ainsi, nous avons choisi de nous intéresser plus particulièrement aux grands affichages composés d’un seul écran en disposition verticale par rapport au sol.

## Les grands écrans verticaux

Les écrans de grand format en disposition verticale peuvent être orientés de deux façons :

- en mode paysage, la base horizontale de l’écran est plus grande que sa hauteur ;
- en mode portrait, la base horizontale de l’écran est plus petite que sa hauteur.

---

<sup>4</sup> Le mot résolution est employé pour décrire la densité de pixel à l’écran, c’est-à-dire le nombre de point pour un pouce (2,54cm). Dans la suite du document, nous parlerons davantage de résolution d’affichage désignant le nombre de pixel sur l’écran.

Nous avons choisi un contexte d'usage dans les lieux publics. Ainsi, nous pensons qu'une orientation de l'écran en mode paysage (présentant une base image horizontale plus importante) va nous permettre de partager l'affichage de manière plus efficace entre les individus. Nous choisissons donc de nous intéresser aux écrans de grand format composés d'un seul écran en mode paysage.

## Type d'interaction avec les grands écrans : directes *versus* non directes

Les recherches sur les écrans de grand format interactifs se classent dans deux grandes catégories :

- les médias ambiants : ces écrans proposent des moyens d'interactions non directes à la périphérie de la perception humaine. L'écran réagit par rapport à l'utilisateur mais celui-ci ne contrôle pas les interactions, ce sont des interactions implicites. L'utilisateur n'est pas conscient que le système réagit par rapport à ses actions et ses besoins (Mankoff, Dey et al. 2003, Vogel et Balakrishnan 2004, Pousman et Stasko 2006, Sadri 2011) ;
- Les écrans de grand format proposant des interactions directes avec le contenu (Endert, Andrews et al. 2011, Alt, Shirazi et al. 2013, Nancel, Chapuis et al. 2013, Vidal, Bulling et al. 2013, Walter, Bailly et al. 2013, Liu, Chapuis et al. 2016). Dans ce cas, l'utilisateur est conscient que ce sont ses actions qui font réagir le système, il a le contrôle. Les interactions sont dites explicites.

Nous avons décidé d'utiliser une technique basée sur la détection et le suivi de la position et de l'orientation de la tête des utilisateurs comme mode d'interaction avec l'écran. Le moyen d'interaction étant la zone d'attention, nous sommes sur des interactions explicites.

Nos travaux s'intéressent donc aux grands écrans verticaux composés d'un seul écran, orientés en mode paysage et autorisant des interactions directes.

## Avantages des écrans de grand format

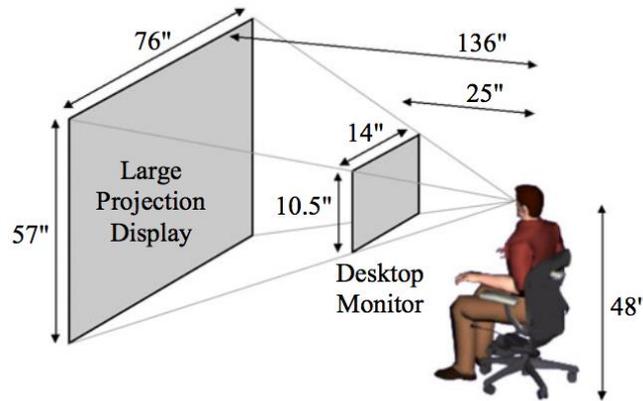
### L'augmentation de la taille de l'écran : quels impacts sur les utilisateurs ?

Plus la taille de l'écran augmente et plus l'écran devient qualitativement différent (Swaminathan et Sato 1997). De nombreuses études ont été menées concernant l'avantage que procure l'utilisation d'un écran plus grand. Ces études s'attachent à montrer les avantages et les inconvénients de l'augmentation de la taille et de la résolution de l'affichage. Par exemple, l'augmentation de la taille physique des écrans aide pour la mémorisation (Tan, Stefanucci et al. 2001, Lin, Duh et al. 2002, Czerwinski, Smith et al. 2003). De plus, l'utilisation de grands écrans améliore les performances dans des situations de tâches multiples<sup>5</sup> (Czerwinski, Smith et al. 2003). Tan et al., montrent dans leur travaux que l'utilisation de grands écrans augmente les capacités d'orientation spatiale et de navigation en 3D des utilisateurs en comparaison avec un écran de même résolution mais de plus petite

---

<sup>5</sup> L'utilisateur réalise une tâche spécifique (tâche principale) qui requiert la réalisation d'une autre tâche (tâche secondaire). Exemple : lors de l'écriture d'un article il est commun d'avoir d'autres documents d'ouverts (références) pour consultation.

taille (Tan, Gergle et al. 2003). Dans leur expérience, l'angle de vue entre les deux tailles d'écran différentes est conservé, comme le montre la *Figure 9* ci-dessous.



*Figure 9 : Configuration physique de l'expérience de Tan et al, issue de (Tan, Gergle et al. 2003).*

Tyndiuk (2005) dans sa thèse de doctorat, montre également que l'utilisation d'un grand écran permet de diminuer l'implication des capacités spatiales<sup>6</sup> dans une tâche en réalité virtuelle et de soutenir un haut niveau de performance pour les utilisateurs ayant des difficultés spatiales (Tyndiuk 2005).

Czerwinski et al., montrent une augmentation significative de la productivité lors de l'utilisation d'écrans plus larges (3 écrans) comparé à un seul écran (Czerwinski, Smith et al. 2003). De plus, il est significativement démontré que la comparaison d'informations fines est mieux réalisée sur des grands écrans (Ball et North 2005).

Simmons montre que plus l'écran est grand (écrans de 15, 17, 19 et 21 pouces de diagonales) plus l'utilisateur est performant pour des tâches d'exploration de tableaux (Microsoft Office Excel) et des situations de tâches multiples (Simmons 2001). L'utilisation de grands écrans facilite la gestion de multiples fenêtres en évitant de passer d'une application à une autre, ce qui a pour effet de réduire les ressources cognitives mises en jeu. En effet, en évitant la navigation virtuelle nécessaire entre deux applications sur petit écran, l'effort de mémorisation des informations contenues dans un document spécifique est moindre ainsi que la concentration. L'utilisation de grands écrans génère une priorisation et une catégorisation spatiale des éléments à l'écran. Les utilisateurs placent généralement les éléments principaux au centre de l'écran et les éléments secondaires autour (Grudin 2001, Ball et North 2005). Ainsi, lors de la réalisation d'une tâche sur grand écran les utilisateurs déplacent les objets correspondant à leur tâche principale au centre de l'écran (le plus directement possible en face d'eux) plutôt que de se déplacer physiquement pour faire face à l'objet. Ces auteurs montrent également que le travail collaboratif (affichage d'images, recherches internet, écriture de documents, etc.) est facilité par l'utilisation des écrans de grand format (Ball et North 2005).

Dans une étude présentée par Huang et al. (2007), les auteurs expliquent que, dans un contexte public, lorsque les utilisateurs ont le choix entre deux contenus identiques d'information affichés sur des écrans de tailles différentes, ceux-ci favorisent les écrans les plus petits (Huang 2007). Ce constat est établi dans plusieurs lieux observés au cours de cette expérience. Les petits écrans plus intimistes et plus habituels que les grands écrans attirent donc plus facilement les utilisateurs. Il est donc important de porter une attention toute particulière lors de la conception et de l'implantation de ces écrans, aux sentiments

<sup>6</sup> Les capacités spatiales font généralement référence à la capacité à produire, représenter, transformer et rappeler des informations spatiales. Linn, M. C. and A. C. Petersen (1985). "Emergence and characterization of sex differences in spatial ability: A meta-analysis." *Child development*: 1479-1498.

d'exposition (voir partie *Embarras social*) et de confidentialité éprouvés par les utilisateurs. En effet, ces éléments auront une forte incidence sur l'expérience, la perception et le confort. Ainsi, proposer différentes tailles d'écran en fonction des contextes d'usages est une solution à considérer.

## Les grands écrans catalyseurs des interactions sociales

De nombreuses études démontrent le caractère « brise-glace » des écrans de grand format et le bon accueil réservé par les utilisateurs à ces technologies (Churchill, Nelson et al. 2003). Les grands écrans incitent les interactions sociales entre les utilisateurs (Agamanolis 2003, Churchill, Nelson et al. 2003, Grasso, Mühlenbrock et al. 2003, McCarthy 2003, Peltonen, Salovaara et al. 2007). Borovoy et al. (1998), constatent que les nouveaux arrivants « en dehors du réseau » trouvent très difficile de briser la glace et d'intégrer un réseau existant dans une soirée par exemple. La présence d'un écran crée ainsi un point de référence partagé pour démarrer la conversation (Borovoy, Martin et al. 1998). Les grands écrans facilitent donc les interactions humain-humain en proposant un contexte social. Ils créent un « buzz » social (l'endroit où il faut être pour essayer). Les utilisateurs n'ont pas les mêmes mécanismes sociaux ni spatiaux lorsque ils interagissent en présence d'autres individus ou seul face à une technologie (Bellotti and Sellen, 1993, Dourish, 1993 ; Harper)

## Faiblesses des écrans de grand format

### Absence de considération des écrans de grand format : Display Blindness

#### *La surcharge informationnelle*

L'absence d'utilisation d'un système technologique (ou non) est toujours liée à la représentation que se font les utilisateurs de ce système et de l'utilisation qu'ils pourraient en faire. Pourtant les concepteurs, notamment les concepteurs/exploitants d'écrans de grand format pensent que leurs technologies attirent le regard des utilisateurs de par leur nature. Or, les écrans de grand format sont souvent ignorés (Huang 2007). Le manque d'attention porté à ces écrans est souvent expliqué par le fait que notre attention est hautement sélective. Le monde qui nous entoure produit un flux d'informations à traiter tellement important qu'il peut être difficile dans certains cas de tout comprendre et analyser (fatigue, stress, situation de handicap, etc.). Ce constat est d'autant plus vrai dans les environnements urbains d'aujourd'hui. En effet, Milgram (1970) explique que dans ce type de milieu les individus sont soumis à une surcharge informationnelle (Milgram 1970, Burke, Hornof et al. 2005). Eppler et Mengis (2004) décrivent ce phénomène comme suit : « La surcharge informationnelle est mise en évidence lorsque une personne n'est plus capable de reconnaître, comprendre ou gérer les informations qui lui sont présentées du fait d'une trop grande quantité d'informations à traiter. » (Eppler et Mengis 2004). En d'autres termes, lorsque notre capacité à traiter l'information est inférieure au flux d'informations que l'on reçoit, nous avons de grandes difficultés à traiter ces informations et à identifier leur pertinence. Notre attention devient alors très sélective.

#### *Les préjugés du contexte*

L'attention portée aux écrans est très corrélée à ce que l'on pense y trouver, notamment en matière d'information. De plus, nos attentes sont liées aux contextes dans lesquels nous sommes confrontés aux écrans. En effet, selon Müller et al. (2009), les attentes des utilisateurs

et donc l'attention qu'ils portent aux écrans dépendent de qui « possède » l'écran. Par exemple, dans un magasin les utilisateurs s'attendent à voir des publicités, ils portent donc moins d'attention à l'écran que dans un couloir d'université où ils s'attendent à trouver des informations importantes pour eux (localisation de la salle de cours, agenda, *etc.*)(Müller, Wilmsmann et al. 2009). Dans une autre expérience, Müller et al. (2009) mettent en évidence trois facteurs influençant l'audience des affichages publics : l'emplacement (Churchill, Nelson et al. 2003, Nonnecke et Preece 2003), le moment de la journée (matin, midi, après-midi et soirée) et le type de contenu (Müller, Exeler et al. 2009). Ces trois facteurs ont une influence significative sur le temps passé à regarder un écran. Cependant, leur étude montre que cette durée n'est pas corrélée à la préférence des utilisateurs pour un type de contenu. Ils concluent que le moment de la journée et la localisation sont les facteurs les plus influents sur l'attention portée aux affichages. Le type de contenu quant à lui influence la durée de l'attention portée à un écran. Présenter le bon contenu peut faire que les utilisateurs regardent plus longtemps. Dans leur expérience, la population étudiée est très homogène (employés et usagers d'une université). Ce manque de diversité des utilisateurs réduit l'effet attendu du type de contenu présenté au regard de la localisation et du moment de la journée. En effet, les utilisateurs ne viennent pas regarder les écrans pour y trouver de l'information utile mais seulement parce qu'en passant devant, le contenu a pu attirer leur attention. Cet effet est également amplifié par la durée de présentation de seulement vingt secondes de chaque contenu qui n'incite pas à rester longtemps. Le changement de contenu affiché à l'écran génère une rupture que l'humain considère inconsciemment comme un signal de fin. Ainsi, la fin d'une vidéo, le changement de contenu dans une boucle d'affichage vont nous amener plus facilement à rompre notre attention avec l'écran car nous interprétons cette rupture comme la possibilité de changer d'activité. C'est le même principe lorsqu'un lecteur termine un chapitre et se demande s'il entame le suivant ou arrête sa lecture. Il faut éviter les changements abrupts de contenu afin de maintenir l'audience.

### *L'audience capturée*

De nos jours les écrans sont partout et ils sont souvent placés de manière à ce que les utilisateurs soient forcément face à eux (écrans incrustés dans les caisses dans les bureaux de tabac, les escalators, placés en face d'une file d'attente dans une banque, une boutique, *etc.*)(Müller, Paczkowski et al. 2007). Ces écrans sont pourtant souvent ignorés (Huang 2007). Cependant, Huang et al. (2007), montre que lorsque ces écrans proposent un retour caméra de l'environnement dans lequel évolue l'utilisateur (ex : caméra de sécurité), ils captent le regard. Présenter l'image des utilisateurs sur un grand écran en leur présence aide donc à capturer leur attention.

### *Fugacité de l'attention envers les grands écrans*

Huang et al. (2007), observent que les utilisateurs ne regardent que très brièvement les écrans dans les lieux publics (1 à 2 secondes). Néanmoins, les écrans présentant des vidéos maintiennent légèrement plus l'audience. Des études antérieures (Müller et Rabbitt 1989) menées sur des passants suggèrent que les regards de plus de huit cents millisecondes sont intentionnels. Ces observations amènent à penser que les utilisateurs se font un avis très rapidement sur l'intérêt et la pertinence du contenu affichés sur ces écrans. Il faut donc privilégier des formats de contenus qui permettent aux utilisateurs de déterminer la pertinence des informations affichées en quelques secondes. Ainsi, les textes et même les phrases brèves sont à éviter.

## Panneaux d'affichage physiques *versus* panneaux d'affichage digitaux

Les utilisateurs privilégient les affichages physiques par rapport aux affichages digitaux présentant du contenu dans une boucle d'affichage (Huang 2007). En effet, cette boucle oblige les utilisateurs à attendre la prochaine apparition du contenu qu'ils désirent, s'ils n'ont pas eu le temps de le consulter dans son intégralité. En outre, les écrans possédant une boucle d'affichage et n'autorisant pas les utilisateurs à agir sur le contenu ne parviennent pas à attirer l'attention des utilisateurs et encore moins à la maintenir. Ce résultat montre que la dynamisation de l'affichage à l'aide d'une boucle d'affichage n'est pas une solution viable. Cependant, il ne permet pas de conclure sur le fait que les utilisateurs préfèrent les affichages physiques aux affichages digitaux ni que les affichages statiques sont préférés aux affichages dynamiques. Nous pensons qu'il serait intéressant pour étendre ces résultats de comparer un même affichage tantôt statique tantôt dynamique (avec des interactions autres qu'en boucle d'affichage). Nous présentons cette étude comparative dans le *Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via head tracking avec les écrans de grand format.*

## Embarras social

### *Espace personnel*

Lors de l'utilisation à plusieurs et simultanément des écrans de grand format, une gêne sociale peut apparaître. Cette gêne se manifeste lorsque les autres entrent dans notre espace personnel (Lécuyer 1976, Le Boulch 2001, Maisonneuve 2014) ou lorsque nous avons la sensation d'envahir le leur (Rogers et Lindley 2004). L'espace personnel peut être vu comme une distance physique dynamique entre soi et les autres. En effet, cette distance est variable car elle évolue en fonction du contexte social, de la culture, de la personnalité, etc. Malgré une grande variabilité interindividuelle, nous tolérons certaines personnes dans certains contextes, à des distances physiques plus ou moins grandes.

### *Appréhension de l'inconnu*

Un autre problème très connu de ces nouvelles formes d'affichage est qu'elles peuvent engendrer une appréhension, surtout lorsque les utilisateurs ne sont pas initiés à ce genre de systèmes et qu'ils ne sont pas particulièrement affordants (Ram 1987, Rogers Everett 1995). Par exemple, lorsque ces systèmes sont composés de plusieurs écrans, il peut être difficile de bien comprendre où se placer pour interagir avec l'écran et même où regarder et quoi regarder sur cet ensemble d'écrans.

Ce pose alors le problème de l'intention d'usage mis en avant par Dillon & Morris (1996, 1999) dans leur modèle ci-dessous (Dillon et Morris 1996, Dillon et Morris 1999).

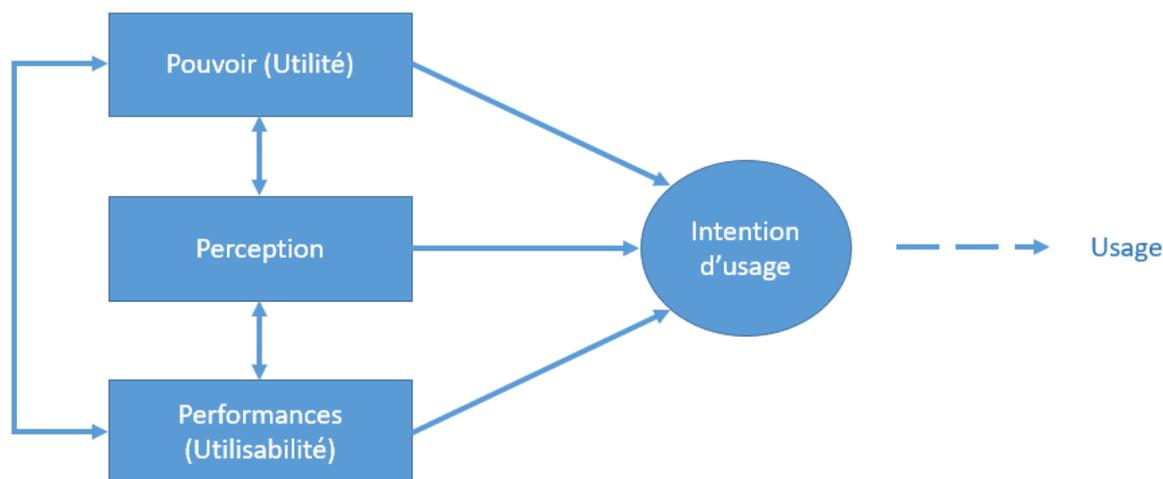


Figure 10 : Modèle P3 proposé par Dillon et Morris (1996, 1999)

Dans leur modèle de l'utilisabilité, Dillon et Morris (1996,1999) utilisent les termes de « pouvoir » pour faire référence à l'utilité (ce que l'outil peut faire), de « performance » pour adresser l'utilisabilité et de « perception » pour mentionner les éléments perçus par les utilisateurs. Il est à noter que dans leur modèle, des effets directs et indirects sur l'usage sont présentés : le pouvoir (utilité), la performance (l'utilisabilité) et la perception ont tous une influence directe sur la décision de l'utilisateur à adopter une technologie. Cependant, le modèle P3 suggère aussi que le pouvoir (utilité) et la performance (utilisabilité) ont également des effets indirects sur l'utilisation, à travers la perception. En d'autres termes, les perceptions sont basées d'une part sur la compréhension de l'utilisateur des capacités proposées par l'outil (pouvoir), et d'autre part sur ses capacités à en exploiter les fonctionnalités (performance). L'usage d'un outil dépend donc de données objectives (utilité et utilisabilité du système) et de données subjectives (perception de cette utilité et utilisabilité). Mais pour qu'il y ait usage, il doit y avoir intention.

Pour éveiller l'intention d'usage et améliorer la perception des utilisateurs (l'utilité et l'utilisabilité du système), la matérialisation de la zone d'interaction dans l'environnement (symbolisation d'une zone au sol) est une solution recommandée. Ainsi, il est possible d'éviter que les utilisateurs soient perdus (voir *Positionnement des écrans*). Une autre technique consiste également à fournir une zone d'attente permettant aux utilisateurs d'être spectateurs, le temps qu'ils se décident à passer acteurs ou non. Ceci permet d'éviter une transition trop brutale de l'espace privé des utilisateurs vers une exposition en publique (Rogers et Rodden 2003). Cette zone d'attente peut être pensée en termes de zone géographique (zone autour de la zone d'interaction par exemple). Ainsi, on obtient deux zones géographiques (la zone d'interaction et la zone d'attente) permettant de définir des contextes d'usages spécifiques et donc des interactions différentes. Ce critère d'espace géographique permet de borner les différents types d'interactions en fonction de chaque espace (Vogel et Balakrishnan 2004). Mais nous pouvons aussi considérer la zone d'attente comme un espace digital (téléphones portables des utilisateurs, terminaux déportés à proximité de l'écran, etc.) leur permettant d'éditer le contenu avant de le publier sur l'écran (espace public) (Müller et Krüger 2007). Ces deux techniques qui ne sont pas exclusives l'une de l'autre facilitent grandement l'interaction avec les grands écrans en permettant aux utilisateurs de contrôler leur passage entre le domaine privé et l'espace public. De plus, les utilisateurs découvrent souvent l'intérêt et l'existence des fonctionnalités d'un système en regardant les autres interagir (Huang, Mynatt et al. 2006).

Une méthode additionnelle consiste à autoriser les utilisateurs à interrompre de manière simple et rapide les interactions entamées avec le système et ceci sans que cela procure un sentiment de gêne envers les autres utilisateurs. Procurer la sensation qu'un engagement faible suffit pour commencer et arrêter d'interagir avec l'écran met les utilisateurs plus à l'aise et facilite leur intention d'usage avec le système.

La dimension « intérêt » est une spécificité spatiale, c'est-à-dire que l'écran étant placé dans un lieu, il faut qu'il propose des fonctionnalités utiles dans ce lieu et surtout dont les utilisateurs se feront une représentation positive de l'utilisabilité. Il doit également avoir une fonctionnalité bien identifiée et dédiée (O'Hara, Perry et al. 2003, Huang, Mynatt et al. 2006).

## Trop d'espace

Ball et al. (2005), montrent un effet de perte de place sur les écrans de grand format. Dans leur étude, ils s'intéressent à l'utilisation d'un mur d'écrans composé de neuf écrans. L'interaction avec le mur d'écrans est réalisée avec une souris et un clavier. Les utilisateurs opèrent des tâches de bureautique comme le traitement de mail, la lecture d'articles, etc. Ils montrent que pour ce type d'usage l'ensemble de l'affichage n'est pas utilisé et une grande partie de la surface d'affichage est laissée pour compte (Ball et North 2005).

## Recommandations générales pour la mise en œuvre des écrans de grand format

### Positionnement des écrans

Huang (2007), a réalisé une étude de terrain consistant en l'observation de quarante-six grands écrans répartis dans plus de vingt-quatre sites différents (gare, librairie, agence de voyage, bibliothèques, université, cafeteria, musée, épicerie, banque et grands magasins) (Huang 2007). L'enquête montre que les usagers portent plus d'attention aux écrans placés à hauteur des yeux. Les écrans suspendus au plafond et ceux disposés en dessous de la ligne du regard n'attirent quasiment pas l'attention. L'auteur remarque également que l'analyse des flux de passage est une très bonne clef de lecture pour disposer au mieux les écrans. Elle préconise d'ailleurs de bien étudier la direction des déplacements des utilisateurs dans le lieu afin d'optimiser le placement des écrans et ainsi de favoriser l'audience. Elle note également que les objets décoratifs et informatifs à proximité des écrans jouent un rôle important d'attraction de l'audience. Il s'agit donc de tirer bénéfice de leur faculté à attirer le regard afin d'amener les utilisateurs à considérer l'écran et surtout de ne pas compter sur le pouvoir attracteur illusoire des grands écrans. De manière générale et dans la mesure du possible, il faut aménager l'environnement dans lequel est situé l'écran dans le but de maximiser son audience. Rogers & Rodden (2003), précisent qu'il ne faut pas hésiter à dessiner la zone d'interaction dans l'espace physique afin que chaque utilisateur comprenne facilement à partir de quel endroit il peut être acteur ou spectateur avec l'écran (Rogers et Rodden 2003).

### L'importance du contenu

#### *Le format du contenu affiché : un atout majeur ?*

Huang (2007), constate que les vidéos sont plus attractives que les images ou les textes, y compris si ces derniers sont animés (Huang 2007, Owens, Chaparro et al. 2011). Proposer des contenus dynamiques permet de capturer l'audience des utilisateurs plus longtemps.

Cependant, les auteurs ne précisent pas si les vidéos engendrent une durée d'audience plus importante par intérêt des utilisateurs pour ce type de contenu ou au contraire car les utilisateurs peinent à trouver l'information pertinente. Dans plusieurs études, les interviews à posteriori des utilisateurs révèlent qu'ils aimeraient trouver du contenu informatif plutôt que du contenu ludique. Nous pensons donc qu'il est nécessaire de dynamiser le contenu proposé. Néanmoins, l'utilisation de vidéos ne nous semble pas le meilleur support pour du contenu informatif dans un contexte de lieux publics.

Churchill et al. (2003), conseillent d'utiliser des couleurs qui se distinguent facilement dans les conditions d'éclairages locales, d'utiliser des animations, de générer du mouvements dans l'interface et d'utiliser des polices de caractères assez larges afin d'attirer l'attention des utilisateurs lorsqu'ils passent même assez loin de l'écran (Churchill, Nelson et al. 2003). Lorsque des informations sont présentées sur un écran et que des mises à jour ont lieu, il est important de bien mettre en évidence quelles informations ont été actualisées (Müller, Paczkowski et al. 2007). Par exemple, dans le cas de boucles d'affichage, les utilisateurs ne savent pas combien d'informations différentes font parties de la boucle tant qu'ils n'ont pas assisté à une boucle complète. Storz et al. (2006), montrent qu'un écran vierge (non éteint) crée implicitement chez les utilisateurs, le sentiment que du contenu va arriver. Il génère ainsi chez l'utilisateur une attente, une envie de ce contenu. Cependant, ils montrent également qu'un écran éteint laisse penser que le système est en panne (Storz, Friday et al. 2006). Dans le but d'inciter les utilisateurs à interagir avec un écran et surtout afin de comprendre qu'ils peuvent interagir avec celui-ci Müller et al., propose d'étudier comment les utilisateurs se rendent compte qu'ils peuvent interagir avec un écran (Müller, Walter et al. 2012). Ils conçoivent un système qui réagit aux mouvements incidents des utilisateurs dû à leur passage devant l'écran. Ils proposent de faire comprendre à l'utilisateur de passage devant l'écran que celui-ci est interactif à l'aide d'indices visuels. Ils testent plusieurs indices visuels :

- une image miroir de l'utilisateur sur l'écran ;
- une image de la silhouette de l'utilisateur.

Ils relèvent une augmentation de 90% d'interactions suite à l'utilisation de la technique d'image miroir en comparaison avec une technique qui affiche un message incitant l'utilisateur à interagir. Cette augmentation est de 47% avec la technique de la silhouette. Ils montrent également que les utilisateurs mettent du temps à identifier l'interactivité de l'écran (« *landing effect* ») et que souvent ils dépassent l'écran puis reviennent sur leur pas pour interagir avec celui-ci. Enfin, ils constatent l'effet « pot de miel » (« *honeyspot* »), ou effet de lancement selon Whyte (Whyte 1980) qui se traduit par l'attraction des utilisateurs vers l'écran dû à la curiosité de voir un autre utilisateur en train d'interagir comme des abeilles qui s'agglutinent sur un pot de miel.

#### *Contenu dynamique en fonction du contexte*

Les informations affichées en fonction du contexte sont très appréciées des utilisateurs. Cheverst et al. (2000) identifient deux types de contextes (Cheverst, Davies et al. 2000) :

- personnel : préférences de l'utilisateur, intérêts, position GPS, historique, etc. ;
- environnemental : heure de la journée, affluence, etc.

Les informations nécessaires dans ces deux types de contextes sont des informations dynamiques c'est-à-dire qui évoluent très rapidement en fonction de plusieurs paramètres (profil de l'utilisateur, moment de la journée, lieu, etc.). Les écrans doivent être en mesure de présenter ces informations dynamiques. Par exemple en fonction du moment de la journée,

l'écran affiche si un restaurant sert encore et quel menu du jour il propose. Müller et Krüger proposent un système présentant de l'information (publicité) en fonction des profils des utilisateurs (Müller et Krüger 2007). La possibilité d'identifier des profils utilisateurs ou de les connaître à l'avance semble être une solution intéressante pour inciter les utilisateurs à regarder les écrans.

### *Mode attractif*

Dans leurs travaux, Wichary et al. (2005), montrent différentes informations aléatoires dispersées sur l'écran pour attirer l'utilisateur (6 à 8 éléments maximum, compris dans des formes circulaires se déplaçant de droite vers la gauche, dont la durée d'affichage et d'une minute et la vitesse de déplacement faible)(Wichary, Gunawan et al. 2005). Ce mode attractif établit grâce aux retours des utilisateurs permet d'attirer les utilisateurs à l'écran afin qu'ils puissent ensuite découvrir plus de détails sur les informations affichées. Créer un mode attractif qui permet de capturer l'audience attire les utilisateurs vers le système pour obtenir plus de détails. De plus, Churchill et al., exposent que l'engagement de l'utilisateur avec un écran est fortement corrélé à la proximité physique avec celui-ci (Churchill, Nelson et al. 2004).

## L'importance de la navigation différenciée

La vidéo offre un flux de contenu en constante évolution qui est plus apprécié des utilisateurs comme présenté dans la partie *Le format du contenu affiché : un atout majeur ?*. Dans cette même expérience Huang observe que les utilisateurs attachent plus d'importance aux dépliants papiers et tableaux d'affichage physiques présents à proximité des écrans plutôt qu'à ces écrans proposant pourtant les mêmes informations (Huang 2007). Ainsi, lors du parcours des dépliants papiers ou des tableaux d'affichage à proximité de grands écrans, les gens naviguent à travers l'information à des vitesses et des niveaux de profondeur variés. En effet, ils sautent les objets qui ne les intéressent pas et se concentrent davantage sur les autres. La possibilité de naviguer à travers l'information affichée et donc d'avoir une expérience dynamique avec l'information favorise l'audience des utilisateurs envers un affichage. Il est donc important de procurer aux utilisateurs la possibilité de naviguer de manière différente à travers le contenu affiché. L'interactivité du contenu est la clef du maintien de l'audience.

## L'acteur unique

Plusieurs études (Jordan et Henderson 1995, Russell et Sue 2003, Trimble, Wales et al. 2003) montrent que les écrans proposant des interactions pour un seul utilisateur amènent une certaine frustration pour les utilisateurs constituant les spectateurs. Le déploiement d'un grand écran requiert une analyse du besoin en amont afin de correspondre au mieux aux besoins des utilisateurs (Beyer et Holtzblatt 1997, Müller, Paczkowski et al. 2007). Il doit apporter un bénéfice pour l'ensemble des utilisateurs (différents profils)(Grudin 1988). Ces constats vont dans le sens de notre démarche. En effet, nous nous attachons dans nos travaux à mettre en œuvre une interaction multi-utilisateur avec les écrans de grand format.

## Fiabilité et robustesse des systèmes techniques

Pour gagner et garder la confiance des utilisateurs, les systèmes proposés doivent être robustes et fiables. Par exemple, une interface qui se bloque doit afficher un message indiquant qu'elle n'est plus fonctionnelle sinon la frustration générée chez les utilisateurs par

son dysfonctionnement non annoncé va réduire considérablement leur taux de confiance envers le système. Un système qui tombe en panne est acceptable à condition que ce soit explicitement indiqué et que la fréquence des pannes soit faible. De plus, si le système a besoin de faire des mises à jour, il est important de les planifier sur des plages horaires de faible activité afin de déranger le moins possible l'utilisation du système (Cheverst, Fitton et al. 2003, Storz, Friday et al. 2006, Müller, Paczkowski et al. 2007). Nous souhaitons souligner ces recommandations générales qui sont valables pour la majorité des systèmes techniques. En effet, nous devons garder en tête ces règles triviales qui sont finalement souvent laissées pour compte.

## Privilégier les habitudes et la spécialisation fonctionnelle pour la collaboration

Huang et al. (2006), recommandent d'utiliser dans la mesure du possible les fonctionnalités identifiées via les habitudes de travail des utilisateurs (Huang, Mynatt et al. 2006). Dans le cas d'utilisation d'un grand écran par un groupe collaboratif de travail, les auteurs conseillent d'utiliser des fonctionnalités déjà identifiées et servant de support à la collaboration plutôt que d'en suggérer des nouvelles. Pour eux, il est peu probable que les utilisateurs prennent le temps de découvrir par eux même d'autres fonctionnalités s'ils n'en connaissent pas déjà certaines qui leur procurent des avantages conséquents dans leur travail de collaboration. Il est judicieux de présenter les fonctionnalités aux utilisateurs dans un premier temps afin qu'ils identifient clairement le potentiel de l'outil. En effet, dans un contexte de travail collaboratif, les utilisateurs prennent moins de temps pour découvrir de nouvelles fonctionnalités. Enfin, ils conseillent que les fonctionnalités supportant la collaboration proposées par le grand écran soient dédiés à une activité particulière plutôt qu'à des activités de collaboration générales. Les fonctionnalités proposées doivent impérativement aider à la réalisation de tâches cruciales pour les utilisateurs et fournir un avantage clair par rapport aux anciennes méthodes utilisées pour remplir ces tâches. Le *MERBoard* est un bon exemple de l'intérêt de la spécification fonctionnelle pour une adoption et une utilisation du grand écran par les utilisateurs (Trimble, Wales et al. 2003). Notre cas d'étude étant les lieux publics, il est difficile de connaître à l'avance les habitudes des utilisateurs. Néanmoins, ces résultats soulignent que les utilisateurs ne prennent pas le temps de découvrir des fonctionnalités. C'est pourquoi, nous devons nous concentrer sur le développement d'interactions simples et potentiellement déjà connues des utilisateurs.

## Lignes directrices pour la conception d'affichage de grand format

### Introduction aux méthodes de conception d'expérience utilisateur

Nos travaux s'articulent autour d'une démarche de Conception Centrée Utilisateur (CCU). Comme vu dans le *Chapitre 1 : Introduction*, cette méthode consiste à évaluer de manière itérative un système tout au long de son processus de conception afin qu'il remplisse au maximum les besoins et attentes des utilisateurs. L'évaluation du système est réalisée selon deux grandes catégories (DIS 2002) :

- l'évaluation n'impliquant pas la participation des utilisateurs finaux : ces méthodes aident le concepteur à se mettre à la place des utilisateurs et/ou se basent sur son expertise dans le domaine de la CCU. Voici, une liste non exhaustive des méthodes existantes : analyse heuristique, ballade cognitive, persona, évaluation automatisée,

évaluation par expertise, analyse de documents et rapports, approches basées sur les modèles cognitifs de l'utilisateur, méthode de créativité, etc. ;

- l'évaluation impliquant la participation des utilisateurs finaux : ces méthodes impliquent directement les utilisateurs en les confrontant à des maquettes et/ou prototypes du système en conception. Puis, l'analyse du comportement des utilisateurs et/ou de leurs avis recueillis *a posteriori* permet d'apporter les changements nécessaires au système. Voici une liste non exhaustive des méthodes existantes : tests utilisateurs, tri de cartes, entretien, focus group, questionnaire, observation, incident critique, méthode de créativité, etc.

Ces méthodes sont complémentaires et souvent couplées. Carine Lallemand et Guillaume Gronier dans leur livre intitulé « *Méthodes de design UX* » exposent en détail ces différentes méthodes (Lallemand et Gronier 2016).

## Analyse heuristique

L'analyse heuristique<sup>7</sup> consiste à utiliser des listes de critères/règles à respecter dans la conception des interfaces. Leur objectif est de fournir une évaluation ergonomique des interfaces. Ils permettent donc de catégoriser les problèmes d'ergonomie, de juger de leur impact et de trouver des solutions à ces problèmes avant la mise en service du système.

Ces listes de critères et règles se veulent les plus génériques possibles afin d'être utilisés dans la conception de n'importe quelles interfaces Homme-Système. Cette généralité en fait des outils très utilisés. Voici une liste non exhaustive des plus utilisées :

- les heuristiques de Nielsen (Nielsen et Molich 1990, Nielsen 1994, Nielsen 1994) ;
- les critères de Bastien et Scapin (Bastien et Scapin 1992, Bastien et Scapin 1993) ;
- les règles d'or de Shneiderman (Shneiderman, Plaisant et al. 2016) ;
- la norme ISO 9241-110 (DIS 2006) ;
- les heuristiques de Colombo et Pash (Colombo et Pasch 2012) ;
- les heuristiques de Arhippainen (Arhippainen 2013).

De ce fait, ils deviennent de moins en moins spécifiques des types d'interfaces, ce qui empêche d'identifier certains problèmes.

Ainsi, nous remarquons dans la littérature l'apparition d'extensions et surtout la spécialisation de ces différentes heuristiques pour des types d'interfaces donnés.

Il est important de souligner qu'ils existent de nombreux critères et règles mais seulement quelques-uns d'entre eux ont été validés par des études scientifiques. De plus, selon la langue dans laquelle elles ont été validées, une simple traduction dans une autre ne suffit pas à attester de leur validité scientifique. Vallerand (1989) propose une méthodologie de validation transculturelle permettant de valider un questionnaire d'une langue à l'autre (Vallerand 1989). Cette méthodologie repose sur la traduction du questionnaire par des chercheurs trilingues puis le questionnaire fait l'objet d'un processus de traduction renversée et d'une validation par un comité d'experts. En effet, cette méthodologie peut paraître un peu laborieuse mais elle est essentielle car d'une culture à l'autre et d'une langue à l'autre, le sens des mots et leurs interprétations peuvent varier de manière significative.

---

<sup>7</sup> Jean-Louis Le Moigne, dans "La modélisation des systèmes complexes" (Ed. Dunot, 1991), en donne cette définition : "Une heuristique est un raisonnement formalisé de résolution de problème (représentable par une computation connue) dont on tient pour plausible, mais non pour certain, qu'il conduira à la détermination d'une solution satisfaisante du problème."

## Les heuristiques spécialisées aux écrans de grand format

Dans leur étude sur l'élaboration d'heuristiques spécialisées pour les médias ambiants, Mankoff et al. (2003), adaptent les heuristiques de Nielsen et montrent qu'avec leurs heuristiques spécialisées, ils identifient des problèmes sévères non détectés avec les heuristiques de Nielsen (Mankoff, Dey et al. 2003). Somervell et al. (2003), établissent la même année que Mankoff et al., une liste d'heuristiques pour évaluer l'affichage des écrans de grand format. Cependant dans leur article, ils mettent en avant la validation scientifique et empirique de leur démarche contrairement à celle de Mankoff et al (Somervell, Wahid et al. 2003). Ci-dessous, nous présentons une proposition de traduction de la liste des heuristiques promues par Somervell et al (2003). Cette traduction n'a pas été validée scientifiquement mais elle aide à la compréhension globale de notre démarche.

- **Des systèmes de couleurs appropriés peuvent être utilisés pour soutenir la compréhension de l'information :** Essayez d'utiliser des couleurs froides telles que le bleu ou le vert pour l'arrière-plan ou les bordures. Utilisez des couleurs chaudes comme le rouge et le jaune pour la mise en évidence de l'information.
- **La mise en page doit refléter l'information en fonction de son utilisation prévue.** Les informations basées sur le temps doivent utiliser une disposition séquentielle ; les informations thématiques doivent utiliser des dispositions catégoriques, hiérarchiques ou de grille. L'espace d'affichage doit être divisé en fonction de l'importance de l'information.
- **Une utilisation judicieuse de l'animation est nécessaire pour une conception efficace.** Plusieurs animations distinctes doivent être évitées. Indiquez les emplacements actuels et les emplacements cibles si les éléments doivent être automatiquement déplacés dans l'affichage. Introduisez de nouveaux éléments avec des transitions plus lentes et plus lisses. Mettre en surbrillance des informations connexes est une technique efficace pour montrer les relations entre les données.
- **Utilisez des bannières de texte uniquement si nécessaire.** La lecture de texte sur un grand écran prend du temps et nécessite des efforts. Essayez de placer ces bannières en haut ou en bas de l'écran si elles sont nécessaires. Utilisez des polices sans serif pour faciliter la lecture et assurez-vous que les tailles de police sont suffisamment grandes.
- **Indiquez la présence d'informations, mais pas les détails.** Utilisez des icônes pour représenter de plus grandes structures d'information, ou pour donner un aperçu de l'espace d'information, mais pas les informations détaillées ; afficher les détails de l'information est mieux adapté aux interfaces de bureau. La grandeur ou la densité de l'information dicte les mécanismes de représentation (texte vs icônes par exemple).
- **L'utilisation d'écrans cycliques peut être utile, mais il faut être prudent dans leur mise en œuvre.** Indiquez «où» l'affichage en est dans le cycle (c'est-à-dire premier des 5 éléments par exemple ou une barre de progression). Les délais (à la fois pour la présence d'un seul élément et le temps de cycle total) des cycles doivent être appropriés aux informations présentées et permettre aux utilisateurs de comprendre le contenu sans être distrait.
- **Évitez l'utilisation de l'audio.** L'audio est distrayante et, sur un grand affichage public, peut nuire aux autres. En outre, le manque d'audio renforce l'idée de s'appuyer sur le système visuel pour échanger des informations.
- **Éliminez ou masquez les contrôles de configuration.** Les grands affichages publics doivent être configurés une fois par un administrateur. Permettre à plusieurs utilisateurs de modifier les paramètres de configuration de l'affichage peut augmenter

la confusion et la distraction causées par celui-ci. Les modifications répétées de l'interface empêchent les utilisateurs d'apprendre son fonctionnement.

Ces grandes lignes directrices en tête, nous avons conçu un affichage sur écran de grand format qui repose un maximum sur ces règles. L'affichage en question est présenté dans le *Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via head tracking avec les écrans de grand format.*

## Modes et moyens d'interaction avec les écrans de grand format

Comme vu précédemment dans le *Chapitre 1 : Introduction*, les modes et moyens d'interaction nous permettent de communiquer et de contrôler l'interface. Dans son article, Bierz (2006) présente les trois interactions les plus importantes pour agir sur un environnement virtuel (Bierz 2006) :

- la navigation : elle consiste en l'orientation du point de vue dans l'environnement virtuel et en la détermination du chemin à parcourir pour réaliser son action ;
- la sélection : après la navigation vers l'objet désiré vient l'étape suivante, la sélection de cet objet ;
- la manipulation : une fois l'objet sélectionné vient le moment de manipuler l'objet (translation, rotation, etc.).

Dans notre contexte, nous nous intéressons plus particulièrement à la sélection et la manipulation. Comment interagir avec l'écran ou sélectionner de l'information et agir dessus ? Ces deux interactions sont donc l'objectif principal que nous adressons à l'aide des modes d'interaction.

### Proposition de catégorisation des modes d'interactions avec les écrans de grand format

Ils existent de nombreux moyens d'interaction (la souris, le clavier, le stylet, nos mains, nos yeux, *etc.*) avec les écrans de grand format. Dans un premier temps, dans le but de catégoriser les modes d'interaction et de situer notre démarche, nous proposons de définir quatre types d'interaction :

- Interactions directes *versus* indirectes :
  - interactions directes : Il existe une relation directe entre l'entrée<sup>8</sup> et la sortie<sup>9</sup> produite par l'action de l'utilisateur. L'action produit une sortie à l'endroit même où l'entrée a été détectée. C'est le cas avec un écran tactile (Dietz et Leigh 2001, Shen, Vernier et al. 2004) ;
  - interactions indirectes : L'action produit une sortie à un endroit différent de là où l'entrée est détectée. Il existe une distance physique entre l'entrée et la sortie. C'est le cas avec l'utilisation d'une souris.

---

<sup>8</sup> Entrée : donnée envoyée ou action exercée sur le système

<sup>9</sup> Sortie : réponse du système à une entrée

- Interactions instrumentées *versus* libres :
  - interactions instrumentées : ces techniques d'interaction nécessitent la manipulation et/ou le port d'artefacts physiques afin d'interagir avec l'écran. C'est le cas avec l'utilisation d'un stylet sur les surfaces tactiles, l'utilisation du clavier et de la souris sur un ordinateur ou encore les techniques de détection de mouvements (main, tête, regard, corps, etc.) nécessitant le port de marqueurs comme *Optitrack* (NaturalPoint 2017) ou la *Wii mote* de Nintendo ;
  - interactions libres : à l'inverse ces techniques d'interaction ne nécessitent ni le port ni l'utilisation d'artefacts physiques pour interagir avec l'écran. La *Kinect* de Microsoft est un très bon exemple (Kinect 2017).

En croisant ces 4 types d'interaction, nous obtenons une première catégorisation des modes d'interaction comme le montre la *Figure 11* ci-dessous. La zone verte indique la partie dans laquelle nos travaux s'inscrivent.

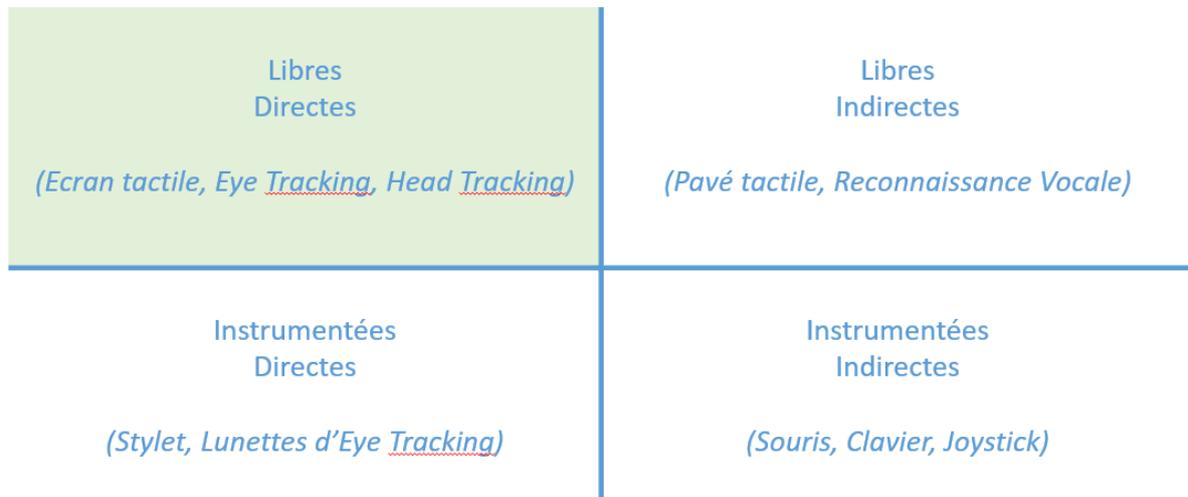


Figure 11 : Catégorisation des modes d'interaction en fonction des types d'interactions.

L'association des différents types d'interaction forment des familles permettant une première catégorisation des modes d'interaction.

Nous considérons l'*Eye Tracking* (détection de la position et de l'orientation des yeux) et le *Head Tracking* dans la famille « Libres/Directes », lorsqu'ils sont utilisés comme modes d'interaction. En effet, nous estimons que la projection dans le plan écran de l'endroit regardé, que ce soit en fonction des yeux ou de la tête, est une interaction directe.

Le pavé tactile et la reconnaissance vocale appartiennent à la famille « Libres/Indirectes » car ces modes d'interaction ne nécessitent pas de porter ou de vêtir de capteurs ou d'artefacts pour interagir avec l'écran et l'entrée produite ne produit pas une sortie à l'endroit physique où elle est capturée (microphone pour la voix et déplacement ou appui du doigt sur le pavé numérique).

Le stylet permet des interactions directes avec l'écran dans le cas où il est utilisé sur la surface tactile d'affichage. Les lunettes d'*Eye Tracking* permettent des interactions directes avec l'écran également. Cependant, ces deux moyens d'interaction nécessitent l'utilisation d'artefacts pour interagir. Nous les considérons donc dans la famille des interactions « Instrumentées/ Directes ».

Enfin, les moyens interactionnels très connus comme le *joystick*, la souris ou encore le clavier sont rangés dans la famille d'interactions « Instrumentées/Indirectes ». Ils nécessitent l'utilisation d'artefact autre que le système d'affichage et les sorties produites par ces systèmes ne s'effectuent pas à l'endroit physique où se sont déroulées les entrées (le mouvement de la souris sur la table déplace la souris sur l'écran et l'appuie sur une touche du clavier génère une action sur l'écran).

Cette première catégorisation réalisée, nous souhaitons apporter une distinction supplémentaire dans la famille d'interactions « Libres/Directes » afin de démarquer nos travaux (suivi du regard) des interactions tactiles. Nous séparons alors cette famille en deux sous-catégories relatives à la distance d'interaction. En effet, la distance d'interaction est le critère qui nous paraît le plus naturel. Ainsi, les deux nouvelles sous catégories sont :

- interactions de contact : ce type d'interaction oblige l'utilisateur à être en contact avec l'écran et donc relativement proche de l'écran ;
- interactions à distance : ce type d'interaction autorise l'utilisateur à être à distance de l'écran.

Comme le montre la *Figure 12*, nous obtenons une catégorisation affinée des modes d'interaction possibles avec les écrans. En vert, la famille d'interactions dans laquelle s'inscrivent nos travaux.

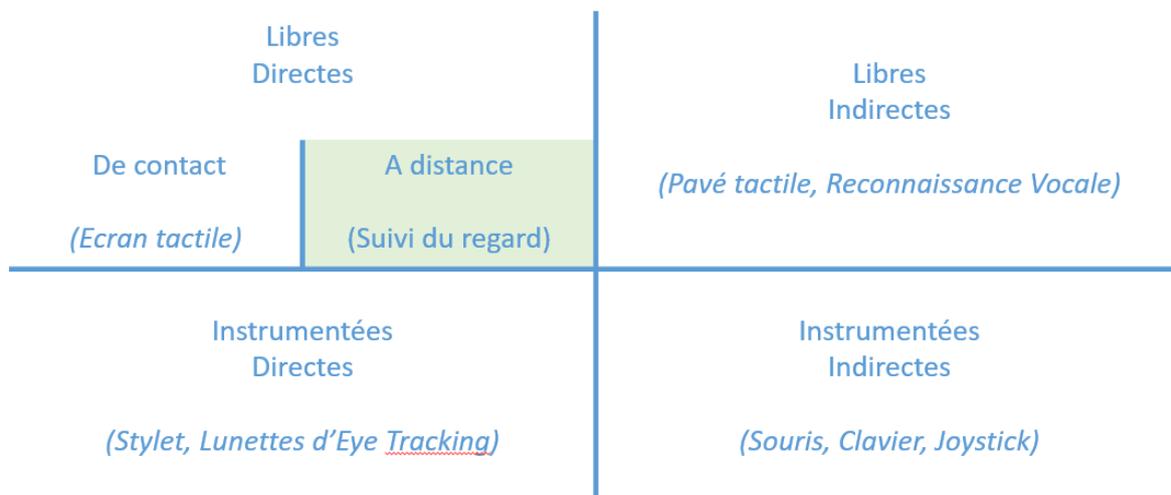


Figure 12 : Catégorisation affinée des modes d'interaction en fonction des types d'interactions.

Fitzmaurice George W. (1997) dans sa thèse de doctorat, présente les différentes étapes d'interaction nécessaires en fonction du type d'interface utilisé (Fitzmaurice et Buxton 1997). Terrenghi Lucia (2007) propose une extension de son concept appliqué aux interfaces tactiles (Terrenghi 2007). La *Figure 13* ci-dessous est l'extension de notre part de leur concept appliqué aux interfaces dont les modes d'interactions sont basés sur le suivi du regard des utilisateurs. Nous nommons ces interfaces, les interfaces attentionnelles.

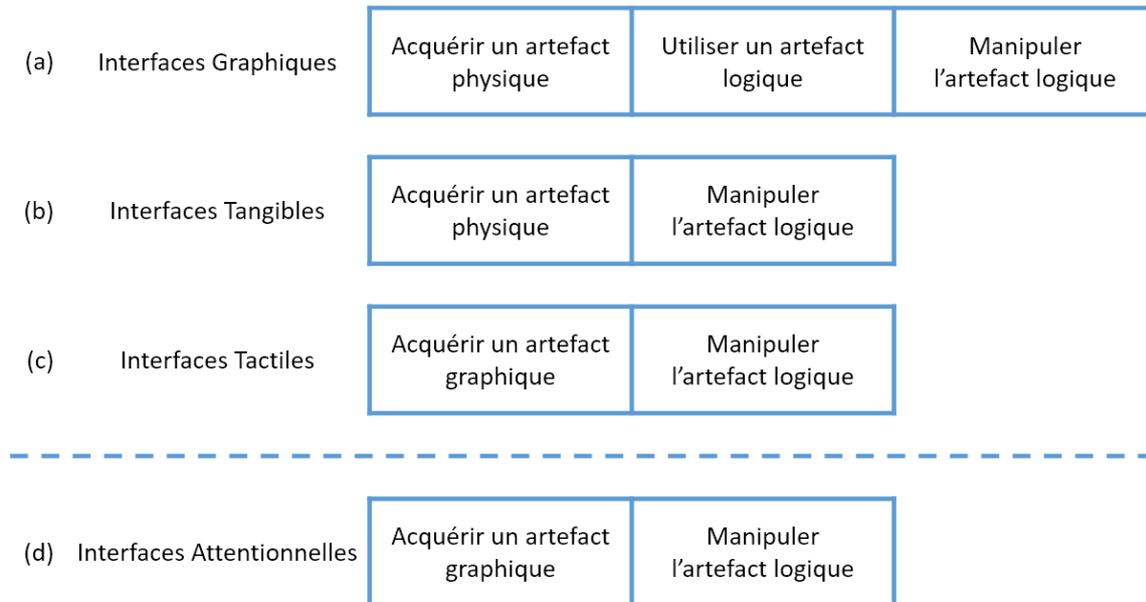


Figure 13 : (a) et (b) représente la réduction du nombre de phases d'interaction vu par Fitzmaurice (Fitzmaurice et Buxton 1997). (c) représente l'extension du concept de Fitzmaurice aux interfaces tactiles par Terrenghi (Terrenghi 2007). La dernière partie (d) représente notre vision du concept de réduction du nombre de phase d'interaction appliqué aux interactions attentionnelles.

Les interfaces graphiques sont les interfaces présentes sur nos ordinateurs. Elles représentent les objets à manipuler sous formes de pictogrammes à l'écran.

Les interfaces tangibles sont des systèmes dans lesquels l'interaction avec l'information numérique est réalisée avec des objets physiques du quotidien. Le but de ces interfaces tangibles est de briser les barrières entre le monde réel et le monde virtuel. De faire du monde qui nous entoure, une interface à part entière afin de connecter le monde réel au monde virtuel. L'idée est d'utiliser les objets physiques du quotidien, dont nous maîtrisons l'utilisation, pour réaliser des actions dans le monde virtuel comme nous le ferions naturellement dans le monde réel en utilisant toute la richesse des différents sens de l'humain. Ainsi, l'intérêt est de profiter des capacités humaines de saisie et de manipulation d'objets en donnant à l'information digitale une forme physique. Ces interfaces tangibles encouragent la collaboration, facilite l'apprentissage et la conception. Cependant, ce mode d'interaction rencontre les mêmes contraintes physiques que le monde réel. Par exemple, il n'est pas possible de sauvegarder une configuration physique d'éléments comme le permet une configuration virtuelle. De plus, l'utilisation d'objets physiques pour l'interaction a pour effet de spécialiser ces objets (un objet → une fonction). En effet, les objets virtuels sont faciles à créer/fabriquer, à modifier, à supprimer ou à copier contrairement aux objets physiques. L'utilisation d'objets physiques rend également difficile le retour en arrière, qui est pourtant très utilisé dans le cas des objets virtuels, ou encore la répétabilité d'une action (Ishii et Ullmer 1997, Ishii 2007, Shaer et Hornecker 2010).

Les interfaces tactiles sont maintenant très répandues avec l'avènement des smartphones et tablettes tactiles. Ces interfaces permettent de s'abstraire du dispositif de pointage (souris pour les ordinateurs) nécessaire aux interfaces graphiques non tactiles des ordinateurs. Ces interfaces impliquent par contre que l'utilisateur soit proche du système.

Enfin, les interfaces attentionnelles, peu déployées pour le grand public (exception faite du domaine des jeux vidéo), utilisent les techniques de vision par ordinateur permettant une détection de la position et du mouvement des yeux ou de la tête par rapport à un capteur (caméra, marqueurs magnétiques, marqueurs infrarouges, etc.). Une fois la position et l'orientation des yeux et/ou de la tête identifiées par rapport au capteur et en connaissant la position du capteur par rapport à l'interface, nous pouvons déterminer de manière plus ou moins précise la zone attentionnelle sur l'écran.

Comme vu dans la *Figure 13*, le choix de ce mode d'interaction nous semble un choix judicieux, car il réduit les phases d'interaction nécessaire avec l'écran. De plus, il a un avantage considérable par rapport aux interfaces tactiles. En effet, il permet aux utilisateurs d'interagir à distance avec l'interface évitant ainsi les problèmes d'obstruction bien connus liés à l'utilisation du tactile. Ce mode d'interaction se démarque également de ceux de la famille « Libre/Indirect ». En effet, utiliser une interface vocale à plusieurs dans un lieu public peu sensiblement causer des problèmes d'embarras social comme expliqué précédemment dans la partie *Embarras social*. De plus, même si les systèmes de reconnaissance vocale ont largement progressé ces dernières années (Siri d'Apple, Cortana de Microsoft ou encore Watson d'IBM, etc.), ils sont encore loin de pouvoir gérer des interactions vocales multi-utilisateurs. Enfin, pour les deux autres familles qui ont en commun l'instrumentalisation des utilisateurs, ces modes d'interaction limitent le nombre d'utilisateur pouvant interagir au nombre d'artefacts présents. De plus en plus de recherches s'intéressent aux interactions possibles avec les grands écrans via nos smartphones. Le principe est de considérer que n'importe quel écran peu devenir un support d'affichage ou nous proposer d'interagir avec lui (Scheible et Ojala 2005, Peltonen, Salovaara et al. 2007, José, Otero et al. 2008, O'Hara, Perry et al. 2013). Nos smartphones s'apparentent ainsi à des télécommandes universelles pour les télévisions. Il s'agit de problématiques que nous n'avons pas choisi d'aborder pendant nos travaux.

## Le suivi du regard comme mode d'interaction

Les techniques de détection de la position et de l'orientation des yeux (Hansen 2003, Jacob et Karn 2003, Poole et Ball 2006, Duchowski 2007, Hansen et Ji 2010, Yan, Wang et al. 2011, Wang, Liu et al. 2015) et de la tête de l'utilisateur (Zhao, Chellappa et al. 2003, Murphy-Chutorian et Trivedi 2009, Perreira Da Silva, Courboulay et al. 2009, Fanelli, Gall et al. 2011, Fanelli, Gall et al. 2012, Ariz, Bengoechea et al. 2016, Ranjan, Patel et al. 2016, Rossi, Leone et al. 2016, Yan, Ricci et al. 2016, Drouard, Horaud et al. 2017, Eichenlaub 2017, Jeni, Cohn et al. 2017, Khalidov et Odobez 2017) sont principalement basées sur les technologies de poursuite optique (*tracking*). Celles-ci sont basées sur des capteurs optiques (caméras) qui déterminent la position et l'orientation de la tête et/ou des yeux avec des précisions, des taux de rafraîchissement (mise à jour des données) et une mobilité des utilisateurs suffisants (Carranza, Theobalt et al. 2003). Une amélioration importante de la méthode de détection est possible à l'aide de marqueurs. Cependant, comme expliqué précédemment, notre but est de ne pas instrumenter les utilisateurs.

Il existe deux types de méthodes de *tracking* optique : les méthodes basées sur la lumière ambiante nommées méthodes passives et les méthodes basées sur une illumination de la scène par une lumière additionnelle (infrarouge/laser/etc.) nommées actives.

- Méthodes actives : L'illumination infrarouge (IR) repose sur la longueur d'onde des faisceaux lumineux compris entre 0,7  $\mu\text{m}$  et 1 mm. Ces systèmes sont composés d'un couplage capteur-émetteur. L'émetteur envoie des rayons IR dans toutes les directions qui illuminent les objets ;
  - calcul de contraste : le capteur (caméra) est capable de détecter des différences de contrastes importantes entre les objets réfléchissant ou non l'IR. Par exemple, pour les systèmes d'eye tracking basés sur l'IR, les rayons se réfléchissent sur la cornée de l'œil humain et entraînent un fort contraste entre la rétine et la pupille. Par alternance des réflexions contrastées de la pupille (clair/foncé), il devient donc possible « d'accrocher » le mouvement de l'œil. Cependant, une calibration préalable du système est nécessaire ;
  - temps de vol : le capteur (caméra) calcule le temps que l'éclair (illumination IR) de la scène prend pour effectuer le trajet entre l'objet et la caméra. Le temps de vol de cet éclair est directement proportionnel à la distance entre l'émetteur-récepteur (très proches l'un de l'autre) et l'objet mesuré. Cette mesure du temps de vol est effectuée pour chaque pixel de la caméra ce qui permet d'obtenir une image complète en 3D de l'objet mesuré. La Kinect V2 de Microsoft utilise cette méthode afin de construire une image 3D de la scène (distance entre les objets et le capteur) et utilise ensuite des méthodes passives pour l'identification de la position et des orientations de la tête. Cette méthode ne nécessite pas de calibrage préalable et permet donc d'obtenir la distance entre les objets et la caméra.

Les deux principaux problèmes de ces méthodes sont : 1) que leur utilisation est limitée à des contextes lumineux d'intérieur et 2) qu'elle nécessite d'être assez proche de l'émetteur pour que l'illumination fonctionne. En effet, la lumière extérieure trop intense réduit considérablement leur précision, voire empêche complètement leur fonctionnement.

- Méthodes passives : Elles sont principalement basées sur des algorithmes de traitements de l'image récupérée via le capteur. Ils existent trois grandes approches :
  - approches reposant sur un modèle (descripteurs) : ces techniques déterminent un modèle de l'objet à détecter, puis tentent de l'identifier dans l'image ;
  - approches reposant sur des éléments remarquables : ces techniques reposent sur l'identification de points caractéristiques (points du visage, zones de couleurs, etc.) ;
  - approches qui fusionnent les deux autres.

La première approche est peu sensible aux variations interindividuelles et permet de poursuivre plusieurs "cibles" dans une même image du fait de la rapidité d'exécution. Ceci est un avantage en ce qui concerne l'aspect multi-utilisateurs et l'absence de calibrage. Par contre, la simplicité des modèles généralement employés pour la tête (ellipse, cylindre ou même plan) produit des résultats peu précis et peu robustes (inclinaison de la tête, occlusions, distracteurs dans l'image, etc.) (Basu, Essa et al. 1996, Birchfield 1997, Schodl, Haro et al. 1998, Russakoff et Herman 2002). Quand le modèle utilisé est plus complexe (modèle 3D de la tête), le problème se pose de sa généralité : comment choisir une tête "standard" ? (Prêteux et Malciu 1998, Zivkovic et Van der Heijden 2001, Yang et Zhang 2002, Luo et Eleftheriadis 2003).

Dans la seconde catégorie de techniques (traque d'éléments remarquables), les détections sont très souvent utilisées pour segmenter l'image et ainsi augmenter les performances en ne traitant que les zones qui contiennent a priori l'image d'une tête. Ceci pose problème pour un contexte multi-utilisateurs ouvert (Birchfield 1998, Hager et Belhumeur 1998, Ahlvers, Rajagopalan et al. 2005, Zhong, Xue et al. 2006).

La seconde et la troisième catégories nécessitent toutes les deux des phases d'apprentissage du fait de l'approche probabiliste employée. Ceci pose à nouveau le problème du choix du modèle employé pour l'apprentissage (tête "standard" ? grand nombre de sujets jugés représentatifs de la population ? Etc.) (Toyama et Horvitz 2000, Ba et Odobez 2004, Ba et Odobez 2011)

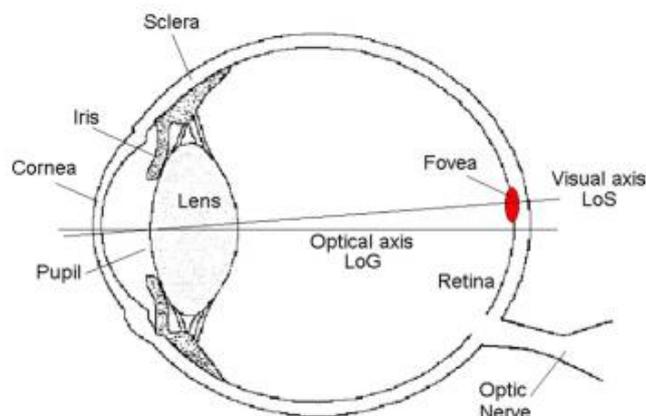
Le principal problème commun à ces techniques de détection est l'occlusion. Par exemple, lorsqu'un utilisateur passe sa main sur son visage ou qu'un autre utilisateur se positionne entre l'utilisateur et le capteur, il devient difficile de capturer la position et l'orientation de la tête et des yeux. Pour pallier ce problème, il existe par exemple des techniques basées sur l'utilisation de filtres de Kalman (Moustakides 1989, Bishop et Welch 2001) ou de Wiener (Wiener 1949) qui déterminent la position et l'orientation de la tête en fonction des précédentes. De plus, le positionnement du capteur par rapport aux utilisateurs est essentiel pour minimiser ce problème. L'utilisation de plusieurs capteurs permet également de résoudre en partie le problème mais la fusion des données issues de nombreux capteurs, notamment en temps réel (traitement vidéo) est un autre challenge.

#### *Suivi du regard : Interactions et paradigmes d'interface*

Le suivi du regard permet de déterminer l'endroit regardé par un utilisateur sur un écran. En effet, une fois la position et l'orientation des yeux et de la tête calculées par rapport à l'écran, si nous sommes en mesure de connaître la distance entre l'utilisateur et l'écran, il est aisé de calculer la projection sur l'écran de la zone regardée au vu des données d'angle de champ de vision de l'être humain (approche trigonométrique).

Mais avant cela, expliquons rapidement comment fonctionne l'œil humain (Hécaen 1972).

Pour voir un objet clairement, nous devons bouger nos globes oculaires dans le but de faire apparaître l'objet sur la petite zone centrale de la rétine appelée fovéa (1° du champ visuel).



*Figure 14 : Structure de l'œil humain, issue de (Morimoto et Mimica 2005)*

C'est pourquoi, la position des yeux fournit une très bonne indication de la zone qu'elle est en train d'examiner dans son environnement. Ils existent deux types de mouvements :

- les mouvements de saccades sont très rapides. Ils sont suivis d'un temps de fixation de l'ordre de 200-600 millisecondes. C'est durant cette période de fixation que les objets peuvent être vus. Mais même durant ce temps de fixation (relativement court), nos yeux continuent de bouger ;
- les mouvements de poursuite sont moins rapides que les saccades. Ils se produisent en réponse à un en mouvement dans notre champ visuel.

Les techniques de suivi du regard s'attachent donc à détecter les fixations et les saccades des yeux de l'utilisateur. Mais elles doivent constamment discriminer les mouvements non volontaires qu'effectuent les yeux afin de ne considérer que les mouvements volontaires et conscients.

### Navigation et sélection d'objet

L'interaction via le suivi du regard (Hutchinson, White et al. 1989, Majaranta et Rähkä 2002, Morimoto et Mimica 2005, Turner, Alexander et al. 2015) s'avère être le dispositif de pointage le plus rapide et le plus naturel. Cependant, l'*eye tracking* et encore plus l'*head tracking* ne permettent pas de sélectionner de très petits objets sur l'interface (Ware et Mikaelian 1987). En effet, lorsque l'utilisateur est à un mètre de l'écran la projection de la zone attentionnelle sur l'écran pour l'*eye tracking* est d'environ 1°, soit un carré sur l'écran mesurant environ 1,7 cm de côté. Pour l'*head tracking* la zone attentionnelle est d'environ 30° du champ visuel, soit un carré de 53,5 cm de côté. Ainsi, en considérant un écran de 24 pouces de diagonales (61 cm) au format 16/9 (53,1 x 29,9 cm) avec une résolution de 1920x1080 pixels, nous obtenons une zone attentionnelle de 89x89px (1° du champ visuel) sur l'écran de précision, ne permettant pas une sélection aussi précise qu'avec une souris. Ce manque de précision, nous amène au constat que les interfaces proposant des interactions via le suivi du regard (*eye tracking* et ou *head tracking*), doivent être adaptées à cette contrainte.

Hyrskykari et al. (2012), présentent deux techniques de sélection utilisant le suivi de regard. La première basée sur le temps de fixation et la deuxième technique basée sur des mouvements particuliers. La première technique ne permet pas de régler les problèmes de précision (l'objet fixé doit être dissociable des autres éléments) et conduit à des interactions non désirées si le temps de fixation nécessaire est trop court ou au contraire à des interactions pouvant être considérées comme trop lentes si le temps de fixation nécessaire pour interagir est trop important. Basée sur des mouvements particuliers et identifiables, la deuxième permet de s'abstraire des problèmes d'interactions non volontaires. De plus, les mouvements des yeux étant très rapides, le temps nécessaire pour réaliser l'interaction est le temps pour réaliser le mouvement particulier. Enfin, au lieu de sélectionner une interaction représentée par un élément visible sur l'interface, il suffit d'associer un mouvement particulier à une fonctionnalité. Ceci implique que l'utilisateur apprenne et retienne ces mouvements particuliers (Hyrskykari, Istance et al. 2012). Bednarik et al. (2012), présentent une méthode permettant de détecter les mouvements intentionnels d'interaction (80% de précision) (Bednarik, Vrzakova et al. 2012). A l'aide de la norme ISO-9241-9, Zhang et al. (2007) comparent trois techniques de navigation et sélection des objets à l'aide du suivi du regard par rapport à la souris, moyen traditionnel d'interaction. Les deux premières sont basées sur des temps de fixation plus ou moins important. La troisième associe la navigation au suivi du regard et la sélection à l'appui sur une touche du clavier. Ils montrent que c'est cette troisième technique la plus rapide et la plus confortable (Zhang et MacKenzie 2007). Dans notre contexte de lieux publics, utiliser des mouvements de tête particuliers et identifiables ne nous semble pas adapté. En effet, ceci implique que les utilisateurs aient connaissance de ces

mouvements en amont. De plus, il nous apparaît que réaliser des mouvements de tête particuliers et identifiables à la vue de tous pourrait engendrer une certaine gêne (cf. *Conserver la zone de confort des utilisateurs*). Nous avons donc préféré mettre en œuvre une interaction basée sur le temps de fixation. Nous expliquons plus en détail les raisons (contexte d'utilisation et tâche à réaliser) dans le *Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via head tracking avec les écrans de grand format*.

### Manipulation d'objet

Il n'est pas naturel de surcharger un canal perceptif tel que la vision pour exécuter des tâches motrices comme la manipulation d'objets (Zhai, Morimoto et al. 1999). Utiliser, les techniques de suivi de regard pour déplacer un curseur comme nous le faisons avec la souris n'est pas envisageable en l'état. Premièrement, les yeux ne sont quasiment jamais immobiles. Deuxièmement, notre vision nous sert avant tout à scanner notre environnement dans un premier temps avant d'agir dessus (deux actions différentes espacées dans le temps). Nous nous retrouvons face à un problème de taille si les mouvements de nos yeux initient perpétuellement une action sur l'interface (déplacement du curseur) sans donner l'occasion de percevoir l'interface au préalable. Cela va à l'encontre de notre manière d'interagir dans le monde réel. Nous devons donc trouver un moyen de permettre à nos yeux de percevoir l'environnement et d'agir dessus seulement quand nous le décidons (Jacob et Karn 2003). C'est pourquoi dans nos travaux, nous avons choisi d'utiliser la position et l'orientation de la tête comme indicateur d'interaction. Ainsi, les yeux peuvent balayer librement l'environnement sans causer d'interaction. De plus, nous avons placé l'initiative du déplacement de l'information du côté de l'écran et plus du côté de l'utilisateur. C'est l'écran qui déplace l'information en fonction de l'endroit identifié comme regardé et de la position de l'utilisateur face à lui.

### Head Tracking versus Eye tracking

Intéressons-nous aux différences entre l'*eye tracking* (détection et suivi du regard) et l'*head tracking* (détection et suivi de la position et de l'orientation de la tête).

### Calibration, lumière, distance utilisateur-capteur

Les techniques de suivi du regard imposent une calibration du système avant utilisation. De plus, les techniques de suivi de regard utilisant des méthodes passives et actives imposent que les utilisateurs soient assez proches du capteur (< 1 mètre de distance). Pour les méthodes actives, cette proximité engendre un assèchement des yeux lié à l'illumination des rayons infrarouge (Schnipke et Todd 2000, Hansen, Hansen et al. 2001) et elles sont assez sensibles aux conditions de luminosité comme vu précédemment. Néanmoins, ces techniques de suivi du regard permettent d'obtenir une grande précision (proche de 1°).

Les techniques de suivi de l'orientation de la tête (Wang, Xiong et al. 2006, Pittman et LaViola Jr 2014, Jacob, Bailly et al. 2016) ne nécessitent pas de calibrage et autorisent l'utilisateur à être assez loin du capteur (jusqu'à 4 mètres pour la Kinect par exemple). Cependant, plus la distance entre l'utilisateur et l'écran est importante, plus la zone d'attention calculée sur l'écran est grande et par conséquent plus la précision de l'interaction est faible. De plus, l'estimation de l'orientation de la tête ne suffit pas à déterminer l'endroit regardé par l'utilisateur. En effet, un utilisateur peut tourner la tête vers la droite et regarder vers la gauche ; mais ce genre de comportement n'est utilisé que pour une prise d'information rapide. Dans le cas d'une consultation (lecture d'information, recherche d'information, etc.),

l'orientation des yeux et de la tête se font naturellement de concert afin de conserver des conditions de confort acceptable.

## Paradigmes d'interface et écrans de grand format

### Problèmes du Multi-utilisateur

Une des difficultés majeures est que les utilisateurs comprennent bien comment interagir à plusieurs avec les écrans. Les utilisateurs doivent rapidement identifier qui a accès à l'écran et qui contrôle quoi. En effet, dans le cas de multiples curseurs affichés à l'écran par exemple, il est important que chacun identifie bien le sien. De plus, il est nécessaire de fournir de multiples indices visuels et auditifs permettant aux utilisateurs de comprendre les interactions en cours (Rogers et Rodden 2003). Néanmoins, dans un contexte multi-utilisateur, l'utilisation d'indices auditifs est vite rendue obsolète. En effet, un son émis par l'action d'un utilisateur peut entraîner une incompréhension pour un autre qui pourrait penser qu'il en est à l'origine. L'émission simultanée du même son pour deux utilisateurs différents engendre une superposition des deux sons qui peut conduire à une inhibition non volontaire d'un des deux sons par l'un des utilisateurs. Les techniques de spatialisation du son sont un recours pour gérer ces problèmes mais dans un contexte très bruyant comme les lieux publics l'utilisation du son nous semble compromise.

### Conserver la zone de confort des utilisateurs

Rogers et Rodden (2003) expliquent que pour faciliter l'engagement des utilisateurs envers le système, il faut fournir des modes d'interaction et des paradigmes d'interface déjà connus et maîtrisés (Rogers et Rodden 2003).

Ils préconisent notamment d'utiliser des interactions moins directes que les gestes ou le tactile, comme des claviers, des souris ou encore des stylets. En effet, l'utilisation des interactions gestuelles dans un contexte multi-utilisateurs, peut amener un utilisateur à être très embarrassé lorsqu'un geste est mal réalisé par exemple. C'est pourquoi, nous n'avons pas développé d'interactions basées sur des mouvements particuliers et identifiables de la tête. L'endroit visé par la tête des utilisateurs et leur position par rapport à l'écran sont des indices permettant à l'écran de réorganiser son contenu.

### Transférabilité des interfaces WIMP ?

Les interfaces Windows, Icons, Menus and Pointing device (respectivement Fenêtres, Icones, Menus et dispositifs de Pointage), associées à l'acronyme WIMP, sont le paradigme d'interface le plus communément utilisé et répandu sur les interfaces graphiques usuellement appelée Graphical User Interface (GUI) de nos ordinateurs. Pour les smartphones nous ne parlons plus d'interface WIMP mais simplement de GUI ou d'interface Post-WIMP car l'utilisation des interactions tactiles a conduit à la disparition du dispositif de pointage (Ex : souris + curseur).

Cependant comme le montrent Ball & North (2005), la navigation devient vite difficile avec une souris et un clavier sur grand écran. Les paradigmes d'interface WIMP sont plutôt dédiés aux écrans de petite taille. Le transfert de ces interfaces WIMP des petits écrans vers les grands n'est pas aussi efficace qu'espéré.

Par exemple, le mouvement de souris nécessaire pour déplacer le curseur d'un bout à l'autre de l'écran est vite rendu difficile lorsque la taille de l'écran est importante. La fonctionnalité de « glisser et déposer » (drag and drop) se trouve également impactée. De plus, il devient très aisé de perdre le curseur de vue lors de ces déplacements très rapides et très amples. Savoir également dans quelle zone de l'écran se trouve le « focus »<sup>10</sup> devient vite pénible. Un autre exemple de paradigme d'interface qui ne s'adapte pas bien des petits écrans aux grands est la fonctionnalité de recherche de mot (ctrl + F). Sur un petit écran cette fonctionnalité surligne et liste les occurrences du mot recherché. La petite dimension de l'écran entraîne alors la nécessité de scroller pour atteindre les différentes occurrences du mot. Sur un grand écran, si la fonctionnalité se comporte de la même manière, l'espace d'affichage étant beaucoup plus grand (pour une même résolution d'affichage), nous risquons de nous retrouver avec un feu d'artifice de mots surlignés et ainsi perdre le principal intérêt de la fonctionnalité. Son principal intérêt étant de discriminer des informations les unes par rapport aux autres, nous nous retrouvons finalement à perdre en qualité de discrimination avec un affichage plus grand. D'autres problèmes tels que l'apparition des boîtes de dialogue ou l'ouverture d'une application dans des zones de l'écran non attendues entraînent ainsi une absence de détection de leurs ouvertures par l'utilisateur (Ball et North 2005).

### Adaptations des interfaces WIMP aux grands écrans

Pour éviter de perdre le curseur, Baudisch et al. (2003) proposent de réduire la longueur des sauts effectués par le curseur en augmentant le taux de rafraîchissement de l'image du curseur afin d'obtenir un mouvement plus lisse du curseur, comme présenté sur la *Figure 15*. Ils montrent de manière significative que le curseur (h) est le plus apprécié des utilisateurs et augmente leur performance (Baudisch, Cutrell et al. 2003).

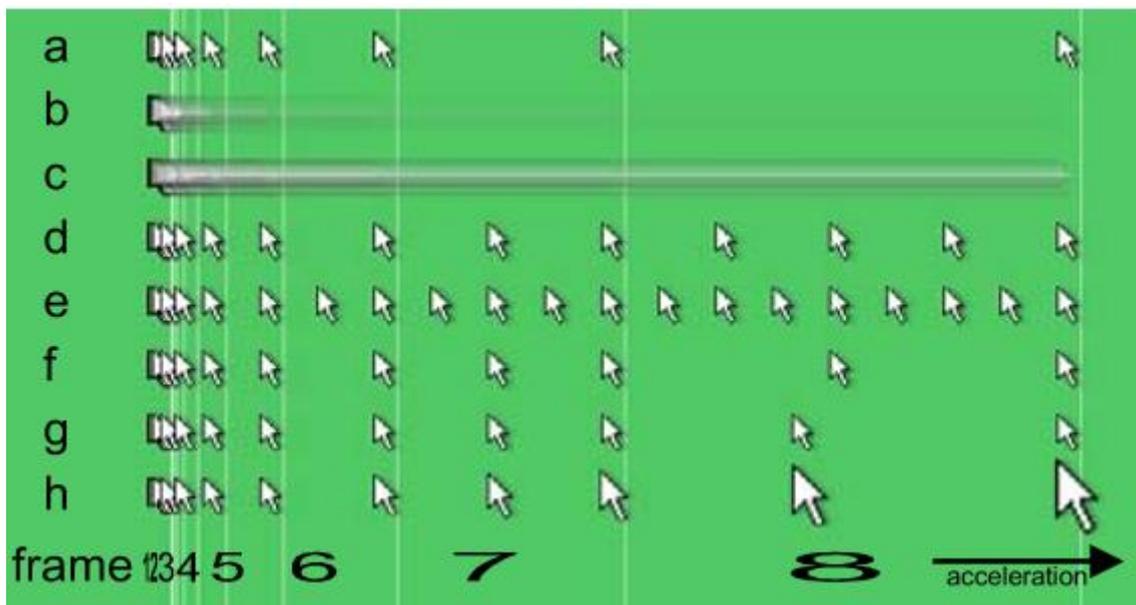


Figure 15 : Proposition de différents comportements du curseur lors d'un déplacement, issue de (Baudisch, Cutrell et al. 2003).

<sup>10</sup> Focus : désigne la cible de saisie dans l'interface utilisateur. L'objet sur lequel s'appliquera la prochaine saisie de l'utilisateur.

La surface d'affichage augmentant (ex : utilisation de deux écrans côte à côte), les utilisateurs ont tendance à ouvrir plus de fenêtres. Ce qui nécessite l'adoption de nouvelles techniques de gestion des fenêtres (Robertson, Czerwinski et al. 2005). Hoffmann et al. (2008), montrent que mettre en surbrillance la fenêtre sélectionnée lors du changement de fenêtre permet de faciliter l'orientation de l'attention de l'utilisateur et d'éviter que celui-ci ne perde le fil (Hoffmann, Baudisch et al. 2008).

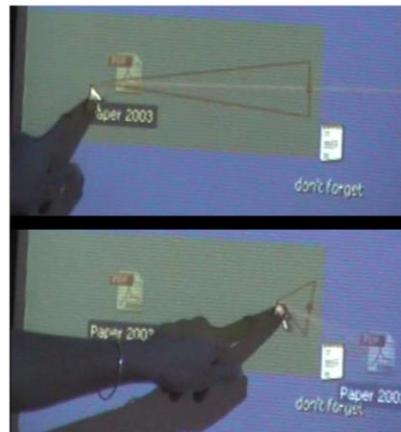
Pour pallier le problème de déplacement sur des longues distances lors de l'utilisation d'écrans de grand format et de modes interactionnels comme la souris et le clavier, les chercheurs proposent différentes techniques nécessitant la combinaison de légers mouvements de souris et l'appui et le maintien d'une touche spécifique du clavier :

- la souris missile : nous lançons le curseur dans une direction avec un léger mouvement de la main, le curseur se déplace de manière continue et peu rapide dans cette direction et il s'arrête lorsque nous bougeons à nouveau la souris ;
- le choix de cible : cette technique permet de sélectionner une fenêtre ouverte parmi plusieurs. Elle a pour vocation de remplacer la fonctionnalité « Alt + Tab », qui n'est plus appropriée lorsque beaucoup de fenêtres sont ouvertes. A la manière de la souris missile, l'utilisateur réalise un mouvement de souris dans une direction (ayant préalablement appuyé sur une touche spécifique du clavier). Ce mouvement de souris directionnel engendre l'apparition d'un rayon sur l'écran qui va sélectionner la fenêtre dont le centre est le plus proche de celui-ci. L'utilisateur peut alors ajuster sa sélection en descendant ou en montant le rayon via de légers mouvements de la souris vers le haut ou le bas ;
- le Drag-and-Pop et Drag-and-Pick : ces techniques ont été développées par Baudisch et al. (2003) dans le but d'améliorer le traditionnel glisser-déposer (Drag-and-Drop), lors d'interactions tactiles sur les écrans de grand format (Baudisch, Cutrell et al. 2003). Lorsque l'utilisateur initie un Drag-and-Drop au lieu de devoir déplacer l'objet sélectionné jusqu'à la destination voulue (par exemple la corbeille), les autres objets présents dans la direction du déplacement initié lors du Drag-and-Drop de l'utilisateur et compatibles (interactions possibles entre eux), apparaissent à proximité de l'objet maintenu par l'utilisateur en révélant une trainée de couleur indiquant leur emplacement d'origine, comme présentée *Figure 16*. Le Drag-and-Pick utilise le même paradigme d'interface que le Drag-and-Pop à la différence qu'au lieu de sélectionner un objet pour le déplacer, le simple fait de déplacer le curseur de la souris génère les mêmes comportements et permet ainsi de lancer une application même si celle-ci est à l'autre bout de l'écran ;



*Figure 16 : Représentation des techniques Drag-and-Pop et Drag-and-Pick, issue de (Baudisch, Cutrell et al. 2003).*

- le Drag-and-Throw est une technique qui permet également d'améliorer la technique de Drag-and-Drop sur grand écran en évitant à l'utilisateur de réaliser de grands mouvements et même de se déplacer. Elle utilise une variante du paradigme d'interface vu pour le Drag-and-Pop et le Drag-and-Pick. Lors de la sélection de l'objet à déplacer une zone rectangulaire apparaît sous celui-ci permettant de contrôler le déplacement de l'objet (direction et distance de déplacement). Cette zone utilise la métaphore d'une baliste. En effet, plus on s'éloigne de l'endroit pointer au départ plus le déplacement de l'objet est important, comme présenté dans la *Figure 17* ;



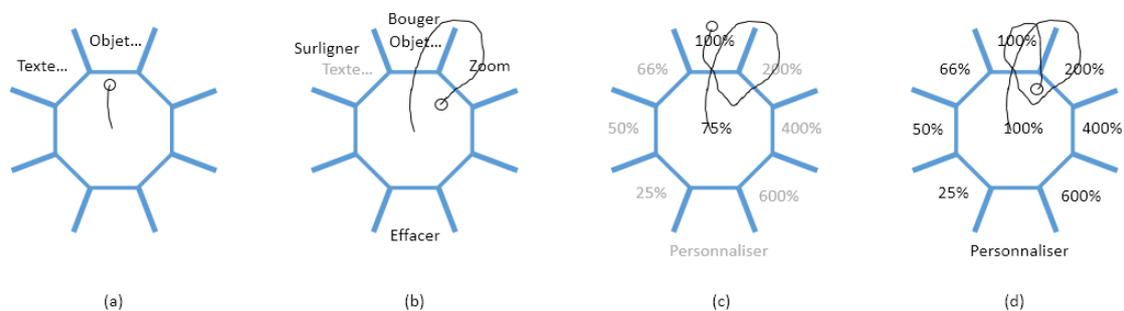
*Figure 17 : Représentation du paradigme d'interface Drag-and-Throw, issue de (Collomb, Hascoët et al. 2005).*

- le Push-and-Pop : Comme le montre Collomb et al. (2005), est la technique la mieux adaptée pour réaliser un Drag-and-Drop sur grand écran (Collomb, Hascoët et al. 2005). Elle consiste à cliquer sur l'objet que l'on souhaite déplacer, il apparaît alors une zone autour de l'objet sélectionné affichant les objets qui lui sont compatibles. Il suffit alors de cliquer sur l'objet désiré pour déplacer l'objet préalablement choisis à côté ou à « l'intérieur » de l'objet désiré. La *Figure 18* montre un enchainement de prise de vue décrivant le mécanisme du Push-and-Pop ;



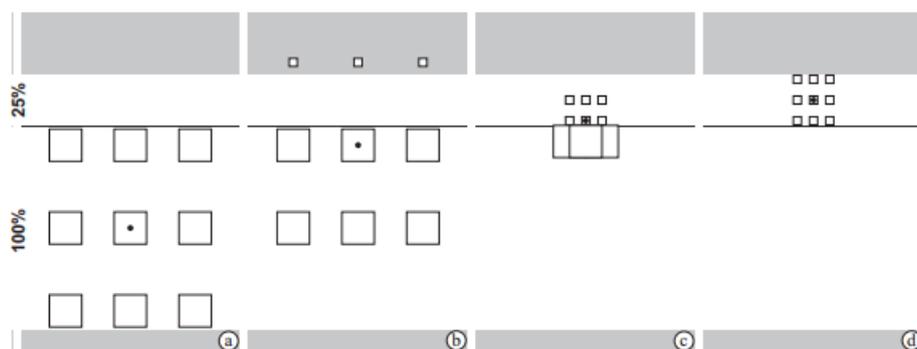
*Figure 18 : Représentation du paradigme d'interaction Push-and-Pop, issue de (Collomb, Hascoët et al. 2005).*

- le Phosphor : est un paradigme d'interface permettant de mettre en évidence durant une courte période les actions préalablement exécutées (surlignage en vert des objets « manipulés »). Baudisch et al. (2006), montre que ce paradigme est très utile lors de tâches collaboratives réalisées sur grand écran (Baudisch, Tan et al. 2006) ;
- le FlowMenu : est une technique développée pour l'interaction avec des stylets. Elle permet d'afficher un menu à l'endroit où le stylet est en contact avec l'interface. Le stylet doit ensuite rester constamment en contact avec la surface pour maintenir le menu visible. Le menu est séparé en 8 zones et les mouvements de va et vient entre la zone centrale et les 8 zones autour permet de sélectionner des fonctionnalités et également dans certains cas de saisir des entrées (textes par exemples). Les entrées sélectionnées apparaissent alors dans la zone centrale et lorsque le stylet est décollé de l'interface, les modifications sont appliquées comme le montre la *Figure 19* ;



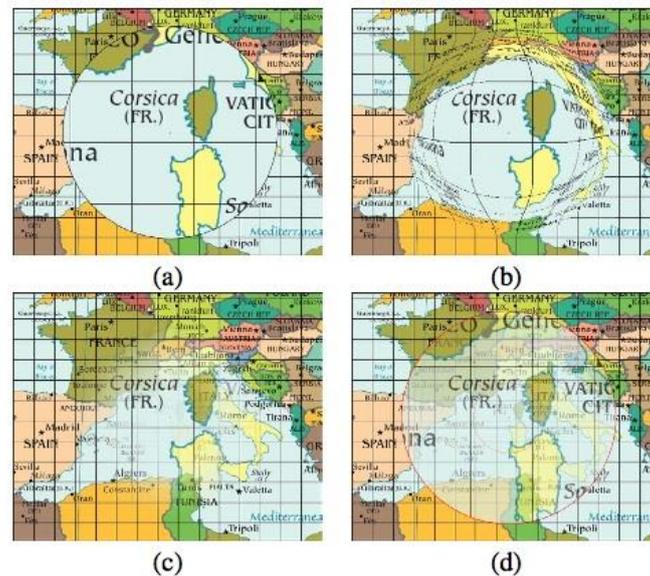
*Figure 19 : Techniques FlowMenu, issue de (Guimbretière et Winograd 2000). (a) pour zoomer l'utilisateur déplace le stylet de la zone centrale vers « Objet... ». (b) Un sous-menu (Surligner, Bouger, Zoom) apparait et la première catégorie du menu (Texte...) est grisée. L'utilisateur choisit « Zoom » et revient vers la zone centrale ce qui fait apparaître la sélection des niveaux de zoom, le zoom actuel est de 75%. (c) L'utilisateur monte et sélectionne la valeur 100%. (d) le retour dans la zone centrale applique le zoom et la valeur 100% est affichée au centre.*

- le ZoomScape : est une technique qui permet de zoomer sur des éléments en fonction de leur emplacement. En effet, la partie supérieure réduit la visualisation des éléments à 25% et la partie centrale affiche ces éléments à 100% comme le montre la *Figure 20* ;



*Figure 20 : ZoomScape technique, issue de (Guimbretière, Stone et al. 2001). De (a) vers (d) l'utilisateur déplace une grille (3x3) carrés vers le haut. Lorsqu'une partie de la grille arrive dans la partie supérieure de l'écran, elle est rétrécie à 25%.*

- le Focus Plus Context : est une technique de visualisation qui affiche une zone de résolution basse (haut niveau de zoom) nommée zone de contexte et une zone de précision (faible niveau de zoom) appelée zone d'attention (Baudisch, Good et al. 2001). Il devient possible de consulter de l'information en détail dans la zone d'attention tout en gardant le contexte dans lequel se trouve cette information dans la zone de contexte, comme le montre la *Figure 21*.



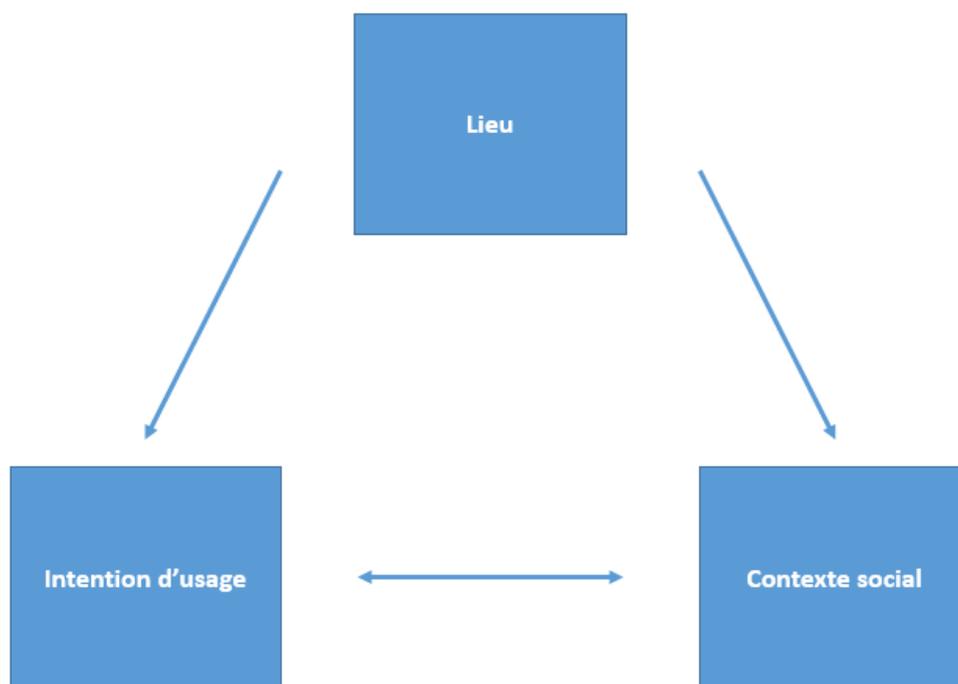
*Figure 21 : Technique SigmaLens utilisant le concept de Focus Plus Context, issue de (Pietriga et Appert 2008).*

Cette présentation des différents paradigmes d'interface utilisés sur les écrans de grand format montrent leur spécificité en termes d'usages. Il convient donc d'étudier le contexte environnemental dans lequel sera placé l'écran et de bien définir à quoi et à qui il va servir pour ensuite choisir les paradigmes d'interface adaptés.

## Conclusion de l'état de l'art

Nous avons pu voir dans cette partie que les grands écrans occupent une place importante dans notre quotidien et que de nombreuses études ont été menées pour améliorer leur diffusion vers le grand public. Ils existent sous différentes formes (verticales et horizontales) et ont des forces et faiblesses clairement identifiées. Nous avons choisi d'étudier les **écrans de grand format verticaux en mode paysage composés d'un seul écran** car de par leur configuration verticale, ils limitent moins rapidement le nombre d'utilisateurs pouvant les consulter et leur déploiement nécessite moins d'espace au sol.

De nombreuses recherches s'accordent à dire que dans un contexte privé, voire semi-privé, l'augmentation de la taille de l'écran est un facteur influençant de manière bénéfique nos performances à la tâche et améliore également nos capacités mnésiques. Dans un contexte public (notre cas d'usage), de nombreux freins (principalement « *display blindness* » et embarras social suscité) à l'utilisation de ces grands écrans apparaissent bien que plusieurs auteurs soulignent leur effet catalyseur sur les interactions sociales. L'analyse des avantages et des inconvénients des grands écrans nous amène à considérer les trois facteurs essentiels influençant l'expérience utilisateur (UX) ; lieu, contexte social et intention d'usage (*Figure 22*).



*Figure 22 : Relations entre les principaux déterminants de l'UX avec les grands écrans.*

Dans la modélisation que nous proposons *Figure 22*, le lieu définit l'endroit physique dans lequel est positionné le grand écran (centre-ville, bureau, espace de vie, etc.). Le contexte social est défini par les us et coutumes en vigueur. Comme vu précédemment (*Figure 10*), l'intention d'usage est dépendante de la représentation de l'utilité et de l'utilisabilité du système que se font les utilisateurs à travers leur perception. Ainsi, le lieu a un impact direct sur l'intention d'usage et le contexte social. Le contexte social va impacter l'intention d'usage

avec les grands écrans et réciproquement l'intention d'usages va altérer le contexte social. Ces relations entre les trois facteurs vont dicter l'expérience utilisateur avec les grands écrans. Dans nos travaux, le lieu est **un espace public**. Le contexte social sera quant à lui défini par l'intention d'usage que nous arriverons à invoquer chez les utilisateurs grâce aux modes d'interaction et aux paradigmes d'interface que nous proposerons.

La majeure partie des travaux portant sur le sujet, concerne des écrans de grand format autorisant des modes interactions ; soit libres, directs et de contact ; soit instrumentés. L'originalité de notre approche est donc de se placer dans **un mode d'interaction libre, direct et à distance** et d'en étudier la faisabilité et l'expérience utilisateur. Ce mode d'interaction libre, direct et à distance devrait permettre de pallier les problèmes d'occlusion de l'information engendrés par les modes d'interaction libres, directes et de contact. Notre choix s'est porté sur un mode d'interaction basé sur la détection de la position et de l'orientation de la tête, l'**head tracking**. Ce mode d'interaction ne nécessite pas de calibrage du dispositif avant chaque utilisation. Il autorise l'utilisateur à se positionner à une distance importante de l'écran. De plus, lorsqu'on interagit avec un écran de grand format utiliser la détection de la position et de l'orientation de la tête permet de concevoir des interactions plus naturelles. Ces formes d'interaction s'apprennent très rapidement et utilisent les mêmes représentations que celles que nous utilisons pour interagir avec le monde physique (Bierz 2006). Ball et al. (2007), montrent dans leur étude que la navigation physique améliore la performance des utilisateurs sur grand écran en comparaison avec la navigation virtuelle (Ball, North et al. 2007). C'est pourquoi nous avons basé les interactions sur la position des utilisateurs par rapport à l'écran. Ainsi, les interactions sont directement reliées à un paramètre physique maîtrisé par les utilisateurs, leur corps. De plus, utiliser les mouvements naturels du corps (position du corps, déplacement de l'utilisateur, etc.) comme paramètres d'entrée pour l'interaction permet d'interagir de manière directe sans avoir recours à des artefacts dédiés à cette navigation.

Ainsi, à partir des données de position et d'orientation de la tête, nous choisissons de calculer la projection du focus attentionnel (endroit visé par la tête) des utilisateurs sur l'écran. Bien que moins précis que les techniques d'*eye tracking*, les utilisateurs seront alors en mesure de « toucher l'écran du regard ». Nous nous attachons ensuite à décrire des paradigmes d'interface viables permettant la sélection et la manipulation des informations présentes à l'écran. En effet, comme nous l'avons décrit, les paradigmes WIMP bien connus sur les petits écrans ne sont pas exportables en l'état sur les écrans de grand format. Les paradigmes d'interface mis en place prendront en considération une **interaction multi-utilisateur**. Ce travail est appuyé par une démarche de **Conception Centrée Utilisateur**. C'est pourquoi dans les chapitres suivants nous présenterons la mise en œuvre de cette démarche en adoptant les heuristiques de Somervell (Somervell, Wahid et al. 2003) tout au long du processus de conception de nos paradigmes d'interface. Puis, nous évaluons l'ensemble de la dimension UX en lien avec notre système en proposant trois expériences basées sur l'analyse du comportement, de la performance et du ressenti des utilisateurs. En effet, mettre les utilisateurs dans la boucle de conception favorise l'acceptation et l'utilisation du système (Cheverst, Fitton et al. 2003, Churchill, Nelson et al. 2004). Norman et Nielsen (2010) critiquent ouvertement les nouvelles interactions développées par les industriels majeures sur les smartphones en soulignant que ces interactions ne suivent pas les standards de conception d'interaction établis par de nombreuses recherches (Norman et Nielsen 2010). C'est pour ne pas tomber sous le coup de cette critique que nous suivrons les standards de conception des interactions avec les grands écrans tant par l'analyse des études préalablement menées que par la mise en place de tests utilisateurs et l'utilisation d'outils comme les heuristiques via une conception centrée utilisateur.

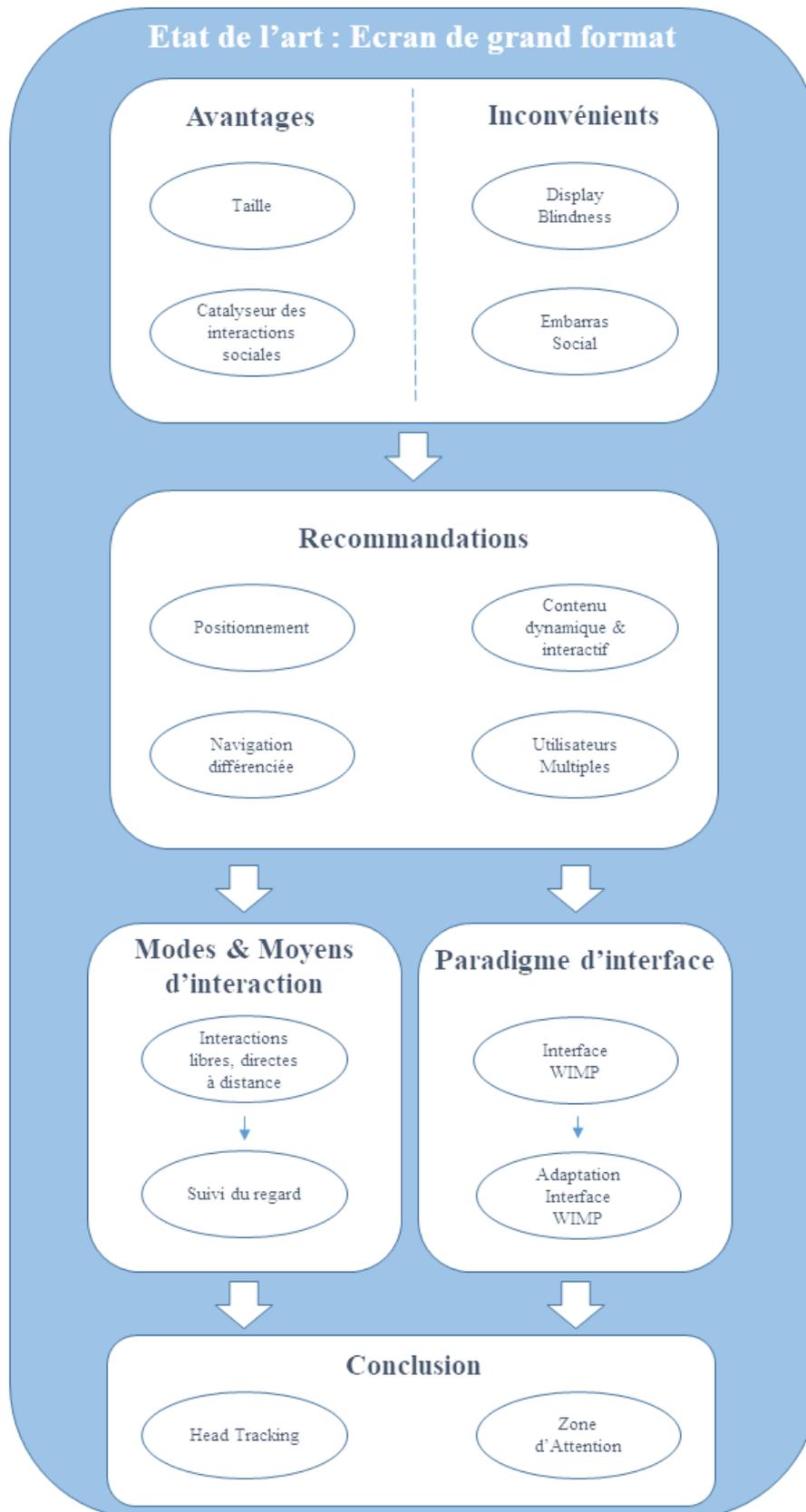


Figure 23 : Schéma général de conclusion de l'état de l'art

# PARTIE II - APPROCHE EMPIRIQUE

---

*« Le savant n'est pas l'homme qui fournit de vraies réponses ; c'est celui qui pose les vraies questions. »*

*- Claude Lévi-Strauss -*

---



# Chapitre 3 : Recherche collaborative d'informations sur écran de grand format

Ce chapitre est consacré à la description de notre première expérience. Dans cette expérience, nous étudions l'influence de la connaissance de la zone regardée par ses pairs sur la performance lors d'une tâche de recherche collaborative d'informations sur écran de grand format. Nous présenterons dans un premier temps la problématique et nos hypothèses. Puis dans un second temps, nous décrirons le protocole expérimental (méthodes et matériels) mis en place afin d'adresser ces premières hypothèses. Enfin, nous présenterons et discuterons les résultats obtenus qui ont guidé notre démarche vers la deuxième expérience.

## Introduction

### Contexte

L'idée de cette première expérience est basée sur l'hypothèse suivante : dans une situation de crise où des collaborateurs ont un même but mais n'ont pas le temps de communiquer oralement car ils sont trop absorbés par leurs propres tâches, connaître ou *a minima* avoir des indices fiables permettant d'inférer sur la prise en charge d'une tâche particulière par un de ses collaborateurs devraient améliorer la résolution collaborative du problème. Savoir quel problème/tâche traite un collaborateur sans avoir à le lui demander permettrait un gain de temps. Cette connaissance évite également les traitements en doublon ou les oublis de traitement. Elle favorise donc une répartition plus efficace de la charge de travail dans le groupe. En d'autres termes, elle améliore la collaboration implicite entre les membres d'un groupe et par conséquent dispense des communications inutiles, tout en conservant une réalisation efficace du but commun.

Mais dans ce cas, quels indices fiables fournir permettant de connaître ou d'inférer sur la tâche traitée par un collaborateur ?

Nous proposons de tester l'influence sur la performance de groupe de la connaissance de l'endroit regardé par ses partenaires sur un écran de grand format. Nous avons choisi comme tâche commune pour le groupe la recherche collaborative d'informations sur un même écran de grand format. Ainsi, la connaissance de l'endroit regardé par ses partenaires sur l'écran est un indicateur permettant d'inférer sur leur recherche en cours et donc d'adapter son comportement et ses actions en conséquence.

Nous pensons qu'un groupe de personnes formant une équipe dont la tâche est de trouver le plus d'informations possibles sur un écran dans un temps imparti, sera plus performant (qualitativement et quantitativement) si chaque individu du groupe a la possibilité de

connaître la zone regardée par ses partenaires sur l'écran dans un contexte où toutes communications orales et gestuelles sont interdites.

Notre hypothèse repose sur la mise en œuvre, volontaire ou non, du mécanisme que nous nommons « vision flottante » par les individus constituant le groupe. Dans la prochaine partie, nous décrivons ce mécanisme.

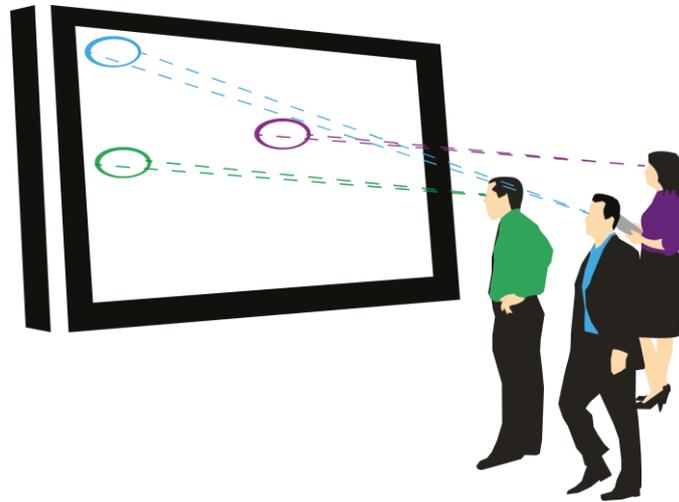
## La vision flottante

Nous désignons par « vision flottante » le mécanisme perceptif et cognitif permettant de diviser ses ressources attentionnelles visuelles de manière volontaire ou automatique. Pour décrire ce mécanisme de vision flottante nous nous appuyons par analogie sur le mécanisme de l'écoute flottante. La vision flottante est à la vision ce que l'écoute flottante est à l'audition. Ainsi la notion de vision flottante, tout comme celui de l'écoute flottante, traduit la faculté de l'humain à diviser son attention. Cette division de l'attention visuelle (Carrasco 2011, Borji et Itti 2013) participe à la construction de la conscience de la situation d'un individu (Endsley 1995, Stanton, Salmon et al. 2017). En effet, la vision flottante permet à un individu de réaliser sa tâche de recherche d'informations tout en traitant d'autres indices visuels, comme par exemple, l'endroit regardé par ses partenaires. Cette conscience de l'endroit regardé par ses partenaires et l'interprétation qui en découle favorise la construction d'une conscience de situation d'équipe (Salas, Prince et al. 1995). Khan et al., (Khan, Matejka et al. 2005) ont exploré une technique permettant de focaliser l'attention d'un groupe de personnes en projetant un faisceau lumineux sur la zone de l'écran nécessitant l'attention de l'audience et en rendant plus sombre le reste de l'écran. Ainsi, ils incitent l'audience à partager une même information. Cette technique favorise l'établissement d'une représentation partagée de la situation (Saner, Bolstad et al. 2009).

Dans le but de générer une situation propice à l'utilisation du mécanisme de la vision flottante et afin de tester l'influence de la connaissance de l'endroit regardé par ses partenaires comme indicateur favorisant la collaboration implicite, nous avons mis en place un paradigme d'interface permettant à chaque utilisateur de percevoir l'endroit de l'écran « regardé » par ses partenaires. Ce paradigme d'interface repose sur l'affichage à l'écran d'une ellipse correspondant à la projection dans le plan image du champ visuel de chaque utilisateur. Cette projection est réalisée en fonction de la position et de l'orientation de la tête de chaque utilisateur, comme expliqué dans la partie

*Conclusion de l'état de l'art.*

Nous parlerons alors de matérialisation de la zone d'attention sur l'écran comme le montre la *Figure 24*.



*Figure 24 : Projection du champ visuel dans le plan écran et affichage de la zone d'attention.*

## Principe de l'expérience

L'expérience consiste en l'affichage d'images très chargées en informations dans lesquelles trois utilisateurs doivent retrouver une liste d'éléments sans communiquer. Les images sont affichées les unes après les autres et sont cadencées par un compte à rebours de 40 secondes. Cette expérience s'apparente à de nombreux jeux de recherche d'éléments comme le jeu des sept erreurs ou encore le célèbre jeu pour enfant « Où est Charlie ? » (Handford 1987).

Comme illustrée à la *Figure 25*, la présentation des images se fait selon deux modalités :

- la modalité d'affichage simple (**S**), qui présente la liste des éléments à retrouver dans la partie haute de l'affichage (bandeau supérieur) et l'image dans laquelle les chercher ;
- la modalité d'affichage avec point d'attention (**PA**), qui présente le même affichage que la modalité **S**, avec en complément la matérialisation du point d'attention de chacun des utilisateurs à l'écran.



Figure 25 : Schéma des deux modalités comparées dans notre première expérience : modalités PA et S.

## Hypothèses

L'objectif est d'étudier l'influence de la connaissance de l'endroit regardé sur l'écran par ses partenaires sur la performance du groupe lors d'une recherche collaborative d'informations sur écran de grand format.

Notre première hypothèse est la suivante : « En modalité **PA**, la performance du groupe est amélioré en comparaison avec la modalité **S** ».

Nous étudions la performance du groupe en termes de recherche collaborative d'informations selon plusieurs critères :

- le nombre total d'objets trouvés sur l'ensemble de la série par le groupe ;
- le temps pour trouver le premier objet ;
- le nombre moyen d'objets trouvés par image ;
- le nombre de mauvais clics réalisés sur l'ensemble de la série par le groupe ;
- le nombre de « tout juste »<sup>11</sup> réalisés sur l'ensemble de la série par le groupe ;
- le nombre de « zéro trouvé »<sup>12</sup> réalisés sur l'ensemble de la série par le groupe.

Nous tirons donc de ces critères de performance, les sous-hypothèses suivantes :

- (H1.1) le nombre d'objets trouvés est supérieur avec la modalité **PA** ;
- (H1.2) le temps pour trouver le premier objet est plus faible avec la modalité **PA** ;
- (H1.3) le nombre moyen d'objets trouvés par image est supérieur avec la modalité **PA** ;
- (H1.4) le nombre de mauvais clics est plus faible avec la modalité **PA** ;
- (H1.5) le nombre de « tout juste » est plus important avec la modalité **PA** ;
- (H1.6) le nombre de « zéro trouvé » est plus faible avec la modalité **PA**.

<sup>11</sup> « Tout juste » : le groupe a trouvé tous les éléments à retrouver sur une image.

<sup>12</sup> « Zéro trouvé » : le groupe n'a trouvé aucun élément à retrouver sur une image.

Nous portons une attention particulière à une seconde hypothèse concernant l'ordre de passation de l'expérience. Nous pensons que les groupes commençant l'expérience avec la modalité **PA** (ordre PA-S) auront une performance plus élevée en comparaison avec les groupes démarrant avec la modalité **S** (ordre S-PA). En effet, l'affichage de l'endroit regardé par ses partenaires durant la première série d'image peut amener chaque sujet à se construire une représentation de la stratégie de recherche des autres, d'adapter sa propre recherche en conséquence et de poursuivre ce comportement dans la deuxième série où les points d'attention de chaque participant ne sont plus affichés. Ce comportement n'est pas possible dans l'autre sens car, sur la première série, aucun indice ne permet de se construire une représentation de la stratégie de recherche des autres.

Ainsi nous formulons les sous-hypothèses suivantes :

- (H2.1) le nombre d'objets trouvés est supérieur pour les groupes ayant commencé l'expérience avec la modalité **PA** ;
- (H2.2) le temps pour trouver le premier objet est plus faible pour les groupes ayant commencé l'expérience avec la modalité **PA** ;
- (H2.3) le nombre d'objets trouvés par image est plus important pour les groupes commençant en modalité **PA** ;
- (H2.4) les groupes commençant en modalité **PA** réalisent moins de mauvais clics ;
- (H2.5) le nombre de « tout juste » est plus important pour les groupes commençant en modalité **PA** ;
- (H2.6) le nombre de « zéro trouvé » est plus faible pour les groupes commençant en modalité **PA**.

# Protocole expérimental

## Description Technique

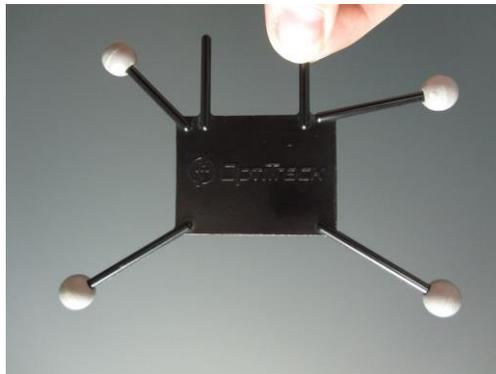
### Head tracking

Afin d'afficher le point d'attention à l'écran de chaque utilisateur extrapolé de la position et l'orientation de la tête des utilisateurs, nous utilisons un dispositif OptiTrack (NaturalPoint 2017). Nous avons utilisé un capteur V-120:Trio composé de trois caméras et de capteur-émetteur d'infrarouge comme le montre la *Figure 26*.



*Figure 26 : OptiTrack – V120:Trio : traqueur optique six degrés de liberté.*

Cette technologie basée sur des méthodes actives (cf. *Le suivi du regard comme mode d'interaction*) permet la détection très précise dans l'espace de la position de marqueurs (cf. *Figure 27*) réfléchissant la lumière infrarouge mais dans des conditions de luminosité contrôlées (contexte d'utilisation en intérieur sous lumière artificielle).



*Figure 27 : Ensemble de marqueurs constituant un Rigid Body.*

A l'aide du Software Development Kit (SDK) et du logiciel Motive:Tracker, il est possible de grouper des marqueurs (minimum 3) afin de définir un élément 3D unique appelé *Rigid Body* : une constellation de marqueurs peut être détectée et traquée par le système, elle est identifiée par sa topologie qui doit être unique. L'hypothèse de détection est basée sur le fait que cette constellation identifie un corps rigide (les distances inter-marqueurs sont constantes), d'où le terme de *Rigid body*. Le logiciel Motive:Tracker calcule alors le barycentre du *Rigid Body* et fournit ainsi sa position et son orientation dans l'espace en temps réel comme le montre la *Figure 28*.

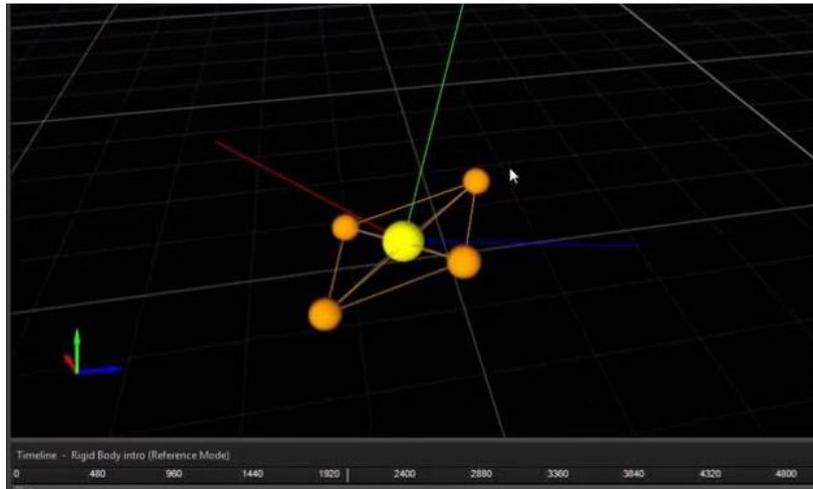


Figure 28 : Exemple d'un Rigid Body composé de 4 marqueurs dans Motive:Tracker.

En s'appuyant sur le SDK, Pierre-Alexandre Favier (co-directeur de cette thèse) a développé bibliothèque de fonctions permettant d'extraire les positions et orientations de plusieurs *Rigid Body* en même temps et en temps réel du logiciel propriétaire. Cela nous permet d'utiliser ces données dans notre propre programme informatique (utilisé pour réaliser l'expérience). Nous avons alors fixé plusieurs marqueurs sur trois casques dans le but de définir trois *Rigid Bodies* uniques dans leur forme comme le montre l'image ci-dessous.



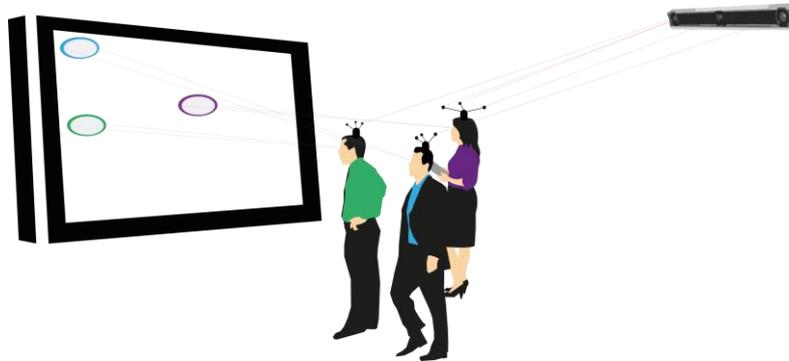
Figure 29 : Casque muni de marqueurs.

L'utilisation de plusieurs marqueurs a pour intérêt de s'abstraire de problèmes d'occlusion liés à la configuration physique du capteur V120:Trio (3 caméras alignées dans le même plan). Ainsi, en plaçant les casques sur la tête des utilisateurs nous sommes en mesure de détecter et suivre de manière précise la position et l'orientation de leur tête.

### Point d'attention

L'étape suivante consiste à projeter dans le plan écran le champ visuel des utilisateurs. Nous avons donc placé 3 marqueurs sur le coin supérieur gauche de l'écran afin de définir un *Rigid Body* fournissant la position et l'orientation de notre plan écran. Le capteur V120:Trio est fixé au plafond derrière les utilisateurs. Le capteur et le *Rigid Body* identifiant l'écran étant fixes, ils définissent notre espace de référence. Par calculs trigonométriques, nous affichons sur l'écran la projection du point d'attention de chaque utilisateur. L'ellipse représentant le point d'attention de chaque utilisateur est transparente et munie d'une double bordure, noire et

blanche, afin d'être toujours visible quelle que soit la couleur du fond des images affichées. Les dimensions de l'ellipse sont calculées afin de représenter la zone de discrimination des couleurs, soit 60° en horizontale de part et d'autre de l'axe optique et de 30° en verticale vers le haut et 40° vers le bas (AW et ROUHET 2004). La *Figure 30* décrit le dispositif.



*Figure 30 : Schéma décrivant le dispositif d'Head Tracking à l'aide de la technologie Optitrack.*

## Programmes informatiques pour l'expérience

Pour réaliser notre expérience nous avons développé deux programmes informatiques. Le premier programme est un éditeur d'image permettant de sélectionner des parties d'une image à l'aide de la souris (sélection classique). A partir de cette sélection, il crée l'image correspondante et référence dans un fichier XML la taille et la position de cette sélection. Grâce à ce programme, nous sommes en mesure d'identifier des zones précises sur une image (zone dans laquelle se trouve l'objet à retrouver sur une image lors de l'expérience) et d'obtenir une image de cette zone (objets à retrouver par les utilisateurs).

Le deuxième programme permet de gérer le déroulement entier d'une expérience, choix de l'ordre des modalités, numéro du groupe, acquisition des données dans un fichier, calcul du score des groupes, etc. Les spécifications fonctionnelles ainsi que les maquettes des écrans de ce programme sont disponibles en annexes (*cf. Annexe II : Spécification fonctionnelle du logiciel de passation (première expérience)*). L'importation dans le programme de passation d'expérience du fichier XML créé dans le programme d'édition d'images nous a permis de placer des zones transparentes cliquables sur les objets à retrouver à l'endroit exact où ils sont localisés dans l'image.

## Déroulement de l'expérience

L'expérience se découpe en quatre temps. Le premier temps est réservé au recrutement et à la passation du test psychologique informatisé Extended CSA-WA (*cf. Indépendance et dépendance à l'égard du champ*). Le deuxième temps, s'organise autour de l'accueil des participants, une brève présentation de l'expérience, la signature du formulaire de consentement libre et éclairé (*cf. Annexe I : Formulaire de Consentement libre, éclairé et exprès (première expérimentation)*), la calibration du système et la présentation en séquence de deux images tests (une en modalité **S** et une en modalité **PA**). Le troisième temps consiste en la passation de l'expérience, présentation de 15 images dans une modalité puis 15 images dans l'autre (*cf. L'expérience*). Enfin, le quatrième temps sert au recueil informel du ressenti immédiat des participants.

## Recrutement

La majeure partie du recrutement de cette expérience a été organisée et gérée par Florian Tual et Yannis Marie dans le cadre de leur projet de fin d'études de troisième année du cycle d'ingénieur de l'Ecole Nationale Supérieure de Cognitique. Nous les remercions grandement pour leur contribution à cette expérience.

## Indépendance et dépendance à l'égard du champ

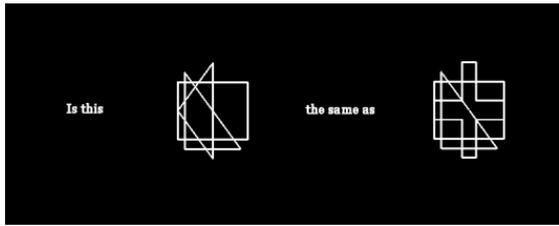
Définissons dans un premier temps ce qu'est un style cognitif. Witkin (Witkin 1965) définit la notion de style cognitif ainsi : « Les recherches récentes ont montré que les gens ont un mode de fonctionnement caractéristique et cohérent dans leurs activités intellectuelles et perceptives, ces styles cognitifs, comme on les a appelés, sont les manifestations dans la sphère cognitive de dimensions encore plus larges du fonctionnement personnel, dimensions que l'on retrouve dans les divers secteurs du psychisme. »

La dépendance indépendance à l'égard du champ quant à elle est une dimension qui permet de distinguer les individus selon leur capacité à percevoir un élément séparé de son contexte et à adopter une attitude analytique dans la résolution de problème (Huteau 1975). Son nom provient de la première expérience ayant mis en évidence cette dimension : il s'agissait d'une expérience sur la perception de la verticalité en l'absence d'indices visuels, elle interrogeait donc l'existence d'une dépendance au champ gravitationnel. Les expériences ultérieures ont prouvé l'extension de ce caractère de dépendance/indépendance à un référentiel dans d'autres processus cognitifs, bien au-delà de la perception de la verticalité.

Cette mesure des styles cognitifs est essentielle afin de former des groupes d'utilisateurs les plus homogènes possibles en rapport avec les différentes stratégies d'exploration visuelles mises en place : les stratégies et les performances d'un individu dans une telle tâche sont liées à sa dépendance au champ (Huteau 1975). En effet nous avons formé les groupes d'utilisateurs de sorte que les scores du ratio de la dimension d'indépendance et dépendance à l'égard du champ de chaque groupe soient les plus proches possibles. Ainsi, nous contrôlons l'introduction d'un biais lié à la capacité de certains groupes à être beaucoup plus performants à la tâche que d'autres grâce à leur style cognitif.

La première partie consiste donc à faire passer le test psychologique informatisé Extended CSA-WA (Peterson et Deary 2006) qui permet de mesurer de la dimension globale-analytique d'un individu. Ce test est une extension (Peterson, Deary et al. 2003) du test « Cognitive Styles Analysis » (CSA) de Riding (Riding et Sadler-Smith 1997). Il est composé de quatre-vingt questions : 40 pour la dimension globale et 40 pour la dimension analytique. Pour mesurer le style cognitif dominant d'une personne, le test compare le temps moyen de réponse mis par l'utilisateur pour les tâches évaluant la dimension globale à celui réalisé en répondant aux tâches analytiques (Peterson, Deary et al. 2005). Ce logiciel fournit donc après passation du test un ratio (score globale/analytique) permettant de déterminer le style cognitif dominant d'une personne. Ce score traduit ainsi la dimension globale-analytique et plus particulièrement la dimension de dépendance indépendance à l'égard du champ. La *Figure 31* présente un exemple des deux types de tâches issues du test Extended CSA-WA :

(a) Tâche évaluant la dimension globale



(b) Tâche évaluant la dimension analytique

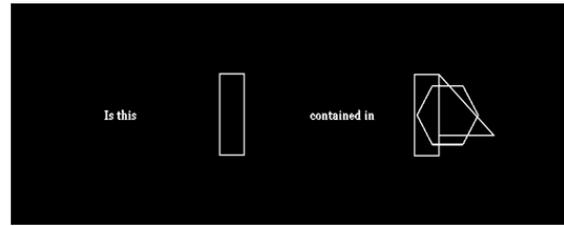


Figure 31 : Exemple de tâches à compléter issues du test Extended CSA-WA (Peterson et Deary 2006). (a) la réponse est « non » car les deux figures ne sont pas les mêmes. (b) la réponse est « oui » car la figure de gauche est présente dans la figure de droite.

## Configuration physique

L'expérience se déroule dans un environnement contrôlé. Les participants sont placés volontairement près de l'écran dans une zone de 3,50 mètres de longueur et de 1 mètre de largeur par rapport à l'écran. Aucune consigne les empêchant de se déplacer dans cette zone n'est formulée. Cette zone est délimitée physiquement par l'écran et le mobilier de la salle d'expérimentation. Cette contrainte garantit que les sujets restent dans la zone de détection. Elle est nécessaire afin que le parcours visuel complet de l'image puisse être réalisé « confortablement » seulement en bougeant la tête et non par un simple balayage du regard (tête immobile). Cette proximité à l'écran impose ainsi le recours à des mouvements assez amples pour être détectables par notre dispositif de *Head Tracking*. Des mouvements plus ténus nécessiteraient l'utilisation d'un système d'*Eye Tracking* multi-utilisateurs, ce qui serait techniquement complexe et coûteux à mettre en œuvre. D'autre part, nous nous intéressons à l'influence de la connaissance de la zone de l'écran observée par les pairs, pas nécessairement sur un point précis de l'affichage. La précision apportée par notre dispositif est donc suffisante à la mise en évidence de l'impact de la matérialisation de la zone d'attention sur l'élaboration de stratégies de recherche visuelle.

Les conditions de luminosité sont également maîtrisées du fait de l'utilisation d'une technologie basée sur des émetteurs et capteurs infrarouges et afin de garantir la stricte équivalence des conditions de passation entre les groupes. Une souris est placée devant les candidats pour qu'ils puissent cliquer sur l'élément trouvé dans l'image. Quand le clic désigne bien un élément à trouver, la zone cliquée correspondante est mise en surbrillance. Simultanément, l'élément trouvé disparaît de la liste des objets à rechercher au niveau du bandeau supérieur. Dans le cas contraire, le clic est comptabilisé comme un mauvais clic et il ne se passe rien à l'écran. A chaque clic des participants, une fenêtre événementielle apparaît sur l'écran de l'expérimentateur qui doit alors renseigner le numéro de l'utilisateur ayant cliqué. Nous filmons le déroulement entier de chaque expérience à l'aide d'une caméra de type GoPro placée derrière les utilisateurs face à l'écran. Enfin, une calibration du dispositif est réalisée avant chaque passation. Cette calibration consiste à identifier l'ensemble des marqueurs disposés sur chaque casque porté par chaque utilisateur pour définir des *Rigid Bodies* uniques. Nous pouvons nous dispenser de cette calibration mais nous avons choisi de la réaliser à chaque fois afin de s'assurer de la robustesse du *tracking* et de l'unicité des *Rigid Bodies*.

## Consignes

Dans le but d'introduire le moins de biais possible, les consignes sont affichées à l'écran et doivent être lues par les utilisateurs. Leur énoncé est donc strictement identique d'un groupe à l'autre. Elles sont disponibles en annexes (*cf. Annexe III : Consignes affichées à l'écran (première expérience)*). En résumé, les consignes sont les suivantes :

- les utilisateurs n'ont pas le droit de communiquer entre eux via la parole, le geste ou le regard ;
- plus les utilisateurs trouvent rapidement des éléments « à retrouver » dans une image plus ils marquent des points ;
- ils ont 40 secondes par image pour trouver un maximum d'objet ;
- une transition affichant un compte à rebours de 3 secondes est présentée entre chaque image ;
- plus les utilisateurs cliquent sur des éléments qui ne sont pas à retrouver sur l'image, plus ils perdent des points.

Afin de rendre l'expérience plus ludique et plus engageante, un score collectif est calculé durant l'expérience. Ce score est présenté comme un score d'équipe prenant en compte la date du clic, le nombre d'objets trouvés et le nombre de mauvais clics. Le mauvais clic n'enlève en fait pas de points au score mais la consigne dit explicitement le contraire afin d'éviter les stratégies de clics à l'aveugle pour trouver un maximum d'objets.

## L'expérience

Les groupes sont composés de trois individus pour que le partage de la surface d'affichage ne soit pas implicite.

Nous présentons dans un premier temps deux images tests en séquence, une image avec la modalité **S** et une image avec la modalité **PA**. Ces deux images servent de repères aux utilisateurs afin qu'ils aient bien compris la tâche à réaliser et le principe de déroulement de l'expérience. Le début de l'expérience est annoncé par une phrase sur l'écran. L'expérience démarre immédiatement, nous présentons alors deux séries de quinze images par groupe. Une série avec la modalité **S** et une série avec la modalité **PA**. L'ordre des séries, et donc des modalités, est contrebalancé pour chaque groupe d'utilisateur. A l'intérieur d'une modalité, l'ordre de présentation des images est conservé pour l'ensemble des groupes. Les numéros de groupe pairs commencent avec la modalité **S** et inversement pour les numéros de groupes impairs. Sur chaque image, les utilisateurs doivent retrouver six éléments en quarante secondes. Les six éléments à retrouver sont affichés au-dessus de l'image et placés en nuage pour éviter une répartition implicite des éléments entre les individus.

De plus, les utilisateurs sont placés près de l'écran et donc pour éviter les gênes éventuelles, au niveau de la nuque, causées par des mouvements trop répétitifs de haut en bas, les objets à retrouver sont placés dans un bandeau en haut de l'écran ainsi que le compte à rebours de quarante secondes. Sur les six éléments à retrouver, il y a toujours un élément non présent dans l'image, un "leurre", afin d'inciter les participants à chercher pendant toute la durée de présentation d'une image. Entre chaque présentation d'image un écran présentant un compte à rebours de trois secondes s'affiche. De plus, avant le passage d'une série à l'autre, les utilisateurs sont avisés du changement de modalité pour la série suivante, pendant vingt secondes, avec affichage du point d'attention si le groupe a commencé par la modalité **S** ou disparition du point d'attention si le groupe a commencé avec la modalité **PA**. Concernant,

l’affichage des points d’attention, nous avons choisi de leur attribuer la même couleur pour les trois sujets afin qu’il n’y ait pas de stratégie compensatoire vis à vis d’un seul individu du groupe (en cas d’a priori sur la performance des autres membres). A la fin de l’expérience, un débriefing informel est réalisé. L’expérimentateur demande aux sujets comment ils ont vécu l’expérience, s’il y a des choses qu’ils souhaitent évoquer.

Il est important de noter qu’aucun utilisateur ne s’est déplacé dans la zone plutôt restreinte durant l’expérience bien qu’aucune consigne n’ait été formulée dans ce sens. Les utilisateurs sont restés en ligne et ont conservé l’ordre de placement du début à la fin de l’expérience. L’utilisateur numéro un est toujours à gauche, le numéro deux au centre et le numéro trois sur la droite de l’écran.

## Matériel

Les images sont projetées sur un écran de 2.40 mètres de base horizontal, au format 16/9, permettant une projection par l’arrière. Un capteur Optitrack V120 Trio placé derrière les utilisateurs, au plafond de la salle, détecte les mires infrarouges placées sur les casques portés par chacun des participants. L’expérience est filmée à l’aide d’une caméra GoPro afin de lever les éventuelles ambiguïtés durant le dépouillement des données. Le corpus d’images présentées est de type varié : dessins, photos, infographies, nuages de tags. L’utilisation de différents types d’image dans le corpus est en adéquation avec les guides de conception évoqués par Alt et al., (Alt, Schneegaß et al. 2012). Toutes les images utilisées sont libres de droits et vont servir à constituer une banque de données qui sera mise à disposition.

## Participants

Cinquante-quatre personnes ont participé à l’expérience : dix-huit groupes de trois personnes (n=18), majoritairement des étudiants. Les groupes ont été formés au hasard des disponibilités des individus en veillant tout de même à respecter l’homogénéité des ratios de chaque groupe concernant le style cognitif d’indépendance et dépendance au champ. Comme expliqué précédemment les groupes pairs ont commencé l’expérience avec la modalité **S** et les groupes impairs avec la modalité **PA**.

*Tableau 2 : Description de la population et de la répartition selon l’ordre de passation des modalités.*

	<b>Age</b>	<b>Sexe</b>	<b>Main dominante</b>	<b>Œil directeur</b>
<b>Population</b>	26 ± 8,6 ans	25F & 29H	38D & 16G	38D & 17G
<b>Groupes pairs</b>	24,5 ans	14H & 13F	19D & 8G	19D & 8G
<b>Groupes impairs</b>	28 ans	15H & 12F	19D & 8G	19D & 8G

# Résultats

## Analyses des données

Les données ont été analysées avec le logiciel XLSTAT version 18.01.35929, logiciel développé par Addinsoft et s'intégrant dans Microsoft Excel. Nous prendrons comme seuil de significativité  $\alpha = 0,05$  pour l'étude statistique.

## Modalité PA versus Modalité S

### H1.1 : Nombre total d'objets trouvés (NTOT)

Dans un premier temps, nous présentons la comparaison du nombre total d'objets trouvés par groupe en fonction de la série d'image présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30), indépendamment des modalités d'affichage **S** et **PA**. Les données ne suivent pas une distribution normale. Le *Tableau 3* présente les statistiques descriptives issues des données étudiées.

*Tableau 3 : Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère NTOT.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NTOT   Série1	18	39,000	58,000	51,722	4,535
NTOT   Série2	18	36,000	55,000	47,056	5,755

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) pour comparaison d'échantillons dépendants. Nous obtenons une forte différence significative entre la Série1 et la Série2 pour la variable NTOT ( $p\text{-value} < 0,0001$  pour le test du signe et  $p\text{-value} = 0,0002$  pour le test de Wilcoxon signé).

Nous avons donc approfondi l'analyse, la différence significative entre la Série1 et la Série2 est observée seulement en modalité PA ( $p\text{-value} = 0,006$  pour le test de Mann-Whitney).

Dans un second temps, nous présentons la comparaison du nombre total d'objets trouvés en fonction des modalités **PA** et **S**.

Les données ne suivent pas une distribution normale. Le *Tableau 4* présente la description des données étudiées.

*Tableau 4 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère NTOT.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NTOT   PA	18	36,000	57,000	48,444	6,195
NTOT   S	18	39,000	58,000	50,333	4,994

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons

dépendants. Nous n’obtenons pas de différence significative ( $p$ -value = 0,072 pour le test du signe et  $p$ -value = 0,113 pour le test de Wilcoxon signé) entre la modalité PA et la modalité S pour la variable NTOT.

## H1.2 : Temps pour trouver le 1<sup>er</sup> objet (T1)

Dans un premier temps, nous présentons la comparaison du temps pour trouver le 1<sup>er</sup> objet en fonction de la série d’image présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30) indépendamment des modalités d’affichage **S** et **PA**. Nous obtenons une distribution normale des données. Le *Tableau 5* présente la description des données étudiées.

*Tableau 5 : Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère T1.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>T1   Série1</b>	18	28,933	33,786	31,196	1,205
<b>T1   Série2</b>	18	26,267	32,933	30,304	1,965

Dans les résultats présentés, la valeur du temps correspond à la valeur du compte à rebours au moment du clic et non au temps mis par les utilisateurs à cliquer sur le premier objet trouvé. Si nous prenons pour exemple la valeur de la moyenne pour la Série1 indiquée dans le *Tableau 5*, soit 31,135, il faut comprendre que les utilisateurs ont mis en moyenne 40 - 31,135 secondes à trouver le premier objet sur chaque image dans la Série1, soit 8,865 secondes en moyenne.

Nous sommes dans le cas d’échantillons appariés. Le test de Fisher de comparaison des variances indique une différence significative entre les deux échantillons ( $p$ -value = 0,026). Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison des lois de distribution. Nous obtenons une différence significative ( $p$ -value = 0,240 pour le test du signe et  $p$ -value = 0,037 pour le test de Wilcoxon signé) entre la Série1 et la Série2 pour la variable T1. En approfondissant l’analyse, nous ne trouvons pas de différence significative entre la Série1 et la Série2 en modalité PA, ni entre la Série1 et la Série2 en modalité S.

Dans un second temps, nous présentons la comparaison du temps pour trouver le 1<sup>er</sup> objet en fonction des modalités **PA** et **S**.

Nous obtenons une distribution normale des données. Le *Tableau 6* présente la description des données étudiées.

*Tableau 6 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère T1.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>T1   PA</b>	18	26,267	33,786	30,570	1,840
<b>T1   S</b>	18	26,667	32,933	30,930	1,511

Nous sommes dans le cas d’échantillons appariés. Le test de Fisher de comparaison des variances n’indique pas de différence significative entre les deux échantillons. Nous effectuons donc deux tests paramétriques (test t et test z) de comparaison des moyennes. Nous

n’obtenons pas de différence significative (p-value = 0,239 pour le test t et p-value = 0,234 pour le test z) en moyenne entre la modalité PA et la modalité S pour la variable T1.

### H1.3 : Le nombre moyen d’objets trouvés par image (NbMO)

Dans un premier temps, nous présentons la comparaison du nombre moyen d’objets trouvés par image en fonction de la série d’image présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30) indépendamment des modalités d’affichage **S** et **PA**. Les données ne suivent pas une distribution normale. Le *Tableau 7* présente la description des données étudiées.

*Tableau 7 : Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère NbMO.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>NbMO   Série1</b>	18	2,600	3,867	3,448	0,302
<b>NbMO   Série2</b>	18	1,933	3,667	3,063	0,475

Nous sommes dans le cas d’échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous obtenons une forte différence significative (p-value < 0,0001 pour le test du signe et p-value = 0,0002 pour le test de Wilcoxon signé) entre la Série1 et la Série2 pour la variable NbMO. Nous avons donc approfondi l’analyse, la différence significative entre la Série1 et la Série2 est observée seulement en modalité PA (p-value = 0,004 pour le test de Mann-Whitney).

Dans un second temps, nous présentons la comparaison du nombre moyen d’objets trouvés par image en fonction des modalités **PA** et **S**.

Les données ne suivent pas une distribution normale. Le *Tableau 8* présente la description des données étudiées.

*Tableau 8 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère NbMO.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>NbMO   PA</b>	18	1,933	3,800	3,156	0,513
<b>NbMO   S</b>	18	2,600	3,867	3,356	0,333

Nous sommes dans le cas d’échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous n’obtenons pas de différence significative (p-value = 0,072 pour le test du signe et p-value = 0,081 pour le test de Wilcoxon signé) entre la modalité PA et la modalité S pour la variable NbMO. En approfondissant l’analyse, nous obtenons une différence significative entre la modalité PA et la modalité S seulement pour la Série2 (p-value = 0,034 pour le test de Mann-Whitney).

## H1.4 : Le nombre de mauvais clics (NbMC)

Dans un premier temps, nous présentons la comparaison du nombre de mauvais clics en fonction de la série présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30) indépendamment des modalités d'affichage **S** et **PA**. Les données ne suivent pas une distribution normale. Le *Tableau 9* présente la description des données étudiées.

*Tableau 9 : Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère NbMC.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbMC   Série1	18	2,000	40,000	14,444	9,581
NbMC   Série2	18	1,000	29,000	8,944	6,593

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous obtenons une différence significative (p-value = 0,004 pour le test du signe et p-value = 0,005 pour le test de Wilcoxon signé) entre la Série1 et la Série2 pour la variable NbMC. Nous avons donc approfondi l'analyse, la différence significative entre la Série1 et la Série2 est observée seulement en modalité **S** (p-value = 0,002 pour le test de Mann-Whitney).

Dans un second temps, nous présentons la comparaison du nombre de mauvais clics sur la série en fonction des modalités **PA** et **S**.

Les données ne suivent pas une distribution normale. Le *Tableau 10* présente la description des données étudiées.

*Tableau 10 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère NbMC.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbMC   PA	18	2,000	29,000	10,889	6,525
NbMC   S	18	1,000	40,000	12,500	10,360

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous n'obtenons pas de différence significative (p-value = 0,240 pour le test du signe et p-value = 0,316 pour le test de Wilcoxon signé) entre la modalité **PA** et la modalité **S** pour la variable NbMC. En approfondissant l'analyse, nous n'obtenons pas de différence significative entre la modalité PA et la modalité S pour la Série1 (p-value = 0,051 pour le test de Mann-Whitney) et pour la Série2 (p-value = 0,071 pour le test de Mann-Whitney).

## H1.5 : Le nombre de « tout juste » réalisés (NbTJ)

Dans un premier temps, nous présentons la comparaison du nombre de tout juste en fonction de la série présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30) indépendamment des modalités d'affichage **S** et **PA**. Les données ne suivent pas une distribution normale. Le *Tableau 11* présente la description des données étudiées.

Tableau 11: Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère NbTJ.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbTJ   Série1	18	1,000	6,000	3,222	1,166
NbTJ   Série2	18	0,000	6,000	2,500	1,823

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous obtenons une différence significative (p-value = 0,038 pour le test du signe) entre la Série1 et la Série2 pour la variable NbTJ.

Dans un second temps, nous présentons la comparaison du nombre de tout juste en fonction des modalités **PA** et **S**.

Les données ne suivent pas une distribution normale. Le *Tableau 12* présente la description des données étudiées.

Tableau 12 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère NbTJ.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbTJ   PA	18	0,000	4,000	2,278	1,447
NbTJ   S	18	0,000	6,000	3,444	1,464

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous obtenons une différence significative (p-value = 0,038 pour le test du signe et p-value = 0,003 pour le test de Wilcoxon signé) entre la modalité **PA** et la modalité **S** pour la variable NbTJ.

## H1.6 : Le nombre de « zéro trouvé » réalisés (NbZT)

Dans un premier temps, nous présentons la comparaison du nombre de zéro trouvé en fonction de la série présentée (Série 1 : image 1 à 15 et Série 2 : image 16 à 30) indépendamment des modalités d'affichage **S** et **PA**. Les données ne suivent pas une distribution normale. Le *Tableau 13* présente la description des données étudiées.

Tableau 13 : Statistiques descriptives des données étudiées pour comparer Série1 vs Série2 vis-à-vis du critère NbZT.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbZT   Série1	18	0,000	1,000	0,167	0,383
NbZT   Série2	18	0,000	3,000	0,333	0,767

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous n'obtenons pas de différence significative (p-value = 0,5 pour le test du

signe et p-value = 0,294 pour le test de Wilcoxon signé) entre la Série1 et la Série2 pour la variable NbZT.

Dans un second temps, nous présentons la comparaison du nombre de zéro trouvé en fonction des modalités **PA** et **S**.

Les données ne suivent pas une distribution normale. Le *Tableau 14* présente la description des données étudiées.

*Tableau 14 : Statistiques descriptives des données étudiées pour comparer PA vs S vis-à-vis du critère NbZT.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbZT   PA	18	0,000	3,000	0,222	0,732
NbZT   S	18	0,000	1,000	0,278	0,461

Nous sommes dans le cas d'échantillons appariés. Nous effectuons donc deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons dépendants. Nous n'obtenons pas de différence significative (p-value = 0,188 pour le test du signe et p-value = 0,286 pour le test de Wilcoxon signé) entre la modalité **PA** et la modalité **S** pour la variable NbTJ.

### Synthèse : modalité PA vs modalité S

Dans le *Tableau 15*, nous présentons l'ensemble des résultats obtenus concernant la comparaison des variables selon les différents critères. Les résultats indiqués représentent la valeur de la p-value. Les résultats en **bleu** indiquent une différence significative entre les deux variables.

*Tableau 15 : Synthèse des résultats de comparaison entre les séries d'images (Série1 vs Série2) et les modalités (PA vs S).*

	Critères de comparaison					
	NTOT	T1	NbMO	NbMC	NbTJ	NbZT
<b>Série1 vs Série2</b>	0,006 S1PA > S2PA	0,037 S1 > S2	0,004 S1PA > S2PA	0,002 S1S > S2S	0,038 S1 > S2	0,294
<b>PA vs S</b>	0,072	0,234	0,034 SS2 > PAS2	0,240	0,003 S > PA	0,188

Pour rappel :

NTOT : Nombre total d'objets trouvés  
T1 : Temps pour trouver le 1<sup>er</sup> objet  
NbMO : Nombre moyen d'objets trouvés par image  
NbMC : Nombre de mauvais clics  
NbTJ : Nombre de tout juste  
NbZT : Nombre de zéro trouvé

## Ordre PA-S versus Ordre S-PA

La même étude statistique que pour la comparaison des séries (Série1 vs Série2) et modalités (PA vs S) a été effectuée pour comparer l'ordre de passation. Nous présentons donc seulement la synthèse des résultats obtenus.

### Synthèse : ordre PA-S vs ordre S-PA

Dans le *Tableau 16*, nous présentons l'ensemble des résultats obtenus concernant la comparaison des variables selon les différents critères. Les résultats indiqués représentent la valeur de la p-value. Les résultats en **bleu** indiquent une différence significative (moyenne ou médiane) entre les deux variables pour la première ligne. Pour la deuxième ligne, les résultats en **bleu** indiquent une différence significative des variances des deux variables.

*Tableau 16 : Synthèse des résultats de comparaison entre les ordres de passation de l'expérience (PA-S vs S-PA).*

	Critères de comparaison					
	NTOT	T1	NbMO	NbMC	NbTJ	NbZT
PA-S vs S-PA	0,200	0,434	0,144	0,035 S-PA > PA-S	0,321	0,418
	0,034 S-PA > PA-S	0,290	0,007 S-PA > PA-S	0,002 S-PA > PA-S	0,311	0,041 S-PA > PA-S

Nous présentons ensuite la comparaison des variables d'ordre de passation sous forme de *box plots* permettant de mettre graphiquement en évidence les différences notoires entre les deux variables pour chaque critère de comparaison.

Commençons avec le critère NTOT :

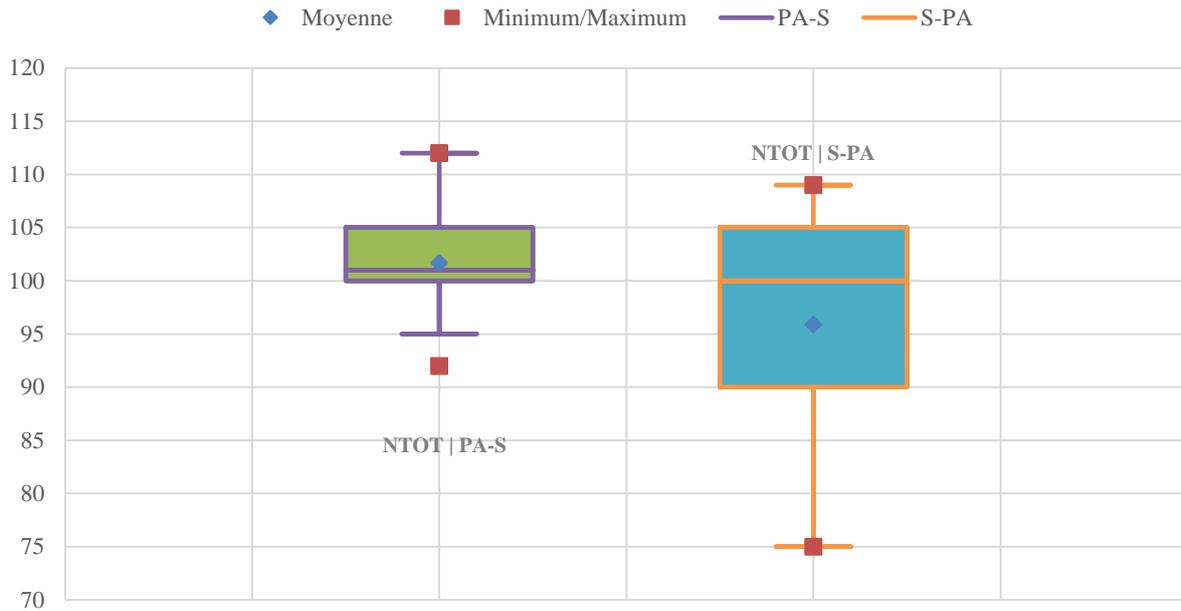


Figure 32 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère NTOT.

La Figure 32 montre que les groupes ayant passé l'expérience dans l'ordre des modalités PA puis S, ont une moyenne du nombre total d'objets trouvés sur l'ensemble de l'expérience supérieure à celle des groupes ayant réalisé l'expérience dans l'ordre des modalités S puis PA. Le maximum et le minimum totaux d'objet trouvés sont également supérieurs. Enfin, comme nous l'avons présenté dans le Tableau 16, nous obtenons une différence significative pour le test de Fisher de comparaison des variances entre les deux variables. En effet, nous retrouvons cette différence sur le *box plots*, les groupes ayant passé l'expérience dans l'ordre PA-S sont beaucoup moins dispersés par rapport au nombre total d'objets trouvés que les groupes ayant réalisés l'expérience en S-PA.

Poursuivons avec le critère T1 :

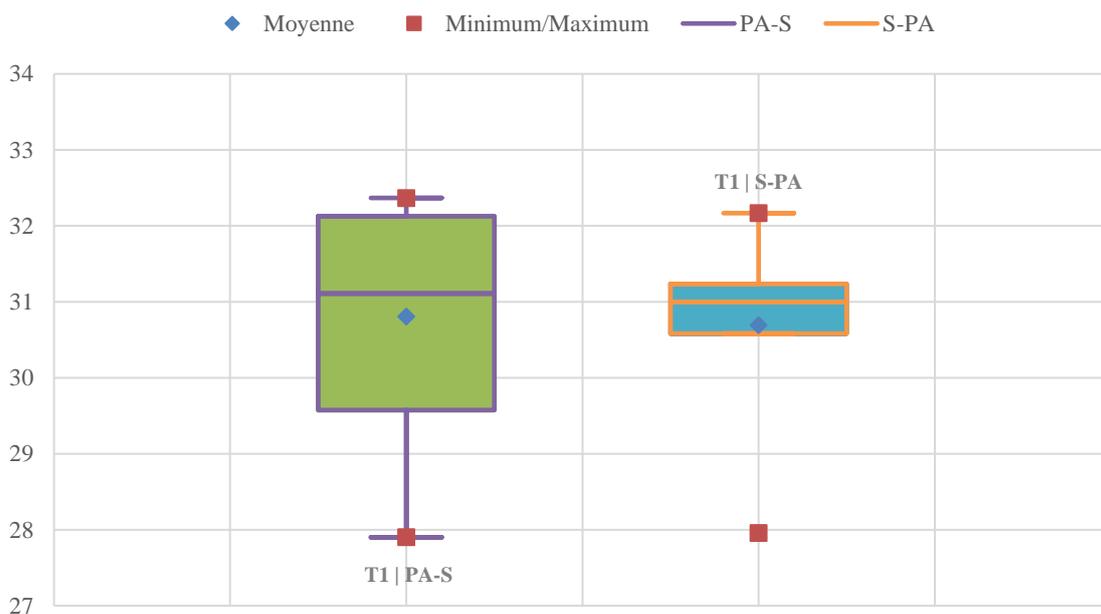


Figure 33 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère T1.

Pour rappel, un compte à rebours de 40 secondes est présent pour chaque image, donc si un groupe a pour résultat 32 secondes, cela signifie que ses membres ont trouvé le 1<sup>er</sup> objet en 8 secondes en moyenne sur l'ensemble de l'expérience.

La *Figure 33*, montre que les groupes ayant passé l'expérience dans l'ordre des modalités PA puis S, ont trouvé le 1<sup>er</sup> objet légèrement plus vite en moyenne que les groupes ayant réalisé l'expérience dans l'ordre des modalités S puis PA. Nous pouvons observer que contrairement au critère NTOT, les groupes ayant passé l'expérience dans l'ordre S-PA sont moins dispersés que les groupes ayant réalisé l'expérience en PA-S, mais pas de manière significativement différente au seuil  $\alpha = 0,05$ .

Continuons avec le critère NbMO :

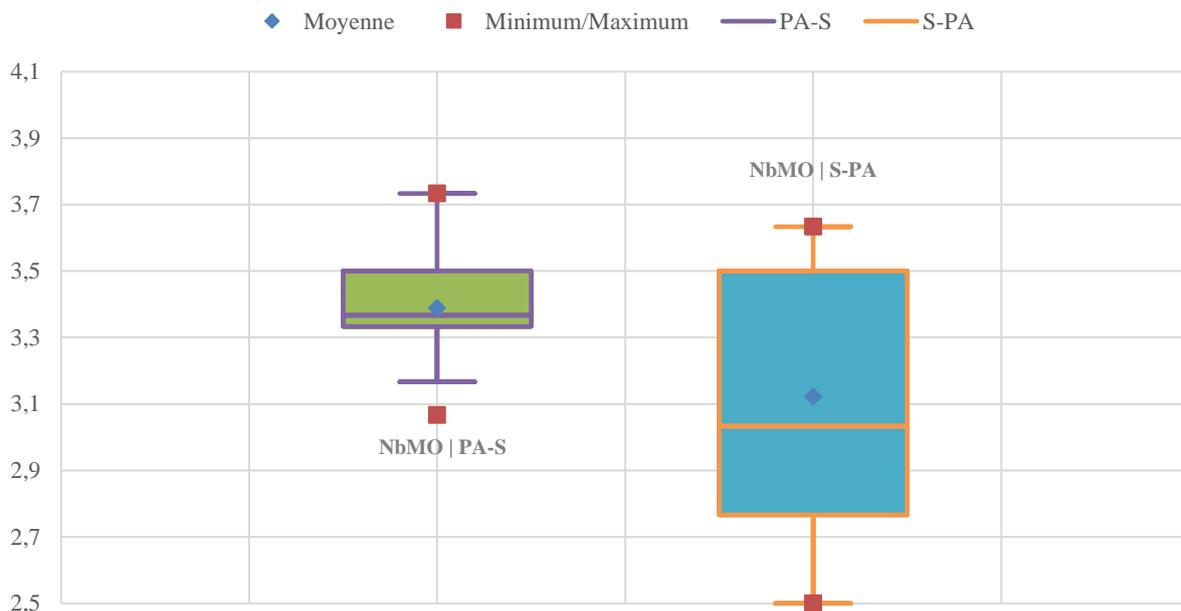


Figure 34 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère NbMO.

Sur la *Figure 34*, nous pouvons constater que les groupes ayant passé l'expérience dans l'ordre des modalités PA puis S, ont une moyenne du nombre moyen d'objets trouvés sur l'ensemble de l'expérience supérieure à celle des groupes ayant réalisé l'expérience dans l'ordre des modalités S puis PA. Le maximum et le minimum d'objet moyen trouvés sont également supérieurs. Enfin, comme pour le critère NTOT, dans le *Tableau 16*, nous obtenons une différence significative pour le test de Fisher de comparaison des variances entre les deux variables. En effet, nous retrouvons cette différence sur le box plots, les groupes ayant passé l'expérience dans l'ordre PA-S sont beaucoup moins dispersés par rapport au nombre moyen d'objets trouvés que les groupes ayant réalisés l'expérience en S-PA.

Voyons maintenant pour le critère NbMC :

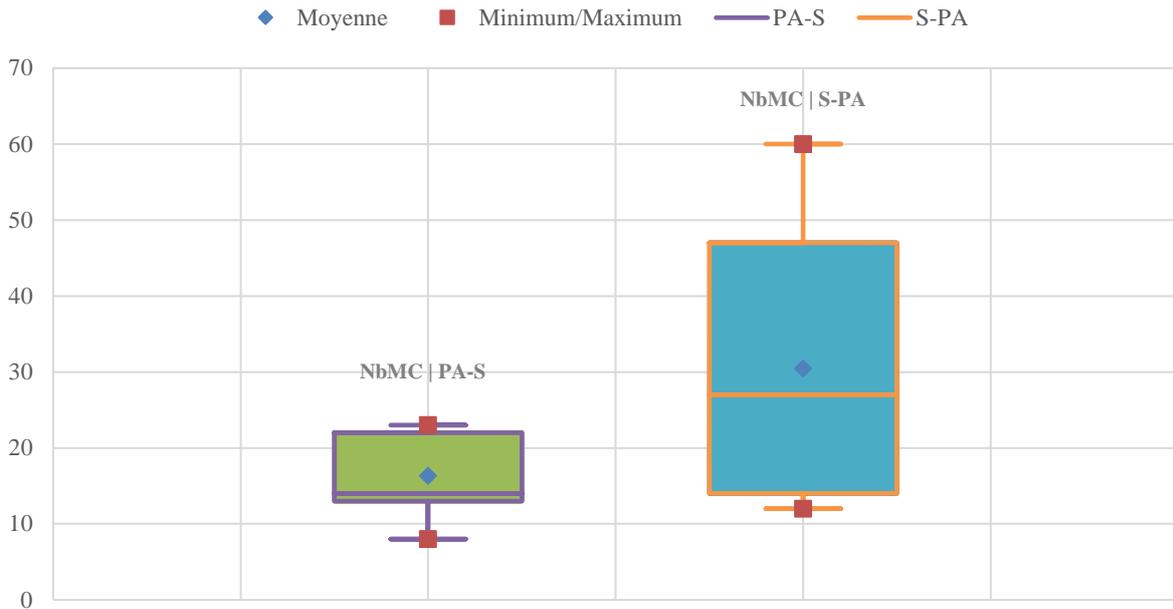


Figure 35 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère NbMC.

De la même façon, la Figure 35 montre qu'en moyenne les groupes ayant réalisé l'expérience dans l'ordre des modalités PA puis S, ont une moyenne de mauvais clics significativement plus faible que les groupes ayant réalisé l'expérience dans l'ordre S-PA. Ce constat est corrélé au résultat exposé dans le Tableau 16. Le maximum et le minimum de mauvais clics sont également inférieurs pour les groupes PA-S. Enfin, comme pour les critères NTOT et NbMO, dans le Tableau 16 nous obtenons une différence significative pour le test de Fisher de comparaison des variances entre les deux variables. En effet, nous retrouvons cette différence sur le box plots, les groupes ayant passé l'expérience dans l'ordre PA-S sont beaucoup moins dispersés par rapport au nombre de mauvais clics réalisés pendant l'expérience que les groupes ayant réalisés l'expérience en S-PA.

Ensuite, exposons les résultats pour le critère NbTJ :

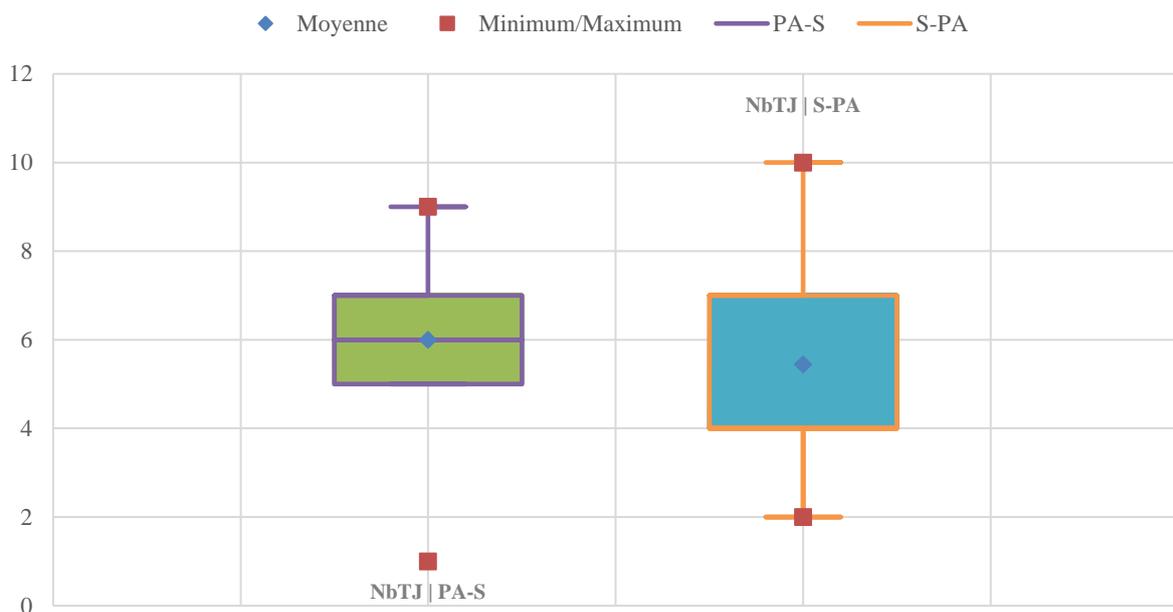
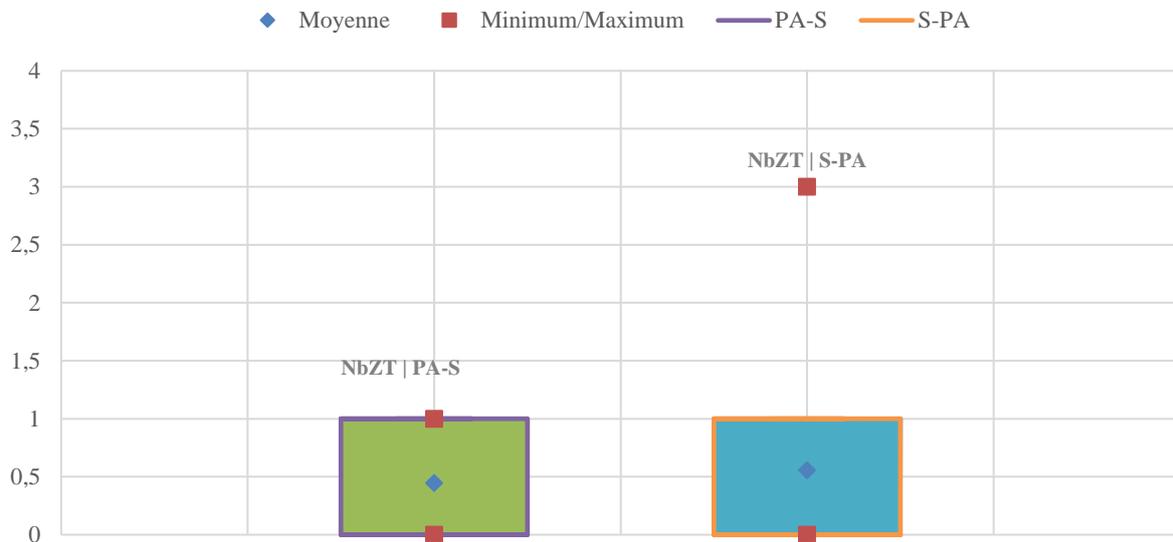


Figure 36 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère NbTJ.

La *Figure 36* expose qu'en moyenne les groupes en PA-S ont trouvé l'ensemble des objets d'une image plus de fois que les groupes en S-PA. Le score le plus élevé de « tout juste » a été réalisé par les groupes en S-PA. Comme pour le critère T1, nous n'observons pas de différence significative en moyenne ou en variance. Cependant, le *box plot* met en évidence que les groupes en PA-S sont moins dispersés que les groupes en S-PA.

Enfin, regardons le critère NbZT :



*Figure 37 : Box plots – Comparaison ordre de passation PA-S versus S-PA pour le critère NbZT.*

Sur la *Figure 37*, nous observons qu'en moyenne, les groupes en S-PA n'ont trouvé aucun objet dans une image plus de fois que les groupes en PA-S. La différence significative de variance observée dans le *Tableau 16* est fortement liée à un groupe ayant réalisé 3 fois la contre-performance de ne trouver aucun objet dans une image.

Dans le but de comparer la performance de manière globale (ensemble des critères de comparaison réunis) pour les ordres de passation, nous avons effectué une normalisation des données des différents critères de comparaison (NTOT, T1, NbMO, NbMC, NbTJ et NbZT). La normalisation des données a été réalisée par l'écart-type non biaisé dont la formule est la suivante :

$$y = (x - \text{Moyenne}) / \text{Ecart-type} ;$$

où :

- $y$  : est la valeur obtenue après normalisation ;
- $x$  : la valeur à normaliser ;
- Moyenne : la moyenne de l'ensemble des données subissant la normalisation ;
- Ecart-type : l'écart-type non biaisé (n-1) de l'ensemble des données subissant la normalisation.

En normalisant ainsi l'ensemble des données de chaque critère, nous pouvons comparer la performance globale des groupes ayant réalisé l'expérience selon les deux ordres de passation. Les données ne suivent pas une distribution normale. Après application d'un test de Mann-Whitney, nous obtenons une différence significative,  $p\text{-value} = 0,021$ .

## Discussion

Rappel de nos deux hypothèses :

H1 : « En modalité **PA**, la performance du groupe est améliorée en comparaison avec la modalité **S** ».

H2 : « Nous pensons que les groupes commençant l'expérience avec la modalité **PA** (ordre PA-S) auront une performance plus élevée en comparaison avec les groupes démarrant avec la modalité **S** (ordre S-PA) ».

### L'effet des Séries d'images : Série1 versus Série2

Dans un premier temps, il nous a semblé important d'étudier d'éventuelles différences de performance pour chaque critère de comparaison entre les séries d'images (Série1 vs Série2). Le but est de contrôler un effet potentiel des séries sur les résultats obtenus, et donc sur notre interprétation des données. En effet, la difficulté des séries d'images n'ayant pas été étudiée en amont de l'expérience, il se peut qu'une des deux séries soit plus « facile » ou « difficile » que l'autre. En d'autres termes, la performance des groupes pourrait être liée à la série. Afin d'éviter d'introduire un tel biais, nous avons contrebalancé les séries et les modalités comme indiqué dans la partie *L'expérience*.

Comme le montre le *Tableau 15*, nous observons des différences significatives pour les critères *temps moyen pour trouver le premier objet* (T1) et *nombre de tout juste* (NbTJ) en faveur de la Série1, c'est-à-dire qu'en moyenne les groupes obtiennent une performance plus élevée en Série1 qu'en Série2. Nous pouvons donc dire que les groupes trouvent les objets significativement plus rapidement en Série 1. De plus, ils ont trouvé tous les objets d'une image (réalisé « un tout juste ») de manière significativement plus élevée en Série1.

Par la suite des différences significatives sont établies en faveur de la Série1, pour les critères *nombre total d'objets trouvés* (NTOT) et *nombre moyen d'objets trouvés par image* (NbMO). Mais ces différences apparaissent seulement en modalité PA. Ces résultats nous laissent penser que pour les critères NTOT et NbMO, la modalité PA a soit un effet facilitateur sur la première série d'image, soit un effet perturbateur sur la deuxième série. Malheureusement, nous ne pouvons pas conclure sur l'effet à ce niveau de l'analyse. Ceci laisse néanmoins présager d'un effet d'ordre.

Pour le critère *nombre de mauvais clics*, en modalité S, nous notons à l'inverse une différence significative entre les séries en faveur de la Série2. En effet, pour cette modalité, les groupes ont réalisé significativement plus de mauvais clics en Série1 qu'en Série2. Ce résultat s'oppose à l'ensemble des résultats apportés jusqu'à présent car il est en défaveur de la Série1. Mais il est en faveur d'un effet bénéfique de la modalité PA sur le nombre de mauvais clics. En effet, nous n'observons pas de différence entre les séries en modalité PA. De plus, les groupes ayant participé à l'expérience dans l'ordre S-PA réalisent plus de mauvais clics en modalité S que les groupes ayant participé à l'expérience en PA-S.

Enfin, aucune différence significative n'est mise en évidence pour le critère *nombre de zéro trouvé* entre les séries.

L'ensemble des résultats présentés nous laissent facilement conclure à un avantage net procuré par la Série1 sur la performance des groupes. Nous pouvons donc dire que la Série1 est plus facile que la Série2. Cet effet mis en évidence, nous pouvons maintenant interpréter les résultats concernant les modalités.

## L'effet des modalités

Suite à l'analyse des résultats des séries, nous pouvons analyser les résultats concernant notre première hypothèse. Seul le critère NbTJ détermine une différence significative en termes de performance entre les modalités, et cette différence est en faveur de la modalité S. Effectivement, les groupes font plus de *tout juste* en modalité S qu'en modalité PA. Ainsi, nous obtenons un résultat contradictoire pour notre hypothèse H1.5.

De plus, nous mettons en évidence une différence significative entre les modalités PA et S par rapport au nombre moyen d'objets trouvés en faveur de la modalité S, mais seulement en Série2. Ce résultat vient contredire notre hypothèse H1.3 mais il va également dans le sens d'un effet d'ordre. Nous ne pouvons donc pas conclure concernant notre première hypothèse sur la performance supérieure des groupes en modalité PA.

## L'effet d'ordre de la passation

Nous allons maintenant analyser les résultats concernant notre deuxième hypothèse sur l'effet de l'ordre de passation sur la performance.

Le premier résultat concluant est celui concernant le critère nombre de mauvais clics (NbMC). Dans le *Tableau 16*, nous exposons une différence significative entre les groupes ayant réalisé l'expérience dans l'ordre S-PA et ceux en PA-S en faveur de l'ordre PA-S. En effet, les groupes en PA-S effectuent significativement moins de mauvais clics que les groupes en S-PA. De plus, nous remarquons une distribution significativement plus regroupée des groupes en PA-S concernant ce même critère. Ces résultats confirment notre interprétation, concernant un effet bénéfique de la modalité PA sur le nombre de mauvais clics. Pour les autres critères nous n'avons pas pu mettre en évidence de différence significative entre les ordres de passation. Cependant, nous constatons que pour les critères NTOT, T1, NbMO et NbTJ, les groupes en PA-S ont une moyenne plus élevée que les groupes en S-PA et, à l'inverse, ont une moyenne moins élevée concernant les critères NbMC et NbZT.

De plus, nous constatons également que pour les critères NTOT, NbMO et NbZT, les groupes en PA-S ont une distribution significativement plus regroupée. Ainsi, en regroupant l'analyse des différents critères en une performance globale à l'aide de la normalisation, nous obtenons bien une différence significative de performance entre les groupes ayant réalisé l'expérience en PA-S en comparaison à ceux en S-PA. Nous pouvons donc conclure que réaliser l'expérience dans l'ordre PA-S permet d'obtenir de meilleures performances. Il semblerait qu'être exposé dans un premier temps à la modalité PA permettrait au groupe de se construire une stratégie collective de recherche, consciemment ou non, qui serait exploitée dans la deuxième partie de l'expérience et qui expliquerait donc leurs meilleures performances. Cependant, nous pouvons également interpréter qu'être exposé à la modalité PA dans un second temps sur une série a priori plus difficile dégraderait la performance collective ou, pour le moins, ne permettrait pas d'améliorer la performance. Ceci peut aussi vouloir dire qu'être exposé à la modalité S dans un premier temps génère un effet de fatigue plus important détériorant fortement la performance dans la deuxième partie en modalité PA. De plus, cet effet de fatigue associé à l'effet de nouveauté de la modalité PA accentuerait les écarts de performance entre les ordres de passation. Pour conclure sur ces différents effets nous avons envisagé à l'issue de cette de construire première expérience d'en construire une nouvelle, décrite dans la suite du document.

## Conclusions<sup>13</sup>

Cette première expérience ne permet pas de conclure concernant les hypothèses d'amélioration de la performance de groupe grâce à l'affichage du point d'attention. Néanmoins, elle permet de confirmer notre idée que la connaissance de l'endroit regardé par ses partenaires a un effet bénéfique sur la collaboration implicite et la performance du groupe. Principalement dans le cas où, les groupes ont cette connaissance de l'endroit regardé dans un premier temps, puis ne l'ont plus après (ordre PA-S). Notre hypothèse sur l'établissement de stratégies de recherches collaboratives involontaires et automatiques suite à la possibilité d'avoir la connaissance de l'endroit regardé par ses partenaires semble donc se vérifier.

La principale conclusion de cette expérience est notre envie d'approfondir le concept de vision flottante et son effet sur la collaboration implicite dans un groupe. Cependant, cette première expérience devait servir uniquement de préparation de la suite des travaux de cette thèse centrés sur le développement et l'évaluation de nouveaux paradigmes d'interface pour le partage collaboratif d'affichage en groupe ouvert. Nous sommes ravis d'avoir mis en lumière un concept intéressant mais nous n'avons pas poursuivi les expériences sur la vision flottante.

Cette première expérience nous a permis de mettre en œuvre un *head tracking* multi-utilisateur appareillé et de tester l'effet de l'affichage du point d'attention à l'écran. L'idée développée dans le cadre de cette expérience est d'utiliser un dispositif de *tracking* instrumenté ("*motion tracking*" par infrarouge) pour émuler le dispositif de *head tracking* "idéal" dans un premier temps et de déterminer les éventuels freins à la mise en œuvre d'un dispositif de *head tracking* multi-utilisateur mais sans calibration ni appareillage. Nous craignons que les utilisateurs ne soient gênés par la matérialisation de leur zone d'attention à l'écran. Les résultats montrent le contraire. De plus, le recueil informel du ressenti immédiat des participants nous a confirmé que les utilisateurs n'étaient pas gênés par la présence de la matérialisation de leur zone d'attention.

De plus, nous avons pu constater lors du débriefing de fin d'expérience que, du point de vue des utilisateurs, l'affichage du point d'attention n'est pas perturbant pour la plupart d'entre eux et qu'ils vont jusqu'à dire qu'ils n'y ont pas porté attention.

Sur la base de ces constats, nous nous sommes lancés dans le développement d'un dispositif à base de Kinect V2 autorisant une interaction multi-utilisateur et non appareillée à base de *head tracking* avec un écran de grand format. Ce dispositif développé, nous avons pu évaluer l'acceptabilité, l'utilité et l'utilisabilité du système à l'aide du ressenti des utilisateurs face aux paradigmes d'interfaces déployés. Nous décrivons la construction et la mise en œuvre de ce dispositif dans les chapitres suivants.

---

<sup>13</sup> Cette étude a fait l'objet d'une présentation orale dans la conférence EPIQUE : Marion, D., P.-A. Favier, C. Bey, S. Lini and J.-M. Andre (2015). Etude outillée de l'attention visuelle flottante dans la recherche collaborative d'informations sur grands écrans. EPIQUE 2015, Aix-en-Provence, France, Arpege Science Publishing.

## Résumé

Dans ce chapitre, nous présentons un système basé sur la détection et le suivi de la position et de l'orientation de la tête qui permet l'affichage sur un écran de grand format du point d'attention de chaque individu. Ainsi, chaque individu sait où regarde à l'écran chacun de ses partenaires. Nous posons l'hypothèse que ce système améliore la recherche collective d'informations en introduisant une collaboration implicite entre les individus. Les résultats de cette expérience préliminaire soulignent la construction, consciente ou non, d'une stratégie performante grâce à la matérialisation du point d'attention.



# Chapitre 4 : Vers une interaction à multiples utilisateurs non appareillés et sans calibrage avec les écrans de grand format.

Ce chapitre est consacré à la description de notre deuxième expérience. Celle-ci a pour objectif la caractérisation des mouvements de la tête lors d'une tâche de consultation d'un affichage de grand format. Elle est conduite à l'aide d'un dispositif de *head tracking* basé sur le Kinect V2. Nous présenterons dans un premier temps la problématique et les hypothèses qui en découlent. Puis dans un second temps, nous décrirons le protocole expérimental (méthodes et matériels) mis en place afin d'adresser ces premières hypothèses. Enfin, nous présenterons et discuterons des résultats obtenus qui ont guidé notre démarche vers la troisième expérience.

## Introduction

### Contexte

La démocratisation récente d'écrans de très grands formats et de la « haute définition » change le rapport des utilisateurs à l'écran. Ces derniers sont en attente de nouvelles technologies qui leur offrent toujours plus de fonctionnalités intuitives et naturelles. De ce fait, de nombreux travaux émergent sur les « interfaces intelligentes » qui s'adaptent, se transforment, en fonction du nombre d'individus les utilisant ou encore en fonction de l'intérêt que peut porter un utilisateur à l'interface en question (Schiavo, Mencarini et al. 2013). Une des principales difficultés du domaine est qu'il est difficile de trouver la technique et le capteur adaptés à ces besoins. Cette dualité, capteur adéquat et technique adaptée engendrent une grande complexité dans la mise œuvre d'une interaction basée sur la détection de l'orientation de la tête. Dans la partie précédente (*cf. Chapitre 3*), nous utilisons la technologie Optitrack afin de mesurer avec précision les mouvements de la tête. Grâce à ces mesures nous extrapolions une zone de focalisation sur l'écran correspondant à la projection du champ de vision de l'utilisateur en fonction de l'orientation de sa tête et non de son regard. Les utilisateurs devaient porter un casque muni de marqueurs réfléchissant la lumière

infrarouge émise par la caméra. Néanmoins, l'avènement récent de capteurs de profondeur (Shotton, Sharp et al. 2013) à bas prix, comme le Kinect V2, permet l'accès, de manière relativement simple et précise (Lachat, Macher et al. 2015), à l'exploration des interactions possibles via la détection de l'orientation de la tête. Mais ceux-ci limitent les tests d'interaction à des contextes relativement bien contrôlés (luminosité, zone d'interaction, nombre d'utilisateurs, etc.).

L'objectif de ce chapitre est de caractériser les mouvements de rotation de la tête lors d'une tâche de consultation d'un affichage de grand format, à l'aide d'un dispositif de *head tracking*. Ainsi, nous pourrions concevoir un système permettant une interaction sans appareillage ni calibrage individuel et multi-utilisateurs lors de la consultation d'un affichage de grand format en milieu ouvert, mais dans des conditions de luminosité d'intérieur. De nombreuses études ont été menées sur les différents types de capteurs et méthodes (Ariz, Bengoechea et al. 2016, Ranjan, Patel et al. 2016, Yan, Ricci et al. 2016, Drouard, Horaud et al. 2017, Eichenlaub 2017, Jeni, Cohn et al. 2017, Khalidov et Odobez 2017) permettant d'estimer l'orientation de la tête des utilisateurs (Murphy-Chutorian et Trivedi 2009, Fanelli, Gall et al. 2012).

Selon nous, de façon idéale, le système proposé devrait permettre :

- une détection précise, proche de  $1^\circ$  ;
- une détection non intrusive/invasive ;
- une détection multi-utilisateur ;
- une détection précise sans calibrage préalable ;
- une réalisation à bas prix.

Dans le contexte d'étude de ce chapitre, les utilisateurs effectuent une tâche de lecture. C'est donc la vision fovéale qui est mise en jeu. La littérature établit cette zone du champ visuel à  $1^\circ$  (Foucambert 2003). Une erreur au-delà du degré nous semble donc compromettante. De plus, l'utilisation de grands écrans implique un rapport de distance élevé entre l'utilisateur et l'écran. Cette distance entre l'utilisateur et l'écran va fatalement réduire les mouvements de la tête au profit des mouvements des yeux. En effet, avec l'éloignement à l'écran, notre zone fovéale de  $1^\circ$  va permettre d'englober plus d'informations et donc de limiter les mouvements de la tête. Ainsi, la distance entre l'utilisateur et l'écran et la sensibilité à la rotation de la tête sont covariantes dans l'estimation de la zone de lecture par l'utilisateur à l'écran. Dans l'expérience décrite ici, nous souhaitons effectuer une caractérisation des mouvements de la tête sans instrumentation de l'utilisateur. Nous testons donc le Kinect V2 et son SDK pour réaliser le *head tracking*.

## Les yeux et la tête : un système masse-ressort

Il existe 3 types de rotation pour la tête comme le montre la *Figure 38* :

- la rotation en lacet (rotation droite-gauche) s'effectuant dans le plan frontal ;
- la rotation en tangage (rotation haut-bas) s'effectuant dans le plan sagittal ;
- la rotation en roulis (la tête se « penche ») s'effectuant dans le plan transverse.

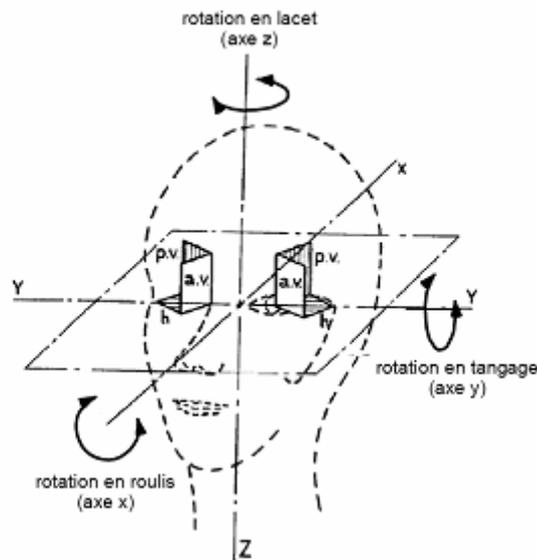


Figure 38 : Les trois types de rotation de la tête, d'après (Benson et Jobson 1973).

Nous posons l'hypothèse que lors de la consultation d'informations sur un écran de grand format, le degré d'orientation de la tête d'un individu varie d'un individu à l'autre pour la consultation d'une même information, à une même distance, depuis une même position, mais oscille dans une plage de valeur moyenne faible. C'est cette plage de valeur pour chaque type de rotation de la tête que nous mettons en évidence ici.

Par exemple, une information affichée à  $15^\circ$  sur la droite et  $10^\circ$  vers le haut par rapport à une position neutre<sup>14</sup> de la tête de l'utilisateur ne nécessitera qu'une rotation moyenne de la tête de  $12^\circ$  vers la droite et de  $5^\circ$  vers le haut afin d'être consultée de manière confortable. Le reste de la rotation étant compensé par les yeux. Nous considérons que l'orientation de la tête d'un individu dont le but est de regarder une information est un système composé des mouvements de la tête et de ceux des yeux. La littérature qualifie ce couple de système masse-ressort (Berthoz 1992, Bingham et Stassen 1994). En effet, la position d'une information dans le champ de vision, va attirer d'abord un mouvement des yeux suivi si nécessaire d'un mouvement de la tête afin d'atteindre l'information désirée dans des conditions de confort optimales, avec une légère inertie. Les conditions optimales de confort sont évidemment propres à chaque individu. D'ailleurs, il existe actuellement une technologie développée par Essilor nommée l'EyeCode (Encaoua et Thomet 2017, Essilor 2017) qui mesure l'anatomie des yeux en 3D dans le but de concevoir des verres ultras personnalisés. En effet, l'œil humain est résolument unique. Certaines caractéristiques comme la position du centre de rotation de l'œil, l'œil directeur et le coefficient tête-œil qui mesure la coordination des mouvements de la tête et des yeux sont révélateur des différences interindividuelles. Un individu qui privilégie le mouvement des yeux sur celui de la tête pourra être appelé « visionaute » tandis qu'à l'inverse un individu favorisant les mouvements de tête pourra être qualifié de « céphalonaute ». Nous souhaitons remercier Alexandre Guédon, opticien qui nous a présenté ce système disponible chez les opticiens engagés partenaires experts Essilor.

Cependant, dans cette partie nous souhaitons qualifier une plage de valeurs moyennes des rotations de la tête correspondant aux conditions de confort optimales de consultation d'informations sur écran de grand format permettant de caractériser au plus près le plus grand

<sup>14</sup> Dans notre cas, la position dite neutre de la tête correspond à une rotation de la tête : dans le plan frontal, appelée rotation en lacet (rotation droite-gauche) de  $0^\circ$  ; dans le plan sagittal, appelée rotation en tangage (rotation haut-bas) de  $0^\circ$  ; dans le plan transverse ou horizontal, appelée rotation en roulis de  $0^\circ$ .

nombre d'individus, qu'ils soient *visionautes* ou *céphalonautes*. La découverte de plages de rotations moyennes va permettre d'être plus précis sur la zone de focalisation calculée sur l'écran en permettant de créer un filtre discriminant les fausses détections. Par la suite, le filtre mis au point va permettre une interaction plus naturelle avec l'écran, via *head tracking*, pour une majorité d'individu.

## Données brutes mesurées par le Kinect V2

Le Kinect V2 et son SDK nous permettent de calculer la zone de focalisation sur l'écran de six individus simultanément en se basant sur la détection de la position et de l'orientation de leurs têtes. Cependant, les calculs basés sur les données brutes issues du Kinect V2 ne permettent pas une interaction précise et naturelle avec un écran de grand format. Par exemple, nous avons pu observer lors de premiers tests que la détection oscille considérablement et que les calculs de la focalisation avec les données brutes ne permettent pas en l'état une interaction correcte, notamment sur la dimension verticale. De ce fait, nous avons tenté d'utiliser un filtre appelé « *one euro filter* » particulièrement adapté au contexte des systèmes interactifs et aux données du suivi du mouvement humain en temps réel (Casiez, Roussel et al. 2012). Ce filtre est un algorithme simple pour filtrer les signaux bruités en conservant une haute réactivité et une importante précision. Il utilise un filtre passe-bas de premier ordre avec une fréquence de coupure adaptée. L'algorithme est simple à mettre en œuvre, utilise très peu de ressources, et est facilement réglable (deux paramètres : la fréquence minimum de coupure et la pente de coupure). Néanmoins, suite à nos premiers tests, nous n'avons pas retenu ce filtre car il n'autorisait pas une interaction fluide et précise basée sur les mouvements de tête. Nous avons alors mis en place le protocole d'expérience décrit dans ce chapitre dans le but de qualifier les données brutes mesurées par le capteur Kinect V2 afin de construire notre propre technique de filtrage.

## Principe de l'expérience

L'expérience consiste en la lecture depuis une distance donnée d'affiches présentées sur l'écran (la *Figure 39* montre une photo de l'installation). Les utilisateurs sont placés dans une des zones numérotées au sol, les épaules parallèles à l'écran et doivent lire les affiches présentées.



Figure 39 : Photo de l'installation de l'expérience concernant la caractérisation des mouvements de la tête lors d'une tâche de consultation d'un affichage de grand format à l'aide d'un dispositif d'head tracking.

Le Kinect V2 placé au-dessus de l'écran enregistre les données de position et d'orientation de la tête de l'individu par rapport à son centre durant chaque session de lecture d'affiche. Ainsi, pour chaque lecture d'affiche depuis une position donnée, nous définissons les cas de consultation pour les rotations en tangage et en lacet :

Pour la rotation en lacet :

- cas 0 : le groupe regarde en face de lui ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 1, 7 et 13 (cf. Figure 42) ;
- cas 1 : le groupe regarde à plus une colonne écran par rapport à sa position ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 2, 8 et 14 ;
- cas 2 : le groupe regarde à plus deux colonnes écran par rapport à sa position ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 3, 9 et 15 ;
- cas 3 : le groupe regarde à plus trois colonnes écran par rapport à sa position ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 4, 10 et 16 ;
- cas 4 : le groupe regarde à plus quatre colonnes écran par rapport à sa position ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 5, 11 et 17 ;
- cas 5 : le groupe regarde à plus cinq colonnes écran par rapport à sa position ;
  - Exemple : les utilisateurs sont en zone au sol numéro 1 et regardent les affiches 6, 12 et 18.

Nous effectuons la même démarche pour la rotation en tangage :

- cas 1 : le groupe regarde la ligne supérieure de l'écran, affiches 1 à 6 ;
- cas 2 : le groupe regarde la ligne centrale de l'écran, affiches 7 à 12 ;
- cas 3 : le groupe le regarde la ligne inférieure de l'écran, affiches 13 à 18.

## Hypothèses

L'objectif de cette expérience est double :

- dans un premier temps, qualifier la précision et les erreurs de détection du Kinect V2. Cet objectif se rapproche de celui de l'étude (Darby, Sánchez et al. 2016) ;
- dans un second temps, vérifier à l'aide des données récoltées la possibilité de caractériser les mouvements de tête des utilisateurs lorsqu'ils consultent des informations sur un écran de grand format.

La principale hypothèse que nous interrogeons dans ce chapitre, dans le but de remplir notre objectif de caractérisation des mouvements de la tête des utilisateurs, est la suivante :

« A une distance donnée, pour une affiche donnée et sur la base des données fournies par le Kinect V2, nous tournons tous la tête quasiment de la même manière, c'est-à-dire que les rotations de la tête nécessaires pour consulter une information ne sont pas significativement différentes d'un individu à l'autre. »

De cette hypothèse générale, nous tirons les hypothèses suivantes :

- les deux premières hypothèses sont basées sur la capacité du Kinect V2 à détecter des différences de rotation significatives pour des cas de consultation différents (précision suffisante du Kinect V2) :
  - (H1) la moyenne en degré de la rotation en lacet mesurée est significativement différente d'un cas de consultation à l'autre ;
  - (H2) la moyenne en degré de la rotation en tangage mesurée est significativement différente d'un cas de consultation à l'autre ;
- Les deux hypothèses suivantes concernent la caractérisation des mouvements de la tête pour des cas similaires de consultation :
  - (H3) la moyenne en degré de la rotation en lacet mesurée n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation ;
  - (H4) la moyenne en degré de la rotation en tangage mesurée n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation ;

# Protocole expérimental

## Matériel

Nous utilisons le Kinect V2 qui est une caméra infrarouge et la version 2.0.1410.19000 du SDK. Celui-ci est placé au-dessus de l'écran en son centre. Le centre du capteur Kinect V2 est à 16 centimètres au-dessus de l'écran et incliné de  $34,8^\circ$  vers le bas par rapport au sol. Cette position et cette inclinaison ont été déterminées afin de maximiser la couverture de la zone d'interaction possible devant l'écran tout en respectant les contraintes physique de notre salle expérimentale.

L'écran servant à l'affichage mesure 3.05 mètres de base horizontale et 1.72 mètre de hauteur. L'image projetée, par projection arrière, en 16/10, mesure 2.96 mètres de base horizontale et 1.66 mètre de hauteur. Elle est surélevée par rapport au sol de 1.27 mètre. Elle est centrée sur l'écran et sa résolution est de 1920x1080 pixels. La zone d'interaction se situe à 1,60 mètre de l'écran. Les dimensions de la zone d'interaction sont 2,81 mètres de longueur et 2,02 mètres de largeur. La zone d'interaction et l'écran sont centrés l'un par rapport à l'autre comme le montre la *Figure 40*.

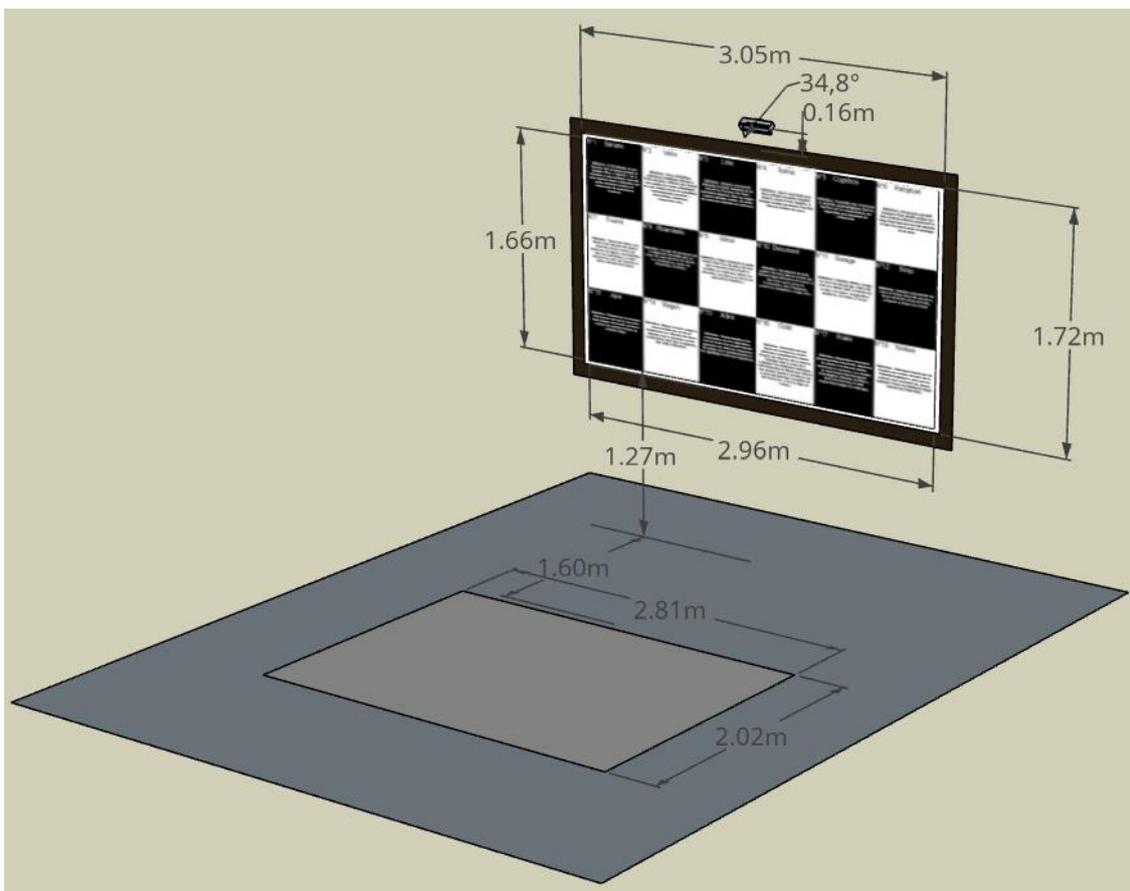


Figure 40 : Modélisation de l'installation avec les côtes en mètre.

## Méthode

*Détails computationnels.* Techniquement, le SDK donne accès aux coordonnées 3D d'un nuage de points caractéristiques sur la personne détectée. Nous utilisons pour la détection de la position dans l'espace de l'utilisateur les coordonnées (x, y et z) issues du nœud « Head » comme le montre la *Figure 41*.

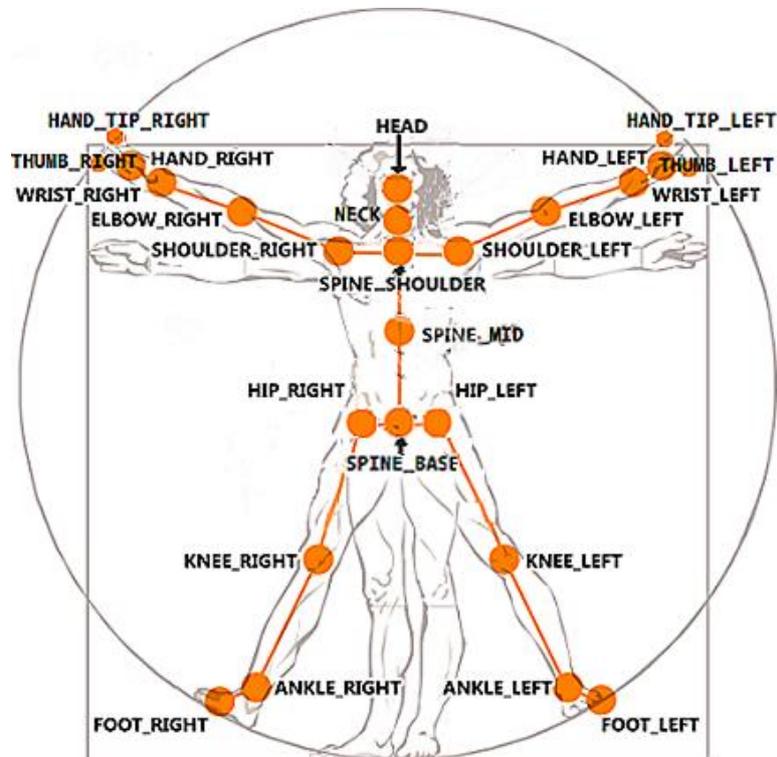


Figure 41 : Body Joints issus du SDK de la Kinect V2.

Pour la détection des angles d'orientation de la tête de l'utilisateur, nous utilisons l'événement de *Face Tracking* et plus particulièrement la propriété *FaceRotationQuaternion* de la class *FaceFrameResult*, permettant d'obtenir les trois rotations en degrés de la tête de l'utilisateur dans le système de coordonnées du Kinect V2. Suite à l'acquisition des coordonnées 3D de l'individu et des angles de rotation, nous calculons ensuite les angles de rotation de la tête par rapport au plan écran et non plus par rapport au centre de la Kinect V2. Enfin, nous calculons grâce à ces données la zone de focalisation sur l'écran de l'individu détecté (cf. *Identification des zones d'attention sur l'écran*).

*Milieu d'implémentation.* La zone d'interaction est divisée en 18 parties identiques, 6 sur la longueur réparties sur 3 lignes comme le montre la *Figure 39*. La taille de la zone d'interaction au sol a été déterminée à l'aide de tests de détection d'*head tracking* pour maximiser la qualité de détection.

La zone d'affichage sur l'écran est divisée de la même manière comme montré par la *Figure 42*.

<p>N°1 Banane</p> <p>Definition : Fruit allongé, un peu incurvé, vert ou jaune, groupé en grappes dites «<i>regimes</i> » sur le bananier, et qui contient, sous une peau épaisse aisée à détacher, une pulpe amygdalée nutritive et savoureuse de consistance fondante.</p>	<p>N°2 Verre</p> <p>Definition : Verre métallique, substance métallique obtenue à l'état amorphe. (Obtenus par trempe très rapide, les verres métalliques se sont révélés riches de retombées technologiques possibles : électronique, matériaux composites, etc.)</p>	<p>N°3 Lilas</p> <p>Definition : Arbuste ornemental (oliacée) de l'Asie tempérée et du sud-est de l'Europe, aux fleurs odorantes blanches, lilas ou rouges, à quatre pétales soudés. (Il existe de nombreux hybrides et cultivars.)</p>	<p>N°4 Tennis</p> <p>Definition : Sport consistant pour deux (simple) ou quatre (double) joueurs, munis d'une raquette, à envoyer la balle au-dessus d'un filet dans les limites du court.</p>	<p>N°5 Cognition</p> <p>Definition : Ensemble des structures et activités psychologiques dont la fonction est la connaissance, par opposition aux domaines de l'affectivité.</p>	<p>N°6 Parapluie</p> <p>Definition : Accessoire portatif composé d'une étoffe tendue sur une armature pliante coulissant le long d'une tige qui sert de manche, et que l'on ouvre pour se protéger de la pluie.</p>
<p>N°7 Fourmi</p> <p>Definition : Fourmi de visite, nom donné à des fourmis des genres <i>anomma</i>, <i>doryle</i>, <i>eciton</i>, qui pénètrent en légions considérables dans les habitations, où elles dévorent les rats, les insectes et autres animaux terrestres.</p>	<p>N°8 Ribambelle</p> <p>Definition : Groupe de personnes qui se déplacent ensemble plus ou moins les unes derrière les autres : Les institutrices suivies de ribambelles d'enfants.</p>	<p>N°9 Miroir</p> <p>Definition : Engin composé de petits miroirs fixés sur une pièce de bois que le chasseur fait pivoter et scintiller au soleil pour attirer les alouettes. (La chasse au miroir est aujourd'hui interdite.)</p>	<p>N°10 Document</p> <p>Definition : Documents de bord, papiers de bord, pièces écrites qui doivent impérativement se trouver à bord d'un navire (acte de francisation, rôle d'équipage, permis de navigation, congé, etc.).</p>	<p>N°11 Garage</p> <p>Definition : Familier. Mettre, ranger sur une voie de garage, mettre de côté une affaire dont on décide de ne plus s'occuper : marginaliser quelqu'un, le mettre à l'écart.</p>	<p>N°12 Sirop</p> <p>Definition : Solution concentrée de sucre et d'eau (sirop de sucre) ou de jus de fruits (sirop de fruit), préparée à chaud ou à froid.</p>
<p>N°13 Jupe</p> <p>Definition : Élément de la structure d'un lanceur qui sert à raccorder deux étages successifs, ou bien à caréner les moteurs/fusées du premier étage.</p>	<p>N°14 Wagon</p> <p>Definition : Wagon couvert, wagon à caisse fermée par un toit et comportant sur chacune des faces latérales une ou plusieurs portes coulissantes ainsi que, en général, des volets d'aération.</p>	<p>N°15 Arbre</p> <p>Definition : Représentation d'un programme, ou d'un algorithme, sous la forme d'un graphe faisant apparaître les diverses séquences de calcul, leurs enchaînements et les sigillages entre ces séquences.</p>	<p>N°16 Octet</p> <p>Definition : Ensemble de huit électrons formant la couche extérieure complète d'un atome. (D'après les théories de la valence chimique dues à Lewis et à Langmuir, les molécules chimiques sont disposées de façon que chacun des atomes tende à constituer une couche périphérique complète de huit électrons : c'est la règle de l'octet.)</p>	<p>N°17 Koala</p> <p>Definition : Mammifère marsupial (phalangéride) d'Australie, dépourvu de queue, à la fourrure dense, arboricole et nocturne. (Les koalas se nourrissent des bourgeons d'eucalyptus. Décimés au XIXe s. pour leur fourrure, ils font aujourd'hui l'objet d'un élevage.)</p>	<p>N°18 Yankee</p> <p>Definition : Soubriquet donné par les Anglais aux colons révoltés de la Nouvelle-Angleterre, puis par les socialistes aux nordistes et, depuis, appliqué à tous les habitants anglo-saxons des Etats-Unis.</p>

Figure 42 : Image affichée à l'écran découpée en 18 affiches numérotées de taille équivalentes présentant un titre et sa définition.

Le découpage en 6 zones sur la longueur (zone d'interaction et d'affichage) a été établi par les contraintes de détection du nombre d'utilisateurs simultanés du Kinect V2, soit 6 utilisateurs maximum.

*Expérience.* Le protocole d'expérimentation consiste pour un individu à se placer au centre d'une zone au sol, indiquée par l'expérimentateur, de façon à faire face à l'écran et à conserver les épaules parallèles à celui-ci. Une fois l'utilisateur correctement positionné dans la zone au sol nous lui demandons de lire des affiches numérotées présentées sur l'écran, en annonçant leur numéro. Lors de la lecture, l'utilisateur ne doit alors plus bouger le corps mais seulement bouger la tête de manière à lire confortablement les affiches annoncées dans un ordre précis et identique pour chaque utilisateur. L'ordre des affiches à lire est le suivant : N°6, 7, 17, 9, 2, 18, 1, 10, 15, 11, 5, 16, 13, 8, 12, 3, 4 et N°14. En donnant les affiches dans l'ordre de lecture occidental, soit de gauche à droite et de haut en bas, nous aurions risqué de suggérer des rotations de la tête très répétitives et systématiques et ainsi biaiser nos données. Les affiches sont donc volontairement désignées dans le désordre afin d'obliger l'utilisateur à rechercher l'affiche énoncée par l'expérimentateur. Nous nous rapprochons ainsi d'une position naturelle et confortable de la tête, lors de la lecture spontanée de l'affiche en question. L'ordre n'a pas été choisi au hasard. Le déplacement de tête nécessaire entre chaque affiche annoncée couvre 18 cas différents sur les 324 cas de déplacements possibles. Chaque utilisateur réalise la lecture de 18 affiches sur l'écran depuis 3 positions au sol indiquées en couleur sur la Figure 43. Ces trois positions correspondent à 3 distances différentes par rapport à l'écran.

Ecran	1	2	3	4	5	6	Légende :	
	7	8	9	10	11	12		Ecran : numéro 1, 2, ..., 18 : numéro affiche
	13	14	15	16	17	18		
Sol	1	2	3	4	5	6	Sol : numéro 1, 2, ..., 18 : numéro zone sol	
	7	8	9	10	11	12		
	13	14	15	16	17	18		GrX : numéro de groupe

GR1	GR2	GR3	GR4	GR5	GR6
-----	-----	-----	-----	-----	-----

Figure 43 : Répartition de la position des groupes par rapport à l'écran durant l'expérience.

Par exemple les sujets du groupe 1 réalisent la lecture des 18 affiches (ensemble des affiches à l'écran) depuis la zone sol numéro 1, puis la lecture de 18 nouvelles affiches depuis la zone sol numéro 8 et enfin encore 18 autres affiches depuis la zone sol numéro 15.

Pour couvrir un jeu de données complet, le nombre d'utilisateur doit être un multiple de 6. Nous avons donc réparti les utilisateurs en 6 groupes. La répartition des groupes en fonction des zones au sol, assure que chaque utilisateur malgré son appartenance à un groupe soit confronté à tous les cas de consultation comparés et étudiés (cf. *Principe de l'expérience*).

Au démarrage de la lecture de l'affiche, l'expérimentateur déclenche l'enregistrement des données une fois que le premier mot présent sur l'affiche est lu et arrête l'enregistrement juste avant la lecture du dernier mot.

*Déroulement.* L'expérience consiste pour chaque utilisateur à lire des affiches à l'écran. Chaque affiche est composée d'un numéro, d'un titre (un nom commun) et de la définition du titre issue du site Larousse comme vu sur la *Figure 42*. Les affiches sont toutes d'égales dimensions. Celles-ci ont été choisies afin que le temps de lecture minimal à voix haute et de manière intelligible soit de 3 secondes. Le Kinect V2 possédant un taux de capture de trente images par seconde (30 FPS), nous obtenons un minimum théorique de 90 enregistrements par utilisateur et par affiche. Dans certains cas, la détection étant trop faible (nombre d'enregistrements strictement inférieur à quatre-vingt-dix), nous avons décidé que les données seraient retirées de l'analyse car jugées « corrompues », la mauvaise détection entraînant des valeurs aberrantes dont le moindre nombre d'enregistrements est un symptôme significatif.

*Enregistrement des données.* Lors de la lecture d'une affiche nous enregistrons les coordonnées 3D du point « head », les données des 3 rotations associées aux mouvements de la tête de l'utilisateur dans le plan du Kinect V2 et par rapport au plan écran. De la même façon, nous enregistrons les coordonnées « écran », en pixels correspondant au centre de la zone de focalisation calculée. La zone de focalisation est calculée par trigonométrie en fonction des rotations de la tête et de la distance de l'utilisateur à l'écran mesurées par le Kinect V2.

## Participants

Quarante-deux personnes réparties en 6 groupes ont participé à l'expérience. Dans le *Tableau 17* nous pouvons apprécier la répartition des utilisateurs dans chaque groupe selon leur âge, leur sexe, leur taille, s'ils sont droitiers ou gauchers de la main, selon leur œil directeur et s'ils portaient des lunettes durant l'expérience.

*Tableau 17 : Description de la population et de la répartition des groupes.*

	<b>Age</b>	<b>Sexe</b>	<b>Taille</b>	<b>Main dominante</b>	<b>Œil directeur</b>	<b>Lunettes</b>
<b>Population</b>	34 ± 14 ans	26H & 16F	1.71 ± 0.08 m	40D & 2G	34D & 8G	23
<b>Groupe 1</b>	39 ± 14 ans	4H & 3F	1.74 ± 0.10 m	6D & 1G	5D & 2G	4
<b>Groupe 2</b>	39 ± 14 ans	4H & 3F	1.71 ± 0.08 m	7D & 0G	5D & 2G	4
<b>Groupe 3</b>	34 ± 11 ans	3H & 4F	1.69 ± 0.10 m	7D & 0G	6D & 1G	4
<b>Groupe 4</b>	27 ± 6 ans	6H & 1F	1.75 ± 0.05 m	7D & 0G	6D & 1G	4
<b>Groupe 5</b>	28 ± 12 ans	5H & 2F	1.71 ± 0.05 m	6D & 1G	7D & 0G	3
<b>Groupe 6</b>	42 ± 16 ans	4H & 3F	1.67 ± 0.07 m	7D & 0G	5D & 2G	4

# Résultats

## Analyse des données

Les données ont été analysées avec le logiciel XLSTAT version 18.01.35929, logiciel développé par Addinsoft et s'intégrant dans Microsoft Excel. Nous prendrons comme seuil de significativité  $\alpha = 0,05$  pour l'étude statistique.

Nous présentons la comparaison des données par groupe. Chaque groupe contient la moyenne des valeurs pour la variable étudiée de chaque individu composant le groupe, lors de la lecture d'une affiche, soit 7 valeurs de moyenne de rotation de tête calculées sur un enregistrement de trois secondes comportant au minimum 90 valeurs.

Nous comparons les différents groupes entre eux en fonction des trois distances à l'écran (LS1, LS2 et LS3) et en fonction des différents cas de consultation des rotations en lacet (6 cas de consultation) et en tangage (3 cas de consultation) de la tête des individus :

- LS1 : correspond à la ligne au sol composée des zones au sol 1 à 6 ;
- LS2 : correspond à la ligne au sol composée des zones au sol 7 à 12 ;
- LS3 : correspond à la ligne au sol composée des zones au sol 13 à 18 ;

## Différence entre les angles de rotation théoriques calculés et les angles de rotation effectifs mesurés par le Kinect V2

Dans la première partie des résultats, nous indiquons les angles de rotation théoriques et effectifs mesurés par le Kinect V2 pour chaque cas de consultation. Nous avons calculé pour chaque individu de notre population l'angle de rotation de leur tête pour chaque cas de consultation, avec pour origine la tête de l'individu (en fonction de leur taille, de la distance à l'écran et de la ligne ou la colonne écran regardée). Les angles théoriques de rotation sont la moyenne des angles de rotation de la tête de l'ensemble des individus pour chaque cas de consultation. Les angles de rotation effectifs mesurés sont les angles donnés par le Kinect V2 mais ramenés à la même origine que les angles de rotation théoriques. Le *Tableau 18* présente les données pour la rotation en lacet et le *Tableau 19* pour la rotation en tangage. Pour maximiser le nombre de données recueillies par rapport au nombre d'utilisateurs que nous avons pu mobiliser, nous exploitons les symétries et translations vis-à-vis de l'écran pour chaque cas de consultation. Si nous prenons pour exemple le cas de consultation numéro 1 (Cas 1) de la rotation en lacet, le nombre de données récoltées pour ce cas est théoriquement de 70 pour une distance donnée. En effet, nous avons 6 groupes composés de 7 individus. Nous avons donc  $6 \times 7 = 42$  données possibles. Ensuite si nous considérons les symétries, nous avons 4 groupes sur 6 qui peuvent regarder à plus une colonne sur la droite et sur la gauche. Nous avons donc  $4 \times 7 = 28$  données supplémentaires. Nous arrivons à un total de 70 données théoriques prisent en compte pour le cas de consultation (Cas 1) de la rotation en lacet. Par construction le test repose donc sur un nombre de données théoriques différent pour chaque cas de consultation. Dans le *Tableau 18* et le *Tableau 19*, la dernière ligne de chaque cas indique le nombre de données prises en compte et le maximum théorique, la différence s'expliquant par les défauts de détections.

Les données du *Tableau 18* et du *Tableau 19* sont présentées comme suit :

moyenne angle théorique en degrés / moyenne angle mesuré en degrés  
 écart entre moyenne angle théorique et mesuré en degrés  
 nombre de mesures réelles / nombre de mesures théoriques.

*Tableau 18 : Comparaison des angles de rotation en lacet théoriques et des angles de rotation en lacet effectifs mesurés par le Kinect V2 en degrés.*

	<b>LS1 = 1,93m</b>	<b>LS2 = 2,6m</b>	<b>LS3 = 3,27m</b>
<b>Cas 0</b>	0 / 1,9 1,9 40 / 42	0 / 1,4 1,4 41 / 42	0 / 0,1 0,1 42 / 42
<b>Cas 1</b>	14,3 / 11,1 3,2 69 / 70	10,7 / 8,4 2,3 68 / 70	8,6 / 6,2 2,4 67 / 70
<b>Cas 2</b>	27,1 / 21 6,1 55 / 56	20,8 / 15,7 5,1 55 / 56	16,8 / 12,6 4,2 49 / 56
<b>Cas 3</b>	37,5 / 28,1 9,4 40 / 42	29,6 / 22,2 7,4 42 / 42	24,4 / 17,9 6,5 36 / 42
<b>Cas 4</b>	45,6 / 33,7 11,9 28 / 28	37,2 / 27,2 10 28 / 28	31,1 / 22,3 8,8 25 / 28
<b>Cas 5</b>	52 / 41,6 10,4 14 / 14	43,5 / 31,6 11,9 14 / 14	37 / 29,1 7,9 13 / 14

*Tableau 19 : Comparaison des angles de rotation en tangage théoriques et des angles de rotation en tangage effectifs mesurés par le Kinect V2 en degrés.*

	<b>LS1 = 1,93</b>	<b>LS2 = 2,6</b>	<b>LS3 = 3,27</b>
<b>Cas 1</b>	25,9 / 12,8 13,1 42 / 42	19,9 / 9,8 10,1 42 / 42	16 / 8,6 7,4 41 / 42
<b>Cas 2</b>	11,3 / 3,4 7,9 42 / 42	8,4 / 2,6 5,8 41 / 42	6,7 / 3,3 3,4 40 / 42
<b>Cas 3</b>	-4,9 / -6,9 2 41 / 42	-3,7 / -4,9 1,2 41 / 42	-2,9 / -2,4 -0,5 41 / 42

## Précision du Kinect V2 suffisante pour distinguer les cas de consultation

Dans un premier temps, afin de vérifier que le Kinect V2 permet de discriminer les angles de rotation de la tête (lacet et tangage) en fonction des différents cas de consultation cités dans la partie *Principe de l'expérience*, nous vérifions que nous obtenons des différences significatives des angles de rotations de la tête selon les 3 distances possibles (LS1, LS2 et LS3) et entre les différents cas de consultation. Autrement dit, lorsqu'un groupe regarde une affiche sur l'écran, par exemple l'affiche numéro 1, depuis une position donnée dans la zone d'interaction, par exemple la zone au sol numéro 1, les données récoltées en moyenne sont significativement différentes de celles observées lorsqu'il regarde l'affiche numéro 2 ou 7 depuis cette même zone au sol.

### Comparaison des groupes selon chaque cas de consultation pour la rotation en lacet en fonction des trois distances

#### *Distance la plus proche de l'écran LS1*

Les données suivent une distribution normale. Le *Tableau 20* présente la description des données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance la plus proche de l'écran (LS1) :

*Tableau 20 : Statistiques descriptives des données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance la plus proche de l'écran (LS1).*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>Yaw LS1   Cas 0</b>	40	-8,560	10,238	1,906	4,691
<b>Yaw LS1   Cas 1</b>	69	-0,952	22,673	11,134	5,399
<b>Yaw LS1   Cas 2</b>	55	5,565	34,113	20,984	6,333
<b>Yaw LS1   Cas 3</b>	40	15,182	40,650	28,126	6,224
<b>Yaw LS1   Cas 4</b>	28	20,093	46,234	33,701	6,732
<b>Yaw LS1   Cas 5</b>	14	29,323	53,253	41,594	8,141

Les échantillons n'ont pas tous des variances égales d'après le test de Levene de comparaison de plusieurs variances. Nous avons donc effectué un test non paramétrique s'appuyant sur des comparaisons multiples par paires suivant la procédure de Steel-Dwass-Critchlow-Fligner :

*Tableau 21 : P-values après procédure de Steel-Dwass-Critchlow-Fligner pour la comparaison des cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance intermédiaire (LS1).*

	Yaw LS1   0	Yaw LS1   1	Yaw LS1   2	Yaw LS1   3	Yaw LS1   4	Yaw LS1   5
Yaw LS1   0	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS1   1	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS1   2	< 0,0001	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001
Yaw LS1   3	< 0,0001	< 0,0001	< 0,0001	1	0,019	0,000
Yaw LS1   4	< 0,0001	< 0,0001	< 0,0001	0,019	1	<b>0,053</b>
Yaw LS1   5	< 0,0001	< 0,0001	< 0,0001	0,000	<b>0,053</b>	1

Le *Tableau 21* met en évidence que le cas 4 et le cas 5 ne sont pas significativement différents au seuil alpha 0.05.

Cependant, un test de Fisher nous permet de montrer que ces deux cas ont des variances comparables (p-value = 0,391). Nous avons donc pu appliquer un test t de comparaison des moyennes. Ce dernier nous permet de mettre en évidence une différence significative entre le cas 4 et le cas 5 avec une p-value = 0,002.

#### *Distance intermédiaire LS2*

Le *Tableau 22* présente les données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance intermédiaire (LS2) :

*Tableau 22 : Statistiques descriptives des données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance intermédiaire (LS2).*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Yaw LS2   0	41	-11,791	8,689	1,406	4,459
Yaw LS2   1	68	-0,712	19,089	8,432	4,295
Yaw LS2   2	55	0,873	26,178	15,704	5,722
Yaw LS2   3	42	7,240	30,716	22,174	5,625
Yaw LS2   4	28	10,995	38,604	27,219	7,076
Yaw LS2   5	14	17,650	42,569	31,619	7,414

Application de la même procédure que précédemment.

*Tableau 23 : P-values après procédure de Steel-Dwass-Critchlow-Fligner pour la comparaison des cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance intermédiaire (LS2).*

	Yaw LS2   0	Yaw LS2   1	Yaw LS2   2	Yaw LS2   3	Yaw LS2   4	Yaw LS2   5
Yaw LS2   0	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS2   1	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS2   2	< 0,0001	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001
Yaw LS2   3	< 0,0001	< 0,0001	< 0,0001	1	0,023	0,002
Yaw LS2   4	< 0,0001	< 0,0001	< 0,0001	0,023	1	<b>0,527</b>
Yaw LS2   5	< 0,0001	< 0,0001	< 0,0001	0,002	<b>0,527</b>	1

Le *Tableau 23* met en évidence que le cas 4 et le cas 5 ne sont pas significativement différents au seuil alpha 0.05 dans le cas d'un test de Steel-Dwass-Critchlow-Fligner.

Cependant, suite à un test de Mann-Whitney pour les cas 4 et 5 nous observons une p-value = 0,045. Nous pouvons donc conclure sur une différence significative.

#### *Distance la plus loin de l'écran LS3*

Le *Tableau 24* présente la description des données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance la plus loin de l'écran (LS3) :

*Tableau 24 : Statistiques descriptives des données étudiées pour comparer les cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance la plus loin de l'écran (LS3).*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Yaw LS3   0	42	-8,844	9,537	0,105	4,248
Yaw LS3   1	67	-4,773	15,884	6,247	3,897
Yaw LS3   2	49	1,627	21,158	12,555	4,813
Yaw LS3   3	36	6,863	25,461	17,886	5,346
Yaw LS3   4	25	7,909	33,465	22,320	7,015
Yaw LS3   5	13	11,923	38,914	29,121	8,207

Application de la même procédure que précédemment.

Tableau 25 : P-values après procédure de Steel-Dwass-Critchlow-Fligner pour la comparaison des cas de consultation par rapport à la rotation en lacet (Yaw) pour la distance la plus loin de l'écran (LS3).

	Yaw LS3   0	Yaw LS3   1	Yaw LS3   2	Yaw LS3   3	Yaw LS3   4	Yaw LS3   5
Yaw LS3   0	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS3   1	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001	< 0,0001
Yaw LS3   2	< 0,0001	< 0,0001	1	< 0,0001	< 0,0001	< 0,0001
Yaw LS3   3	< 0,0001	< 0,0001	< 0,0001	1	<b>0,098</b>	0,001
Yaw LS3   4	< 0,0001	< 0,0001	< 0,0001	<b>0,098</b>	1	<b>0,083</b>
Yaw LS3   5	< 0,0001	< 0,0001	< 0,0001	0,001	<b>0,083</b>	1

Le Tableau 25 prenant en compte le test de Steel-Dwass-Critchlow-Fligner met en évidence que le cas 4 et le cas 3 ne sont pas significativement différents au seuil alpha 0.05 et que le cas 4 et le cas 5 ne sont pas significativement différents au seuil alpha 0.05.

Néanmoins, suite à deux tests de Mann-Whitney entre le cas 3 et le cas 4, et le cas 4 et le cas 5, nous obtenons une p-value = 0,005 pour le premier test et une p-value = 0,004 pour le deuxième. Nous pouvons donc affirmer une différence significative entre le cas 3 et le cas 4, et le cas 4 et le cas 5.

### Comparaison des groupes selon chaque cas de consultation pour la rotation en tangage en fonction des trois distances

Nous avons suivi la même démarche statistique que précédemment pour la comparaison des trois cas de consultation de la rotation en tangage selon les trois distances à l'écran.

Les données suivent une distribution normale. Nous n'observons pas de différence significative de variance suite au test de Levene de comparaison de plusieurs variances.

Nous avons donc effectué un test paramétrique de comparaisons des moyennes de Tukey (cf. Tableau 26).

Tableau 26 : Tableau récapitulatif des p-values après un test de Tukey pour la comparaison des cas de consultation par rapport à la rotation en tangage (Pitch) et selon les 3 distances (LS1, LS2 et LS3).

	LS1	LS2	LS3
Cas 1 vs Cas 2	< 0,0001	0,0004	0,003
Cas 1 vs Cas 3	< 0,0001	< 0,0001	< 0,0001
Cas 2 vs Cas 3	< 0,0001	0,005	0,002

Nous obtenons une différence significative en moyenne sur la rotation en tangage entre les groupes selon les cas de consultation et ceci quel que soit la distance par rapport à l'écran. Les groupes qui regardent la ligne supérieure de l'écran (LE1) ont des moyennes de rotation en tangage significativement différentes des groupes regardant la ligne écran centrale (LE2) ou la ligne écran inférieure (LE3) et ceci quelle que soit leur position (LS1, LS2 et LS3) par rapport à l'écran.

## Caractérisation des mouvements de la tête via la Kinect V2

Dans un second temps, nous réalisons la même démarche statistique afin de tester l'hypothèse suivante : lorsque chaque groupe est dans le même cas de consultation (Ex : les groupes regardent à plus une colonne écran vers la droite en fonction de leur position), il n'y a pas de différence significative entre les groupes. Nous réalisons l'analyse pour chaque cas pour la rotation en lacet et pour les trois cas de la rotation en tangage et ceci en fonction des trois distances possibles à l'écran (LS1, LS2 et LS3).

### Comparaison des groupes cas par cas de consultation pour la rotation en lacet en fonction de chaque distance

#### *Cas 0 : les groupes regardent en face d'eux*

Nous cherchons s'il existe une différence entre les groupes lorsqu'ils regardent en face d'eux. Les données suivent une loi normale, nous avons donc effectué un test paramétrique de comparaisons des moyennes de Tukey pour chaque distance LS1, LS2 et LS3.

Pour la distance la plus proche de l'écran, le test met en évidence une différence significative entre le groupe 3 et le groupe 6, avec une p-value de 0.046. Pour la distance intermédiaire et la distance la plus éloignée, on n'observe aucune différence significative entre les groupes.

Nous allons maintenant présenter les résultats liés aux conditions expérimentales suivantes : les groupes d'utilisateurs regardent à plus une, plus deux, plus trois, plus quatre et plus cinq colonne(s) écran (CE) par rapport à leur position et ceci quelle que soit la ligne écran consultée (trois hauteurs possibles, LE1, LE2 et LE3). Nous présenterons ces résultats en fonction des trois distances (LS1, LS2, LS3) possibles à l'écran comme dans le cas précédent. Suite aux résultats présentés précédemment, nous ne supposons pas dans un premier temps la symétrie de notre système. C'est pourquoi, nous comparons les résultats des utilisateurs regardant à plus une, plus deux, plus trois, plus quatre et plus cinq colonne(s) vers la droite (Ex : le groupe est en colonne 1 au sol et à la distance LS1 et il regarde en colonne 2 de l'écran). Puis, lorsqu'ils regardent à plus une, plus deux, plus trois, plus quatre et plus cinq colonne(s) vers la gauche.

#### *Cas 1 : les groupes regardent à plus une colonne écran par rapport à leur position*

*Tableau 27 : Résultats après test de Tukey comparant les groupes regardant à plus une colonne écran (CE) vers la droite.*

Cas 1 vers la droite					
LS1		LS2		LS3	
CE2 vs CE6	0,007	Pas de Différence Significative (PDS)		CE2 vs CE6	0,036
CE4 vs CE6	0,034				

Le *Tableau 27* montre que, pour la distance la plus proche de l'écran, et selon le test de Tukey, il y a une différence significative en moyenne pour le groupe 1 regardant à plus une colonne vers la droite, soit en colonne écran 2, et le groupe 5 regardant à plus une colonne vers la droite, soit en colonne écran 6. Même constat pour le groupe 3 regardant à plus une

colonne vers la droite, soit en colonne écran 4 et le groupe 5 regardant en plus une colonne vers la droite, soit en colonne écran 6. Nous observons également, une différence significative entre le groupe 1 et le groupe 5, pour la distance à l'écran la plus éloignée.

Pour les groupes, mais regardant à plus une colonne écran vers la gauche, nous n'observons aucune différence significative et ce quel que soit la distance à l'écran.

*Cas 2 : les groupes regardent à plus deux colonnes écran par rapport à leur position*

Quelle que soit la distance à l'écran, nous n'observons aucune différence significative entre les groupes regardant à plus deux colonnes écran vers la droite. De même lorsque les groupes regardent à plus deux colonnes écran vers la gauche.

*Cas 3 : les groupes regardent à plus trois colonnes écran par rapport à leur position*

Quelle que soit la distance à l'écran, nous n'observons aucune différence significative entre les groupes regardant à plus trois colonnes écran vers la droite. De même lorsque les groupes regardent à plus trois colonnes écran vers la gauche.

*Cas 4 : les groupes regardent à plus quatre colonnes écran par rapport à leur position*

*Tableau 28 : Résultats après test de Tukey comparant les groupes regardant à plus quatre colonnes écran (CE) vers la droite.*

Cas 4 vers la droite			
LS1		LS2	LS3
CE5 vs CE6	0,024	PDS	PDS

Le *Tableau 28* montre que, pour la distance la plus proche de l'écran, il y a une différence significative selon le test de Tukey pour les groupes 1 regardant à plus quatre colonnes vers la droite, soit en colonne écran 5, et le groupe 2 regardant à plus quatre colonnes vers la droite, soit en colonne écran 6.

Aucune différence significative n'a été trouvée concernant les groupes regardant à plus quatre colonnes écran vers la gauche.

*Cas 5 : les groupes regardent à plus cinq colonnes écran par rapport à leur position*

Quelle que soit la distance à l'écran, selon le test de Tukey, nous n'observons aucune différence significative entre les groupes regardant à cinq colonnes écran vers la droite. De même lorsque les groupes regardent à cinq colonnes écran vers la gauche.

**Comparaison des groupes cas par cas de consultation pour la rotation en tangage en fonction de chaque distance**

Nous allons maintenant présenter les résultats de l'étude de la variabilité de la moyenne sur la rotation en tangage de la tête de chaque utilisateur par rapport à l'écran, lorsque ceux-ci font face à l'écran et lisent les affiches selon les 3 hauteurs possibles (LE1, LE2 et LE3). Nous souhaitons ainsi démontrer que, quels que soient les groupes d'utilisateurs, la variabilité des mesures prises dans ces cas n'est significativement pas différente d'un groupe à l'autre. Nous présentons donc ci-dessous les moyennes de la rotation en tangage selon les 3 distances possibles à l'écran (LS1, LS2 et LS3) et en fonction des 3 hauteurs (LE1, LE2 et LE3).

*Cas 1 : les groupes regardent la ligne supérieure de l'écran*

Nous n'obtenons aucune différence significative selon le test de Tukey pour la moyenne de rotation en tangage de la tête lorsque les groupes regardent la ligne supérieure de l'écran LE1, et ceci quelle que soit la distance à l'écran.

*Cas 2 : les groupes regardent la ligne intermédiaire de l'écran*

Nous n'obtenons aucune différence significative selon le test de Tukey pour la moyenne de rotation en tangage de la tête lorsque les groupes regardent la ligne intermédiaire de l'écran LE2, et ceci quelle que soit la distance à l'écran.

*Cas 3 : les groupes regardent la ligne inférieure de l'écran*

Nous n'obtenons aucune différence significative selon le test de Tukey pour la moyenne de rotation en tangage de la tête lorsque les groupes regardent la ligne inférieure de l'écran LE3, et quelle que soit la distance à l'écran.

# Discussion

## Le capteur Kinect V2

Les premiers résultats présentés dans la partie *Différence entre les angles de rotation théoriques calculés et les angles de rotation effectifs mesurés par le Kinect V2* montrent que : plus les utilisateurs sont éloignés du capteur Kinect V2, plus l'écart entre les angles théoriques et les angles mesurés par le capteur sont faibles. Ces résultats sont étonnants car avec l'augmentation de la distance entre le capteur et les points mesurés nous devrions obtenir une précision plus faible. Néanmoins, nous observons que : plus les rotations sont importantes, plus les écarts entre angles réels et angles théoriques augmentent. Ce comportement est cohérent. Plus les rotations de tête sont importantes moins l'utilisateur fait face à la caméra car il tourne la tête de manière importante afin d'éviter des angles inconfortable pour les cervicales et donc complique la détection des points caractéristiques du visage nécessaire à un *head tracking* de qualité. Cette occlusion/superposition des points caractéristiques du visage lors des rotations importantes de la tête doit également s'opérer lorsque les utilisateurs sont à la distance la plus proche de l'écran. En effet, le Kinect V2 étant placée sur le dessus de l'écran et inclinée vers le bas, l'angle des rayons incidents arrivant sur le visage des utilisateurs et permettant la détection des points caractéristiques est plus important. Cela explique que plus les utilisateurs sont éloignés (jusqu'à un certain point) par rapport au capteur plus les écarts entre les angles théoriques et mesurés sont faibles.

## Caractérisation des mouvements de la tête

### Observation de différences significative entre les angles de rotation de la tête pour des cas de consultation différents

Les résultats présentés dans la partie *Comparaison des groupes selon chaque cas de consultation pour la rotation en lacet en fonction des trois distances* montrent de manière significative que, dans notre configuration physique, le Kinect V2 est assez précis pour discriminer les différences d'angle de rotation en lacet lorsqu'un groupe regarde en face de lui, à plus une, plus deux, plus trois, plus quatre et plus cinq colonnes écran pour les trois distances. Si nous continuons notre analyse avec la rotation en tangage, les résultats significatifs dans la partie *Comparaison des groupes selon chaque cas de consultation pour la rotation en tangage en fonction des trois distances* montrent que le Kinect V2 est suffisamment précis pour discriminer les groupes regardant la ligne supérieure, intermédiaire ou inférieure de notre écran.

### Absence de différence significative entre les angles de rotation de la tête pour les même cas de consultation

L'ensemble des résultats présentés dans la partie *Caractérisation des mouvements de la tête via la Kinect V2* nous amène à penser que notre installation n'était pas complètement symétrique. En effet, si nous observons les résultats des cas 1 et 4, nous remarquons que des différences significatives ne sont observées que dans les cas de rotation en lacet vers la droite. De plus, en regardant en détail ces résultats, il apparaît que ce sont exclusivement les cas concernant la consultation de la colonne écran 6 qui présentent des différences significatives et cela uniquement lors de rotation en lacet vers la droite. Malgré le soin apporté à

l'installation, il semble qu'une légère inclinaison parasite du capteur ait grevé la symétrie attendue du dispositif. Les cas montrant des différences significatives laissent également penser que nous sommes dans des conditions limites de détection, que ce soit en ce qui concerne le positionnement de l'utilisateur par rapport au capteur ou dans le cas de consultation impliquant des rotations de tête importantes comme expliqué dans la partie *Le capteur Kinect V2*. Ce problème pourrait être dépassé en couplant trois Kinect V2. En effet, en ajoutant un Kinect V2 sur chaque extrémité de l'écran et en les orientant de manière à couvrir la zone d'interaction, les utilisateurs feraient toujours face à un des 3 Kinect V2 quelle que soit leur rotation de tête. Dans une prochaine expérience, nous prévoyons d'utiliser *LiveScan3D* (Kowalski, Naruniec et al. 2015) pour la fusion des données de plusieurs Kinect V2 en temps réel.

Enfin, l'analyse de la rotation en lacet selon chaque cas de consultation, permet d'observer que pour les cas 2, 3 et 5, aucune différence significative n'est constatée entre les groupes. Par ailleurs, l'analyse des résultats pour la rotation en tangage ne met en évidence aucune différence significative selon les 3 distances pour les trois cas de consultation.

L'ensemble de cette étude nous permet de conclure que le Kinect V2 est un capteur assez précis pour réaliser des interactions à base de *head tracking* avec des écrans de grand format dans la configuration physique présentée. En effet, grâce aux données fournies par le capteur concernant les rotations de tête, nous sommes capables de discriminer les différents cas de consultation. Nous confirmons donc statistiquement nos deux premières hypothèses :

- (H1) la moyenne en degré de la rotation en lacet mesurée est significativement différente d'un cas de consultation à l'autre ;
- (H2) la moyenne en degré de la rotation en tangage mesurée est significativement différente d'un cas de consultation à l'autre.

Enfin, les résultats concernant les différents cas de consultation d'informations en fonction des deux rotations de la tête conduisent à penser que, de manière générale, nous orientons tous la tête de manière comparable (à travers les données mesurées par le Kinect V2) pour consulter confortablement une information nécessitant une lecture d'au moins 3 secondes sur un écran de grand format à plus de deux mètres. Ce qui nous conduit également à confirmer nos deux dernières hypothèses :

- (H3) la moyenne en degré de la rotation en lacet mesurée n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation ;
- (H4) la moyenne en degré de la rotation en tangage mesurée n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation.

Ces résultats ne sont pas à comparer directement avec la distinction de styles *visionautes versus céphalonautes* expliquée dans la partie *Les yeux et la tête : un système masse-ressort*. En effet, nous sommes parfaitement conscients qu'ils existent des différences interindividuelles au niveau de l'utilisation du système masse-ressort « tête/yeux ». Dans cette étude, nous nous intéressons à la posture adoptée pour une lecture confortable, pas à la dynamique d'atteinte de cette posture qui est un paramètre pris en compte dans la distinction des styles *visionautes versus céphalonautes*. Néanmoins, à partir d'une certaine distance, ces différences interindividuelles semblent s'atténuer au regard des données mesurées par le capteur Kinect V2.

## Conclusions<sup>15</sup>

Dans un premier temps, cette expérience nous a permis de déterminer que le capteur Kinect V2 est assez précis pour réaliser des interactions via *head tracking* sur écran de grand format. En effet, dans le cas de notre configuration physique, le Kinect V2 est suffisamment précis pour discriminer des cas de consultation différents.

Elle a également mis en évidence la possibilité de caractériser les mouvements moyens de rotation en tangage et en lacet de la tête des utilisateurs, même si un problème de symétrie dans notre système ne nous permet pas de confirmer dans tous les cas nos hypothèses de départ. Grâce à ces résultats nous sommes en mesure de nous abstraire d'une calibration individuelle du système avant chaque utilisation. Nous nous rapprochons ainsi de notre objectif principal à savoir d'autoriser des interactions avec un écran de grand format de plusieurs utilisateurs non instrumentés, sans calibrage.

A l'aune de ces résultats, il apparaît que nous orientons tous la tête de manière comparable pour une tâche de consultation d'informations sur un grand écran nécessitant une lecture. En tous cas, malgré l'existence de différences interindividuelles au niveau des rotations de la tête, celles-ci ne sont en moyennes pas significativement différentes.

Ainsi, dans le chapitre suivant nous décrivons notre système permettant des interactions individualisées sur écran de grand format via un *head tracking* réalisé par un Kinect V2.

## Résumé

Dans ce chapitre, nous avons présenté un système basé sur la détection des mouvements de tête. Il nous permet de caractériser l'orientation moyenne de la tête lors de la consultation d'un affichage de grand format. Les résultats montrent que dans l'ensemble nous orientons tous notre tête de manière comparable lors de la lecture d'une information à une distance donnée. Cette caractérisation des mouvements de tête via le Kinect V2, nous a permis de développer un système d'*head tracking* autorisant une interaction intuitive, sans appareillage ni calibrage individuel qui est décrit dans le chapitre suivant. Ce travail permet, pour plusieurs utilisateurs simultanément, la consultation d'informations individualisées sur un affichage de grand format.

---

<sup>15</sup> Cette étude a fait l'objet d'une présentation à la conférence IHM'17 : Marion, D., P.-A. Favier and J.-M. André (2017). Caractérisation des mouvements de la tête lors d'une tâche de consultation d'un affichage de grand format à l'aide d'un dispositif de head tracking. 29ème conférence francophone sur l'Interaction Homme-Machine, ACM.



# Chapitre 5 : Evaluation d'une interaction multi-utilisateur, non appareillés et sans calibrage via head tracking avec les écrans de grand format.

Dans ce chapitre, nous présentons une étude comparative de deux modes d'affichage. Ceux-ci présentent les mêmes informations. Cependant, dans l'un des deux modes, les utilisateurs peuvent dupliquer et déplacer de l'information à l'aide d'interactions basées sur le calcul de la projection de leur zone d'attention sur l'écran à partir de la détection de la position et de l'orientation de leur tête.

En comparant la performance de consultation et le ressenti des utilisateurs par rapport à ces deux modes d'affichage, nous souhaitons démontrer l'intérêt de rendre les affichages interactifs à l'aide d'interaction basée sur les mouvements de tête. Ceci, dans le but d'améliorer l'expérience utilisateurs.

Ce premier recueil du ressenti des utilisateurs vis-à-vis de notre dispositif, nous a permis de mettre en évidence la pertinence des interactions via *head tracking* proposées mais également de mettre en lumière les premiers faux pas les concernant.

## Introduction

### Contexte

Ces grands écrans font partie intégrante de notre vie (Colley, Pakanen et al. 2016). Cependant, dans la plupart des cas, ils présentent soit des informations statiques (qui n'évoluent pas) soit du contenu en boucle d'affichage. Ainsi, l'absence d'interactivité avec les informations présentées entraîne une faible audience envers ces écrans considérés comme des supports d'information peu intéressants. Dans certains cas, les grands écrans sont des supports d'informations indispensables pour les utilisateurs comme dans les lieux publics (gare, aéroport, musée, etc.). Mais ils deviennent vite difficilement consultables en raison des

conditions de confort peu convenables. En effet, qui ne s'est jamais retrouvé gêné par un amas de personnes agglutinées devant un grand écran contenant les informations dont il avait besoin ? Chacun se met alors à jouer des coudes afin de pouvoir lire les informations désirées. Néanmoins, n'est-il pas possible d'imaginer que l'initiative du déplacement soit donnée à l'affichage. Si, en détectant nos besoins de consultation, celui-ci affichait l'information voulue dans des conditions de confort optimales ? Si le simple fait d'adresser une information la faisait venir à nous ? Effectivement, l'écran en détectant la position et l'orientation de la tête par rapport au contenu affiché à l'écran pourrait calculer la zone de focalisation des individus qui le regardent. Ces zones d'attention individuelles ou collectives déduites, il serait alors en mesure de réorganiser son contenu afin qu'il soit consultable dans les meilleures conditions possibles. Ainsi, ce ne serait plus les utilisateurs qui se déplaceraient devant l'écran mais l'information qui s'organiserait en fonction.

Notre première expérience nous a permis de confirmer la faisabilité d'une interaction via *head tracking* avec les écrans de grand format. La deuxième expérience a mis en évidence que le capteur Kinect V2 est une solution envisageable pour la mise en œuvre d'une interaction multi-utilisateur, non appareillés et sans calibrage avec les écrans de grand format. Il reste donc à identifier et évaluer les règles d'adaptation de l'interface en prenant en compte les principales contraintes d'un *head tracking* sans calibrage et multi-utilisateurs. Comment, sélectionner, désélectionner, déplacer, dupliquer, zoomer ou encore réorganiser les informations sur l'écran sans gêner les utilisateurs. Comment rendre ces interactions affordantes, naturelles et intuitives ?

## Identification des zones d'attention sur l'écran

Le premier enjeu est d'être capable de calculer les zones d'attention des individus sur l'écran. Pour rappel, dans nos travaux, la zone d'attention est la partie sur l'écran qui correspond à la projection directe du champ de vision en fonction de la position et de l'orientation de la tête et non du regard. Cette projection du champ visuel dans le plan écran est caractérisée par une ellipse et se calcule aisément en fonction de la position et de l'orientation de la tête d'un individu. La zone d'attention sert alors d'interacteur avec l'écran.

### Calcul de la zone d'attention sur l'écran

Pour calculer la position (en pixels) de la zone d'attention d'un individu sur un écran, nous devons donc connaître à minima :

- Sur l'écran :
  - ses dimensions physiques ;
  - sa résolution d'affichage.
  
- Sur l'individu :
  - la position x, y et z de sa tête par rapport à l'écran en mètres ;
  - l'orientation de sa tête par rapport à l'écran (en degrés).

Pour rappel, le Kinect V2 fournit directement via son SDK, les coordonnées (x, y et z) de la tête en mètres et les 3 rotations de la tête (en lacet, en tangage et en roulis) en degrés par rapport à son centre, de six individus en même temps.



Figure 44 : Plan de référence des données fournies par la Kinect : Plan Kinect.

La première étape consiste donc à étalonner le Kinect V2 en fonction de sa position par rapport à l'écran. Dans notre cas, celle-ci est placée et centrée au-dessus de l'écran et inclinée de  $34,8^\circ$  vers le sol comme décrit dans l'expérience précédente (cf. Chapitre 4). Nous transformons donc les données brutes fournies par le Kinect V2 par rapport à son centre en données dites rectifiées par rapport au centre de l'écran.

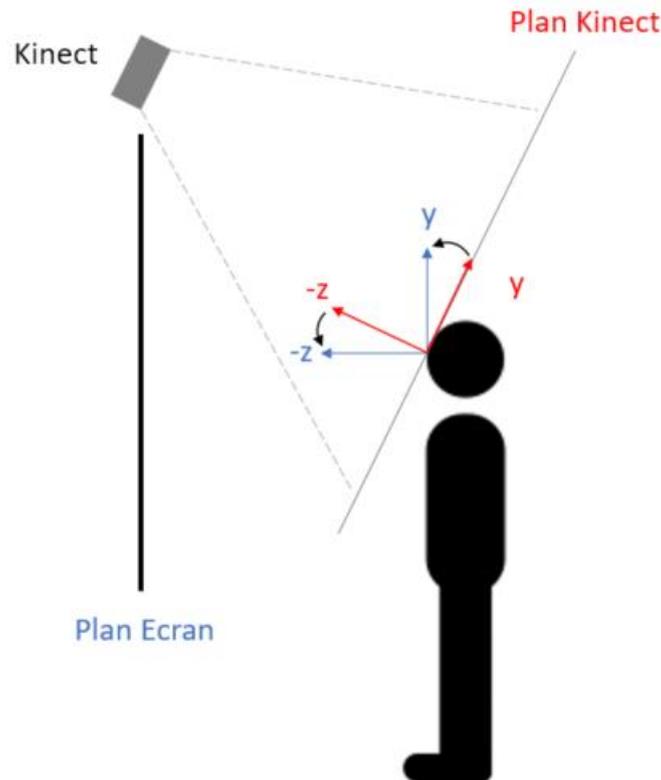


Figure 45 : Schéma du changement du plan de référence, du Plan Kinect vers le Plan Ecran.

Nous obtenons ainsi les coordonnées ( $x$ ,  $y$  et  $z$ ) de la tête et les 3 rotations de la tête (en lacet, en tangage et en roulis) en fonction du Plan Ecran. Nous nommons cette première transformation : Etalonnage.

Une fois l'Etalonnage réalisé, nous calculons la projection de la position ( $x$  et  $y$ ) de la tête sur l'écran (en pixels) selon la normale entre la tête et l'écran de la façon suivante :

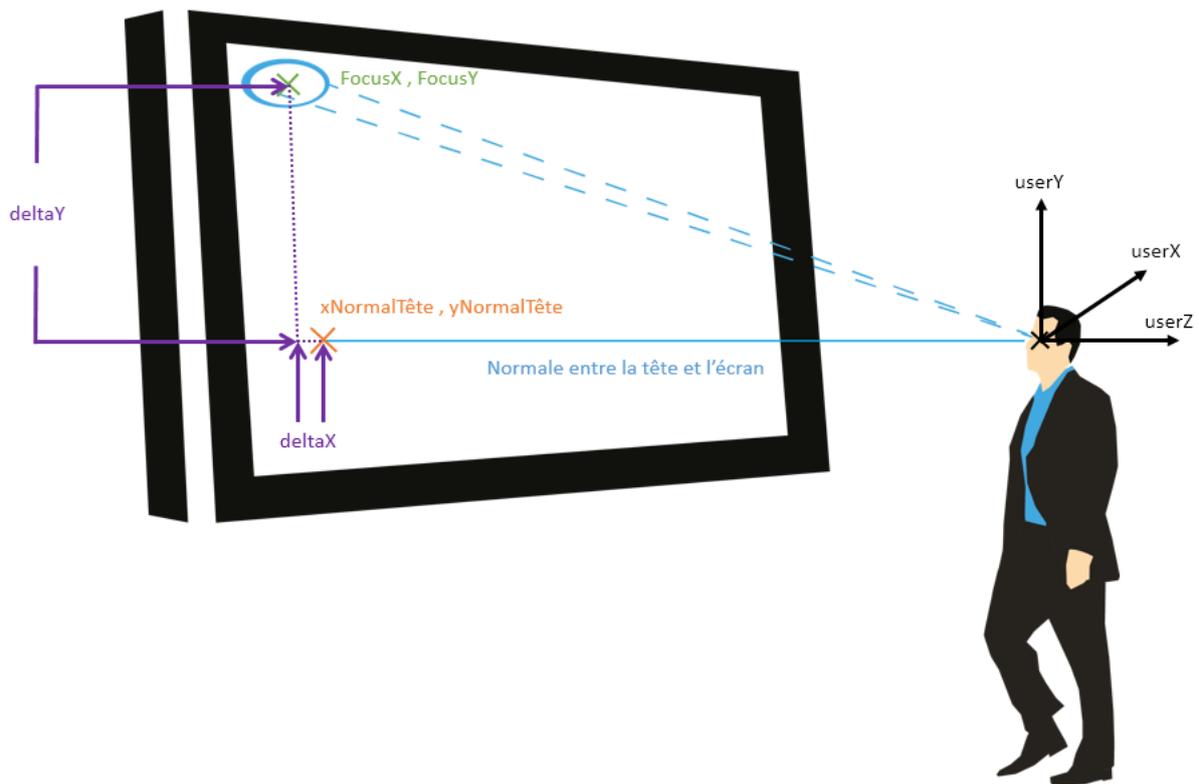


Figure 46 : Calcul de la zone d'attention sur l'écran via les données Kinect V2 après Etalonnage.

Le calcul des coordonnées sur l'écran de l'intersection de la normale entre l'écran et la tête de l'individu est le suivant :

Pour la coordonnée x :

$$xNormalTête = userX * (ResH / L)$$

Avec :

- $xNormalTête$  = position en x de l'intersection de la normale entre l'écran et la tête de l'individu ;
- $userX$  = position en x de la tête ( $userX$ ) en mètres, qualifie la position droite-gauche de la tête par rapport au centre de l'écran ;
- $ResH$  = résolution horizontale de l'écran en pixels, 1920 px ;
- $L$  = largeur physique de l'image affichée en mètres, 2,96 m.

Et pour la coordonnée y :

$$yNormalTête = userY * (ResV / H)$$

Avec :

- $yNormalTête$  = position en y de l'intersection de la normale entre l'écran et la tête de l'individu ;
- $userY$  = position en y de la tête ( $userY$ ) en mètres, qualifie la hauteur de la tête par rapport au centre de l'écran ;
- $ResV$  = résolution verticale de l'écran en pixels, 1080 px ;
- $H$  = hauteur physique de l'image affichée en mètres, 1.66 m.

Nous calculons ensuite le décalage induit par les rotations en lacet (deltaX) et en tangage (deltaY) :

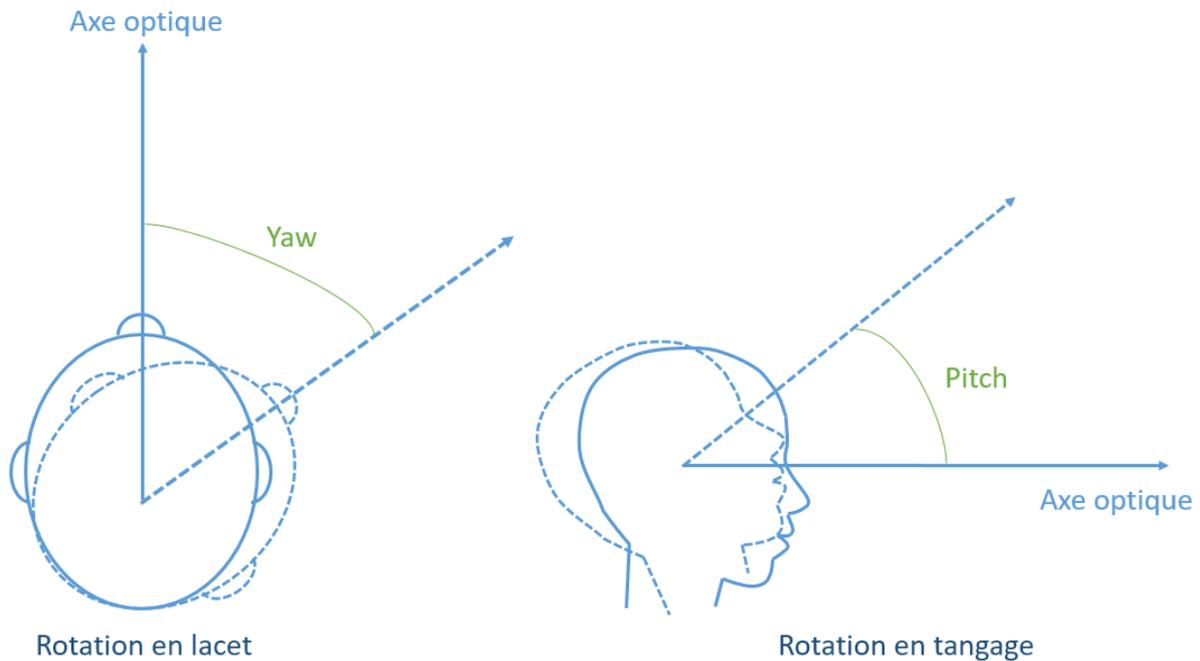


Figure 47 : Description des angles de rotation de la tête.

Pour la rotation en lacet :

$$\text{deltaX} = \text{Tan}(\text{Yaw}) * \text{userZ} * \text{ResH}$$

Avec :

- deltaX = décalage en x induit par la rotation en lacet, en pixels ;
- Tan(Yaw) = tangente de l'angle de rotation en lacet, en degrés ;
- userZ = position en z de la tête (userZ) en mètres, qualifie la distance entre la tête de l'individu et le centre de l'écran ;
- ResH = résolution horizontale de l'écran en pixels, 1920 px.

Pour la rotation en tangage :

$$\text{deltaY} = \text{Tan}(\text{Pitch}) * \text{userZ} * \text{ResV}$$

Avec :

- deltaY = décalage en y induit par la rotation en tangage, en pixels ;
- Tan(Pitch) = tangente de l'angle de rotation en tangage, en degrés ;
- userZ = position en z de la tête (userZ) en mètres, qualifie la distance entre la tête de l'individu et le centre de l'écran ;
- ResV = résolution verticale de l'écran en pixels, 1080 px.

Dernière étape, nous appliquons les décalages induits par les rotations de la tête à la projection de la position de la tête sur l'écran pour déterminer la position de la zone d'attention :

Position de la zone d'attention en x :

$$\text{FocusX} = \text{deltaX} + \text{xNormalTête}$$

Avec :

- FocusX = position en x de la zone d'attention sur l'écran en pixels ;
- deltaX = décalage en x induit par la rotation en lacet en pixels ;
- xNormalTête = position en x de l'intersection de la normale entre l'écran et la tête de l'individu en pixels.

Position de la zone d'attention en y :

$$\text{FocusY} = \text{deltaY} + \text{yNormalTête}$$

Avec :

- FocusY = position en y de la zone d'attention sur l'écran en pixels ;
- deltaY = décalage en y induit par la rotation en tangage en pixels ;
- YNormalTête = position en y de l'intersection de la normale entre l'écran et la tête de l'individu en pixels.

La zone d'attention calculée n'autorise pas encore une interaction précise. Dans le but de calculer de manière précise ces zones d'attention, nous avons utilisé les données récoltées dans l'expérience précédente (*cf. Chapitre 4*). Nous avons également conservé la configuration physique mise en place et le capteur Kinect V2 utilisé.

Nous avons donc testé, dans un premier temps, un réseau de neurones artificiels afin d'estimer de manière précise la zone d'attention de chaque utilisateur à l'écran en fonction de la position et des trois angles de rotation de la tête (*cf. A l'aide d'un réseau de neurones artificiels*). Dans un second temps, les résultats obtenus à l'aide des réseaux de neurones n'étant pas concluants, nous avons utilisé une autre technique afin de déterminer de manière précise la zone d'attention à l'écran des utilisateurs (*cf. A l'aide de moyennes mobiles et des écarts moyens mesurés*).

## A l'aide d'un réseau de neurones artificiels

Un réseau de neurones artificiels (Rosenblatt 1958) est un outil/une technique utilisé dans le domaine de l'informatique décisionnelle (Kimball et Ross 2011) qui s'inscrit dans le courant de pensée de l'intelligence artificielle (IA) (Turing 1950, McCarthy 1960, McCarthy 1960, Mynski et Papert 1969). Cette technique consiste à reproduire informatiquement l'organisation et le fonctionnement d'un réseau de neurones naturels. En effet, les réseaux de neurones artificiels ont été développés à partir de l'hypothèse que le comportement intelligent humain est le résultat de la structure et du fonctionnement des éléments de base du système nerveux central (que sont les neurones).

Les réseaux de neurones artificiels sont principalement utilisés pour de la prédiction, de la classification et de l'analyse de données. Le fonctionnement des neurones artificiels/formels est simple et calqué sur celui des neurones biologiques.

### *Neurone biologique*

Un neurone biologique (*cf. Figure 48*) est une cellule nerveuse constituant la base du système nerveux. Ce neurone est spécialisé dans le traitement des signaux électriques. A l'une de ses

extrémités, il est composé de plusieurs branches nommées dendrites dont la fonction est la réception des informations transmises par d'autres neurones. Ces informations transitent ensuite vers le corps cellulaire (centre de contrôle) qui traite la somme des informations. L'information traitée est alors envoyée sous forme de signaux électriques au travers de l'axone jusqu'à la connexion avec un autre neurone appelée synapse. Un neurone permet donc la communication et le traitement de l'information. Le cerveau humain est un système apprenant basé sur une architecture en réseau des neurones.

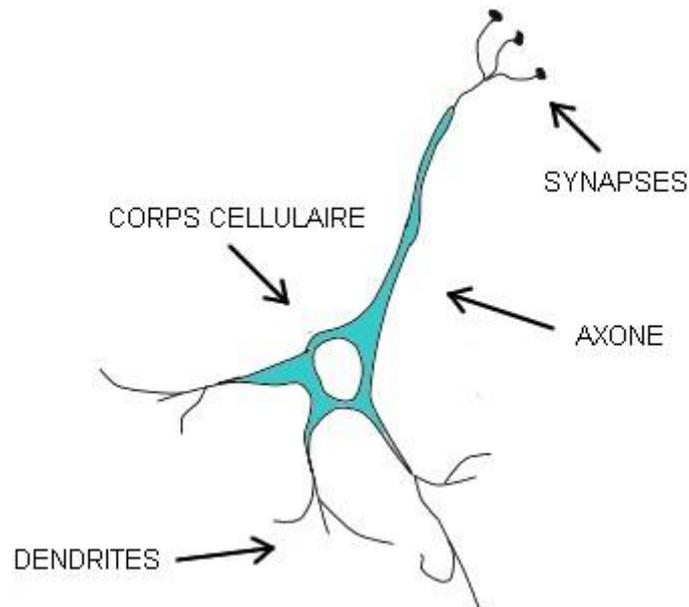


Figure 48 : Schéma d'un neurone biologique.

#### *Neurone formel/artificiel*

Le neurone formel est composé d'un signal d'entrée (dendrites), d'un signal de sortie (axones) et d'un poids de la connexion (synapse) à l'image du neurone biologique.

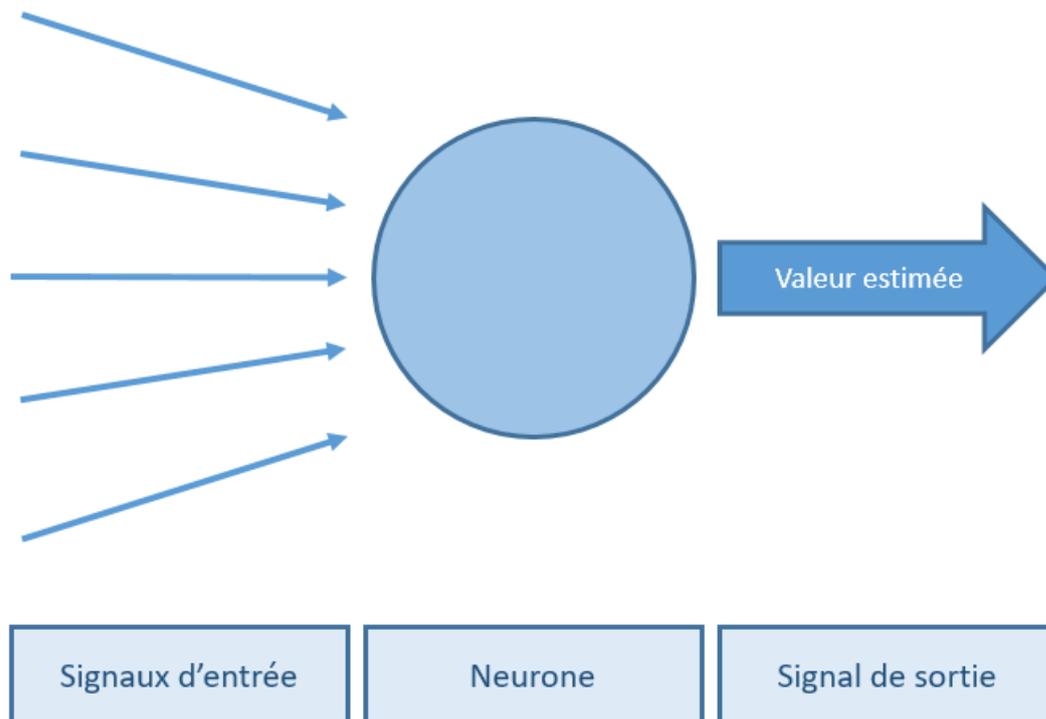


Figure 49 : Schéma d'un neurone formel

Le neurone artificiel est une unité élémentaire qui reçoit plusieurs variables d'entrées provenant des neurones situés en amont. A chacune des variables d'entrée est associé un poids qui représente la force de connexion entre deux neurones. Ce neurone calcule alors une sortie unique basée sur les informations reçues en entrée qui alimentera d'autres neurones en aval. Son action se décompose en deux phases :

- 1<sup>ère</sup> phase : le neurone calcule la somme pondérée des entrées (en fonction de la force des connexions) ;
- 2<sup>ème</sup> phase : à partir de la somme pondérée des entrées, une fonction d'activation (fonction à seuil, sigmoïde, gaussienne, etc.) calcule la valeur de l'état du neurone. Le neurone compare alors sa valeur d'état à une valeur de seuil et fournit une sortie.

#### *Perceptron multicouche avec rétro-propagation du gradient d'erreur*

Ainsi, en associant plusieurs neurones nous obtenons un réseau. Il existe plusieurs types de réseaux (réseau partiellement connectés, à connexion complète ou à couches, etc.) en fonction de la manière dont les neurones sont connectés entre eux. Nous avons choisis d'utiliser un réseau à couches et plus particulièrement un perceptron multicouche avec rétro-propagation du gradient d'erreur (LeCun 1985, Touzet 1992). Ce type de réseau présente deux capacités majeures : la généralisation et l'apprentissage. Pour ce type de réseau, l'apprentissage est dit « supervisé ». La mise en œuvre de ce type de réseau se découpe en 3 étapes :

- la première étape consiste à réaliser l'apprentissage. Le but de cette phase est de modifier le comportement du réseau (force des connexions) afin d'obtenir le résultat désiré. L'apprentissage consiste donc à déterminer la force optimale des connexions en fonction de la sortie souhaitée. Une partie de l'échantillon de données est donc utilisée pour le paramétrage de l'apprentissage. Nous présentons des exemples au

réseau, il calcule alors des valeurs de sortie. Au fur et à mesure, nous calculons l'erreur en sortie et en fonction du taux d'erreur, nous réajustons les poids de connexions des différentes couches jusqu'à obtenir un taux d'erreur satisfaisant.

- la deuxième étape consiste à utiliser un échantillon de données pour la validation de l'apprentissage. Si le réseau a trop appris, il se spécialise et il ne sera pas capable de donner les sorties attendues pour des exemples inconnus (mais reconnaîtra les exemples ayant servi à l'apprentissage avec un faible taux d'erreur). Ce phénomène est appelé sur-apprentissage. La généralisation dépend donc de cette étape. La phase d'apprentissage doit être arrêtée à partir du moment où le réseau fournit les meilleurs résultats de généralisation (meilleur taux de réponse sur des exemples inconnus).
- la dernière étape consiste à tester le réseau sur des échantillons de données inconnus.

Techniquement, nous avons utilisé la librairie AForge.Neuro du framework AForge.NET (AForge.NET 2008-2012) car elle permet une fois l'apprentissage terminé de sauvegarder les réseaux de neurones (structure et poids des connexions de chaque couche) et donc de les réutiliser dans un autre programme.

Nous avons conservé le quadrillage d'affiche à l'écran de l'expérience précédente (*cf. Figure 42*) et souhaitions que le réseau de neurone nous fournisse le numéro de l'affiche sur laquelle est située la zone d'attention de l'utilisateur en fonction des données fournies en temps réel par le Kinect V2. Malheureusement, cette technique qui nous a pris un temps non négligeable dans sa mise en œuvre n'a pas apporté la solution espérée. Malgré un apprentissage réalisé sur les données récoltées lors de la deuxième expérience (*cf. Chapitre 4*) et suite à de nombreux tests, nous n'étions pas capables de calculer en temps réel et de manière efficace la zone d'attention en fonction des données fournies par le Kinect V2. Le principal problème des réseaux de neurones est qu'ils ne fournissent pas d'explications concernant leurs résultats. Cela limite fortement l'analyse des phénomènes sous-jacents. En d'autres termes dans notre cas, les réseaux de neurones pouvaient fournir des numéros d'affiche complètement aberrants par rapport à la position et l'orientation de la tête d'un individu. Par exemple, pour un individu regardant l'affiche numéro 1, le réseau pouvait fournir comme réponse l'affiche numéro 6, 12 ou 18 qui sont complètement à l'opposé sur l'écran. De plus, du fait de l'utilisation des réseaux de neurones pour des données en temps réel (données fluctuant énormément), le taux de sortie souhaité baissait considérablement. Nous avons donc abandonné cette solution au profit de la suivante.

## A l'aide de moyennes mobiles et des écarts moyens mesurés

### *Moyennes Mobiles*

Dans un premier temps, nous avons mis en place un filtre sur les données fournies en temps réel par le Kinect V2. Ce filtre est basé sur le calcul en continu de moyennes pour chaque valeur de positions (userX, userY et userZ) et d'orientations (en lacet et en tangage) de la tête. Prenons comme exemple, la donnée de distance (userZ) entre la tête de l'individu et le Kinect V2. Considérons une moyenne mobile sur 10 valeurs. Le Kinect V2 a un taux de rafraîchissement de 30 images par seconde. Le principe est le suivant :

- (1) : les 10 premières données fournies par le Kinect V2 servent à constituer une liste de valeurs de la donnée de distance (userZ) ;
- (2) : nous calculons la moyenne de ces 10 valeurs ;
- (3) : nous utilisons alors la moyenne obtenue comme valeur de distance (userZ) pour réaliser les calculs nécessaires pour afficher la zone d'attention ;

- (4) : la valeur suivante envoyée par le Kinect V2 étant recueillie, nous retirons alors la 1<sup>ère</sup> valeur de la liste de valeurs (la plus ancienne) et ajoutons la nouvelle ;
- (5) : retour en (2) et ainsi de suite.

Nous obtenons ainsi un calcul de moyenne de la donnée de distance (userZ) en temps en réel, que nous qualifions de moyenne mobile. Nous avons mis en place cette méthode de filtrage par moyenne mobile sur les autres données nécessaires pour le calcul de la zone d'attention (cf. Figure 46). Cependant, cette méthode n'était toujours pas suffisante pour obtenir de manière précise la zone d'attention sur l'écran. Nous l'avons donc couplé avec une autre méthode basée sur la correction du calcul de la zone d'attention par les écarts moyens mesurés pour une lecture d'affiche dans l'expérience précédente.

#### *Procédure d'application des écarts moyens mesurés*

La deuxième méthode est basée sur l'analyse des données récoltées dans notre seconde expérience décrite dans le chapitre précédent.

Pour rappel, voici un schéma représentant les différentes données importantes :

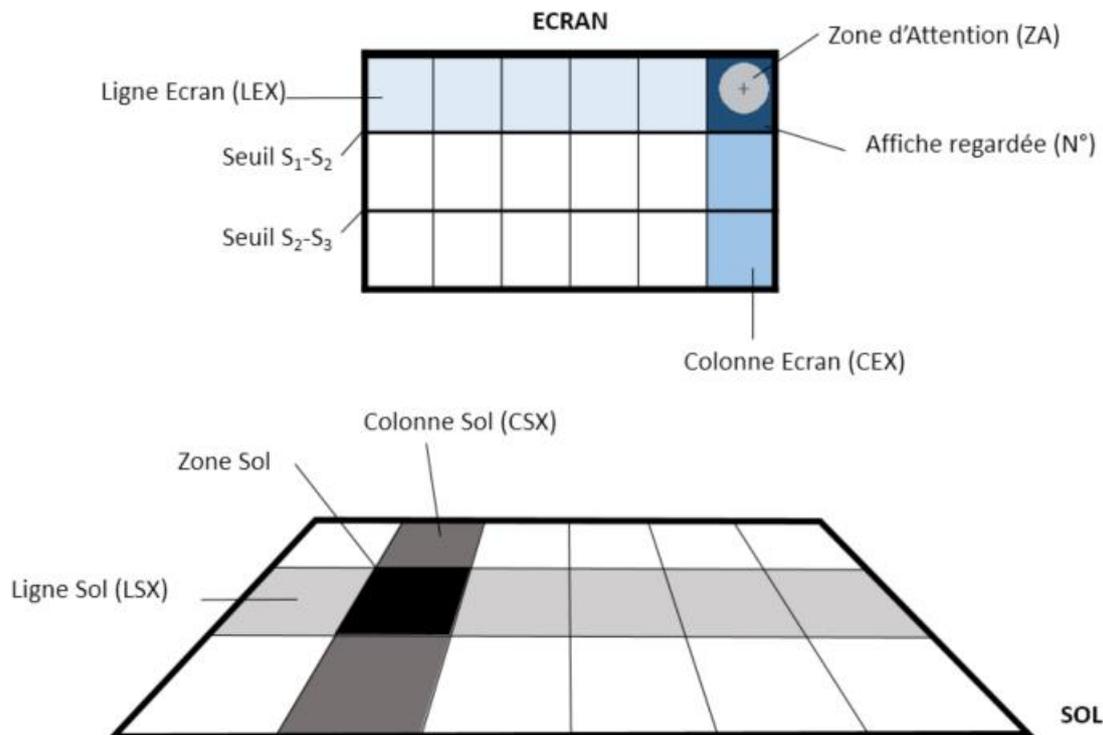


Figure 50 : Description de la configuration physique de l'expérience.

Nous avons donc répertorié ces données dans différents tableaux :

- (a) : dans un premier temps, nous avons classé dans un tableau les valeurs moyennes des angles de rotations en lacet (Yaw), exprimées en degrés, en fonction de chaque distance entre l'utilisateur et l'écran (LS1, LS2 et LS3) et en fonction des cas de consultation présentés dans l'expérience précédente (cf. Chapitre 4 : : *Principe de l'expérience*).

Si nous prenons un exemple de cas de consultation (+4) dans le *Tableau 29*, cette valeur indique que l'individu regarde à plus quatre colonnes écran vers la droite par rapport à sa position. A l'inverse, (-4) indique que l'individu regarde à plus quatre colonnes écran par rapport à sa position vers la gauche. Nous appliquons cet exemple sur la *Figure 50*, si l'individu est en Zone Sol n°8 (ZS8), soit en Ligne Sol n°2 (LS2) et Colonne Sol n°2 (CS2), les Colonnes Sol et les Colonnes Ecran (CE) étant alignées et si l'individu regarde à plus quatre colonnes vers la droite, cela signifie qu'il regarde en CE6.

*Tableau 29 : Moyennes de la rotation en lacet (Yaw), exprimées en degrés, en fonction de la distance et du cas de consultation.*

Cas de consultation	Moyenne Yaw - LS1	Moyenne Yaw - LS2	Moyenne Yaw - LS3
-5	-40	-35	-32
-4	-34	-28	-25
-3	-29	-23	-19
-2	-22	-15	-13
-1	-10	-7	-6
0	1	1	0
1	12	9	6
2	21	16	12
3	27	22	16
4	34	26	19
5	43	28	26

Ce tableau nous a permis d'exprimer trois fonctions polynomiales permettant de calculer en fonction de la distance de l'individu à l'écran et de la valeur de rotation en lacet de sa tête, son cas de consultation.

- (b) : dans un deuxième temps, nous avons construit un tableau indiquant la valeur moyenne en pixels de la frontière séparant la Ligne Ecran n°1 (LE1) de LE2 et la hauteur moyenne séparant LE2 de LE3 en fonction de la distance et du cas de consultation (*cf.* *Tableau 30 : Valeurs moyennes des seuils (S1-S2 et S2-S3) entre les Lignes Ecran (LE), exprimées en pixels, en fonction de la distance à l'écran et du cas de consultation. Tableau 30*).

Tableau 30 : Valeurs moyennes des seuils ( $S_1-S_2$  et  $S_2-S_3$ ) entre les Lignes Ecran (LE), exprimées en pixels, en fonction de la distance à l'écran et du cas de consultation.

Seuils Cas de consultation	LS1		LS2		LS3	
	$S_1-S_2$	$S_2-S_3$	$S_1-S_2$	$S_2-S_3$	$S_1-S_2$	$S_2-S_3$
-5	587	692	373	524	129	300
-4	581	683	395	537	181	323
-3	578	703	389	558	188	329
-2	578	723	400	560	205	375
-1	587	739	394	573	253	409
0	591	756	438	613	258	414
1	649	790	450	620	281	445
2	665	810	494	661	293	447
3	706	848	510	676	330	478
4	706	836	591	742	299	453
5	732	841	594	753	213	374

Ce tableau nous a permis d'exprimer six fonctions polynomiales permettant de calculer en fonction de la distance de l'individu à l'écran et de son cas de consultation, si l'individu regarde la LE1, la LE2 ou la LE3. En effet, en comparant la valeur FocusY du centre de la zone d'attention aux valeurs de seuil données par les différentes fonctions polynomiales en fonction de la distance et du cas de consultation, nous sommes en mesure de déterminer si l'individu regarde au-dessus ou en-dessous des seuils. Ainsi, nous en déduisons la Ligne Ecran regardée.

- (c) : dans un troisième temps, nous avons établi des fonctions polynomiales indiquant l'écart moyen mesuré en pixels entre la position en y du centre d'une affiche et la position en y du centre de la zone d'attention calculée sur l'écran, pour chaque cas de consultation en fonction de la distance à l'écran (LSX) et de la Ligne écran (LEX) regardée. Ces fonctions polynomiales permettent donc d'obtenir l'écart à corriger pour obtenir la coordonnée FocusY du centre de la zone d'attention la plus précise possible.
- (d) : enfin, dans un quatrième temps, nous avons créé un tableau de correspondance indiquant, en fonction de l'Affiche regardée depuis une Zone Sol, l'écart à corriger pour obtenir la coordonnée FocusX du centre de la zone d'attention la plus précise possible.

Ces différentes données agrégées nous permettent la mise en place d'une séquence d'opérations permettant de corriger la position de la zone d'attention sur l'écran. Cette

méthode, que nous nommerons *Procédure d'application des écarts moyens mesurés* se divise en 9 étapes :

- (1) : la première étape consiste, à partir des données de position de la tête fournies par le Kinect V2, transformées par l'*Etalonnage* et moyennées par la *méthode des Moyennes Mobiles*, à déterminer la distance entre l'individu et l'écran et plus précisément de déterminer dans quelle Zone Sol est l'individu. Il s'agit donc d'établir si l'utilisateur est dans la Ligne Sol la plus près de l'écran (LS1 : Zone Sol 1 à 6), dans la Ligne Sol intermédiaire (LS2 : Zone Sol 7 à 12) ou dans la Ligne Sol la plus éloignée (LS3 : Zone Sol 13 à 18). Puis, d'identifier dans quelle Colonne Sol (CSX) il se trouve. Le croisement de la Ligne Sol et de la Colonne Sol nous indiquant ainsi la Zone Sol dans laquelle se trouve l'utilisateur. Pour ce faire, nous utilisons la valeur *userZ* donnée par le Kinect V2 pour déterminer la Ligne Sol et le calcul de la coordonnée (*yNormalTête*) sur l'écran (cf. *Calcul de la zone d'attention sur l'écran*) pour obtenir la Colonne Sol.
- (2) : la deuxième étape consiste à calculer, à partir des fonctions polynomiales établies en (a), si l'individu regarde à plus X colonnes sur la droite ou la gauche. Les fonctions polynomiales établies en (a) requièrent en entrées la Ligne Sol et la valeur de rotation en lacet (*Yaw*) pour calculer le cas de consultation. Nous calculons alors le cas de consultation à l'aide de la Ligne au sol déterminée en (1) et à l'aide de la valeur de *Yaw* fournie par le Kinect V2.
- (3) : la troisième étape consiste à calculer la valeur de seuil à l'aide des fonctions polynomiales établies en (b). Les fonctions polynomiales établies en (b) requièrent en entrées la Ligne Sol et le cas de consultation. Nous obtenons alors la valeur du seuil à l'aide de la Ligne au sol déterminée en (1) et du cas de consultation déterminé en (2).
- (4) : la quatrième étape consiste à comparer la position en y (*FocusY*) de la zone d'attention calculée à l'écran, à la valeur de seuil obtenue en (3). Cette comparaison, nous permet d'établir dans quelle Ligne Ecran l'individu regarde.
- (5) : la cinquième étape consiste à calculer l'écart à corriger pour obtenir la coordonnée *FocusY* la plus précise possible. Ce calcul est réalisé à l'aide des fonctions polynomiales établies en (c). Les fonctions polynomiales établies en (c) requièrent en entrées la Ligne Sol déterminée en (1), le cas de consultation obtenu en (2) et la Ligne Ecran déterminée en (4).
- (6) : la sixième étape consiste en l'application de l'écart à corriger déterminé en (5) à la coordonnée *FocusY* de la zone d'attention. A cette étape, la hauteur sur l'écran de la zone attention est donc corrigée.
- (7) : la septième étape consiste à identifier l'affiche regardée par l'individu. Ce résultat est obtenu en croisant les informations obtenues en (1), en (2) et en (4). En effet, en sachant la Zone Sol dans laquelle se trouve l'individu, son cas de consultation et la Ligne Ecran qu'il regarde, nous obtenons aisément le numéro de l'Affiche regardée.
- (8) : la huitième étape consiste à déterminer l'écart à corriger pour obtenir la coordonnée *FocusX* la plus précise possible. Ce résultat est obtenu à l'aide du tableau de correspondance établi en (d). Pour établir la correspondance, la Zone Sol déterminée en (1) et le numéro d'affiche établi en (7) sont nécessaires. Nous obtenons ainsi l'écart à corriger pour obtenir la coordonnée *FocusX* la plus précise possible.
- (9) : la neuvième étape consiste en l'application de l'écart à corriger déterminé en (8) à la coordonnée *FocusX* de la zone d'attention. A cette étape, la latéralité sur l'écran de la zone attention est donc corrigée ainsi que la hauteur (6).

Voici un schéma récapitulatif de la *Procédure d'application des écarts moyens mesurés* :

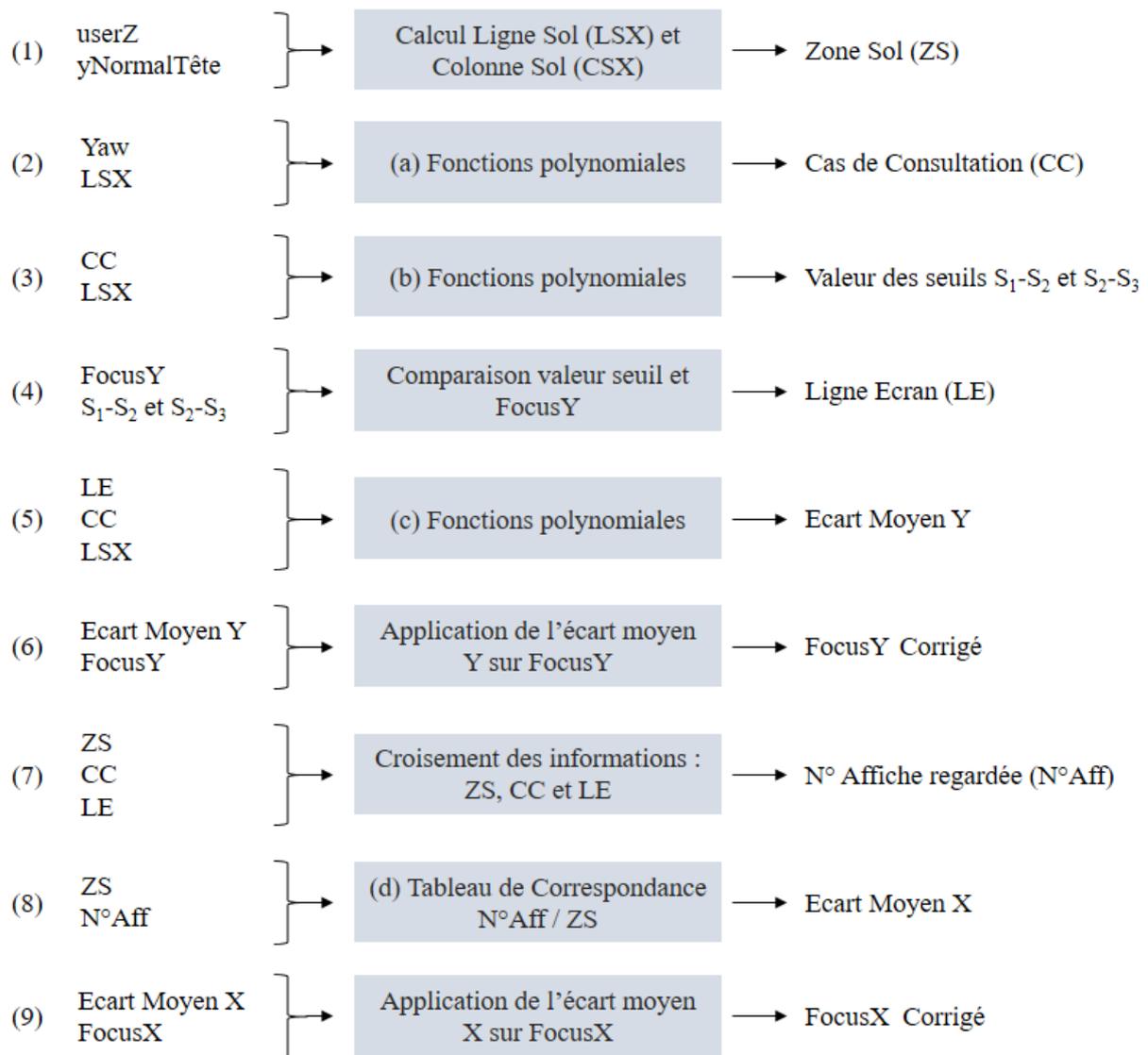


Figure 51 : Schéma de la Procédure d'application des écarts moyens mesurés.

L'addition de la *méthode des Moyennes Mobiles* et de *Procédure d'application des écarts moyens mesurés* est suffisante pour obtenir de manière opérationnelle la zone d'attention sur l'écran permettant ainsi une interaction avec l'écran en situation d'usage.

#### *En résumé*

A chaque image, les données sont étalonnées, puis moyennées à l'aide de la méthode des Moyennes Mobiles. A cette étape, nous calculons les coordonnées du centre de la zone d'attention sur l'écran sachant pertinemment qu'elles ne sont pas correctes. Puis nous appliquons la Procédure d'application des écarts moyens mesurés qui donne en sortie les coordonnées correctes du centre de la zone d'attention.

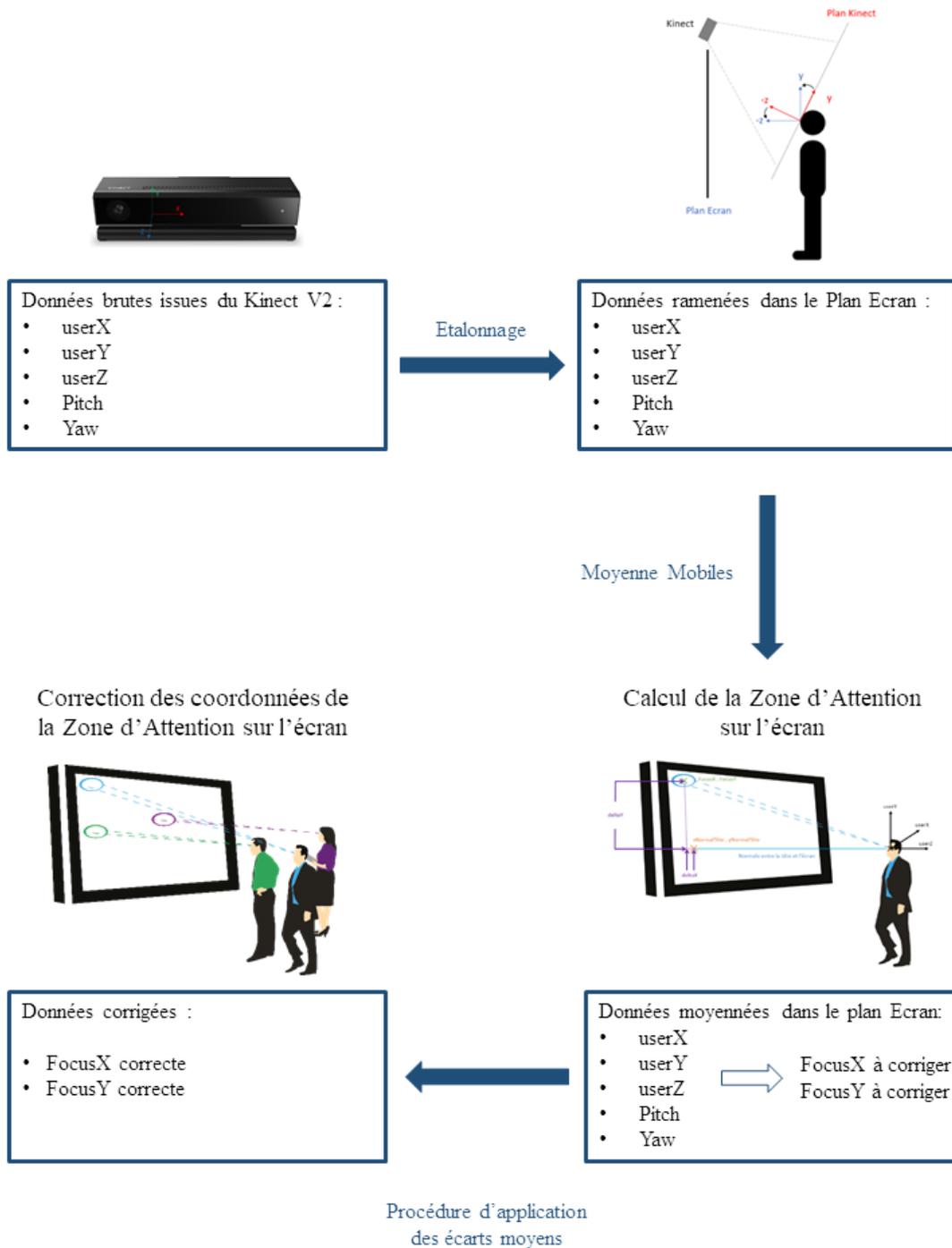


Figure 52 : Procédure complète du calcul de la Zone d'Attention permettant une interaction avec les écrans de grand format.

L'utilisation d'une moyenne mobile amène de la latence dans le système. Après plusieurs tests utilisateurs selon différentes valeurs de moyenne mobile, nous avons finalement appliqué une moyenne mobile sur 18 valeurs pour chaque variable. En effet, c'est le meilleur compromis stabilité *versus* inertie que nous avons obtenu pour proposer des interactions viables.

## Interactions

Une fois la Zone d'Attention correctement estimée sur l'écran, nous avons été en mesure de développer des interactions basées sur celle-ci. L'ensemble des interactions décrites dans cette partie respecte les bonnes pratiques soulignées dans la partie *Etat de l'art sur les écrans de grand format*.

### Paradigme d'Interface n°1 (PI-1): Affichage de l'image de la tête des utilisateurs dans leurs Zones Utilisateur respectives

Lorsque les utilisateurs sont détectés par le Kinect V2, nous affichons en temps réel, dans la ligne inférieure de l'écran, l'image de leurs têtes prise par le capteur. Cette image est placée au centre d'une zone appelée « Zone Utilisateur ». Le contour de la Zone Utilisateur est matérialisé par une couleur différente pour chaque utilisateur. Chaque Zone Utilisateur est attribuée de manière à correspondre à la zone la plus directement accessible, c'est à dire de plus grande proximité entre l'utilisateur et l'écran comme le montre la *Figure 53*. Dans le cas où deux utilisateurs sont alignés sur la profondeur (l'un derrière l'autre) par rapport à l'écran (même colonne), la Zone Utilisateur est alors attribuée à l'utilisateur le plus proche de l'écran.

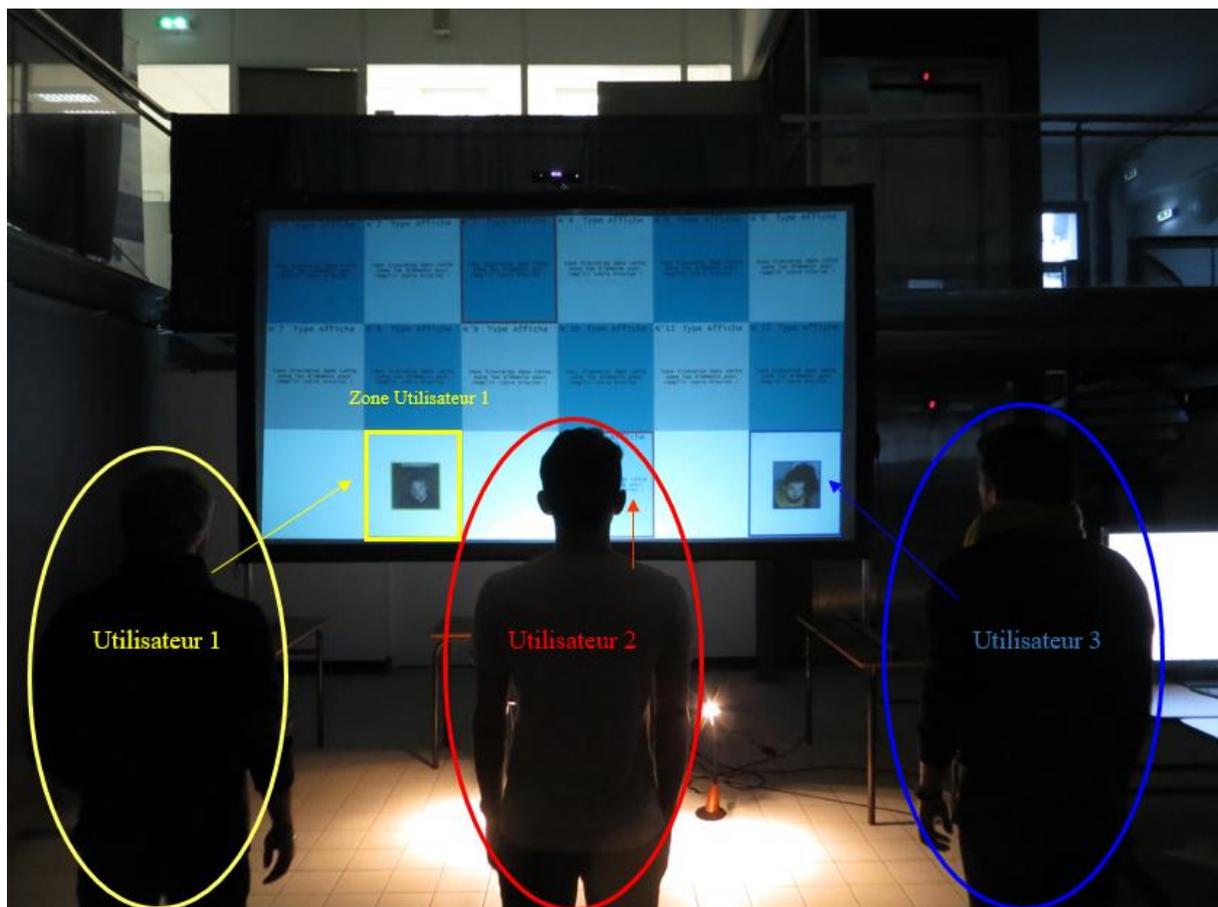
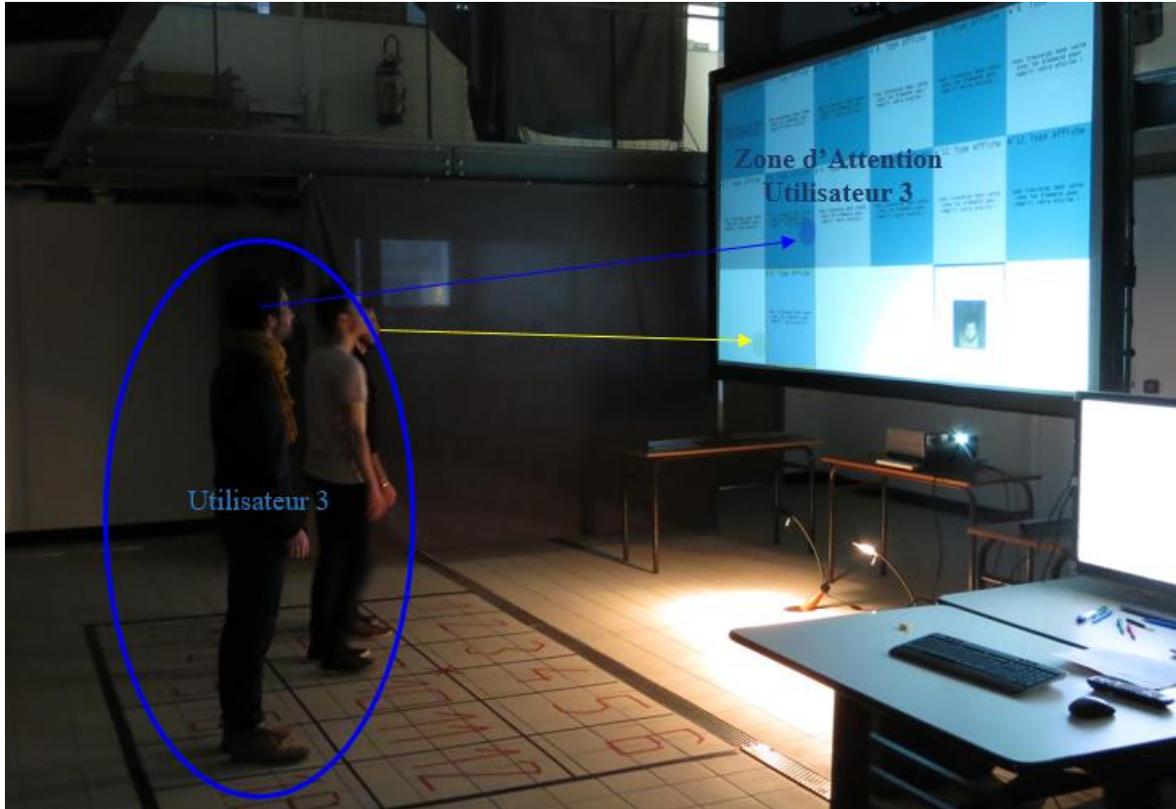


Figure 53 : Description des Zones Utilisateur à l'écran.

### Paradigme d'Interface n°2 (PI-2): Affichage des Zones d'Attention des utilisateurs détectés

La Zone d'Attention est affichée sous forme de cercle de 100 pixels de rayons. Elle est colorée de la même couleur que le contour de la Zone Utilisateur de l'utilisateur en question (PI-1). L'opacité de la couleur de remplissage a été choisie de manière à voir à travers (30%). La Zone d'Attention se déplace en fonction des mouvements de tête propre à chaque individu comme présenté en *Figure 54*.



*Figure 54 : Photo Zones d'Attention sur l'écran.*

Le choix d'utiliser la même couleur pour la bordure des Zones Utilisateurs et la Zone d'Attention permet de soutenir l'association des informations (*cf. Lignes directrices pour la conception d'affichage de grand format*).

### Cas Interactif n°0 (CI-0) : Intégralité de la Zone d'Attention comprise dans une affiche

Lorsque la Zone d'Attention est complètement comprise dans une affiche :

- son opacité diminue jusqu'à ce qu'elle disparaisse au bout de 3 secondes (*cf. Figure 55*). Cette diminution de l'opacité, jusqu'à disparition de la Zone d'Attention permet de ne pas être gêné durant la consultation de l'affiche ;
- le contour de l'affiche apparaît alors coloré de la même couleur que la Zone d'Attention et que le contour de la Zone Utilisateur ;
- dès que la Zone d'Attention n'est plus complètement comprise dans une affiche, elle réapparaît avec son opacité d'origine.

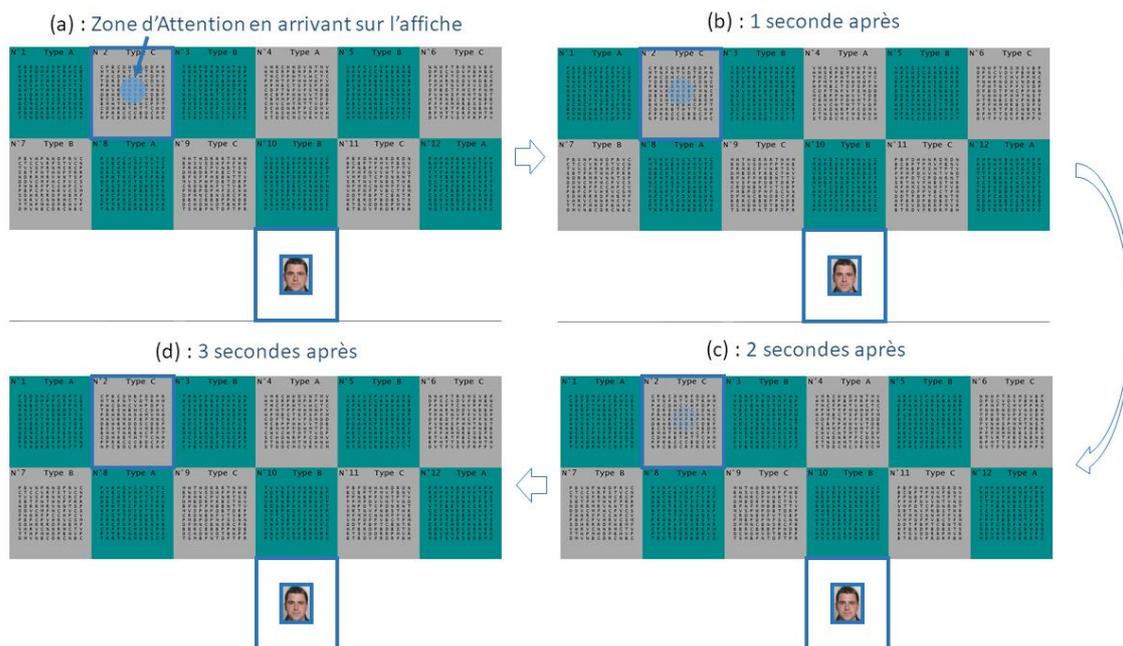


Figure 55 : Atténuation de l'opacité de la Zone d'attention. (a) 30% d'opacité, (b) 20% d'opacité, (c) 10% d'opacité et (d) 0% d'opacité.

Encore une fois, la coloration de la bordure de l'affiche visée par l'utilisateur de la même couleur que celle de sa Zone d'Attention et de sa Zone Utilisateur permet de soutenir la compréhension de son action sur l'interface (cf. *Lignes directrices pour la conception d'affichage de grand format*).

#### CI-0 et Zone utilisateur avec affiche : CI-0-0

Lorsque la Zone d'Attention d'un utilisateur est complètement comprise dans une affiche présente dans la Zone Utilisateur d'un autre utilisateur ou dans une des deux lignes supérieures de l'écran et que sa Zone Utilisateur contient déjà une affiche :

- Il ne se passe rien.

#### CI-0 et Zone utilisateur sans affiche : CI-0-1

Lorsque la Zone d'Attention d'un utilisateur est complètement comprise dans une affiche des deux lignes supérieures de l'écran et que sa Zone Utilisateur ne contient pas d'affiche :

- CI-0-1-1 : une animation de déplacement de l'affiche visée est déclenchée, simulant un déplacement de cette affiche, avec une opacité faible, vers la Zone Utilisateur de l'utilisateur concerné comme illustré en *Figure 56*. La Zone Utilisateur contient alors l'affiche visée par la Zone d'Attention de l'utilisateur mais toujours avec une opacité faible. L'affiche disparaît au bout de quatre secondes, si l'utilisateur ne la vise pas avec sa Zone d'Attention. L'animation permet à l'utilisateur d'identifier l'emplacement cible lié au déplacement de l'affiche (cf. *Lignes directrices pour la conception d'affichage de grand format*).

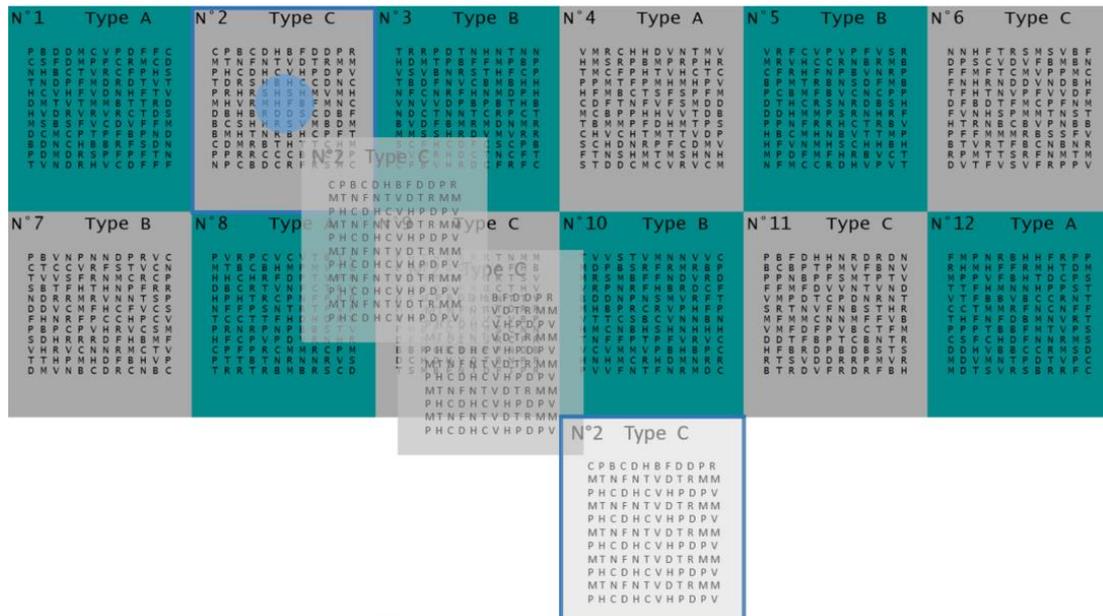


Figure 56 : CI-0-1-1 : Animation de déplacement d'une affiche avec la Zone d'Attention.

Lorsque la Zone d'Attention d'un utilisateur est complètement comprise dans une affiche présente dans la Zone Utilisateur d'un autre utilisateur et que sa Zone Utilisateur ne contient pas d'affiche :

- CI-0-1-2 : l'affiche visée apparaît instantanément, avec une opacité faible, dans la Zone Utilisateur de l'utilisateur concerné comme illustré en Figure 57. La Zone Utilisateur contient alors l'affiche visée par la Zone d'Attention de l'utilisateur mais toujours avec une opacité faible. L'affiche disparaît au bout de quatre secondes, si l'utilisateur ne la vise pas avec sa Zone d'Attention.

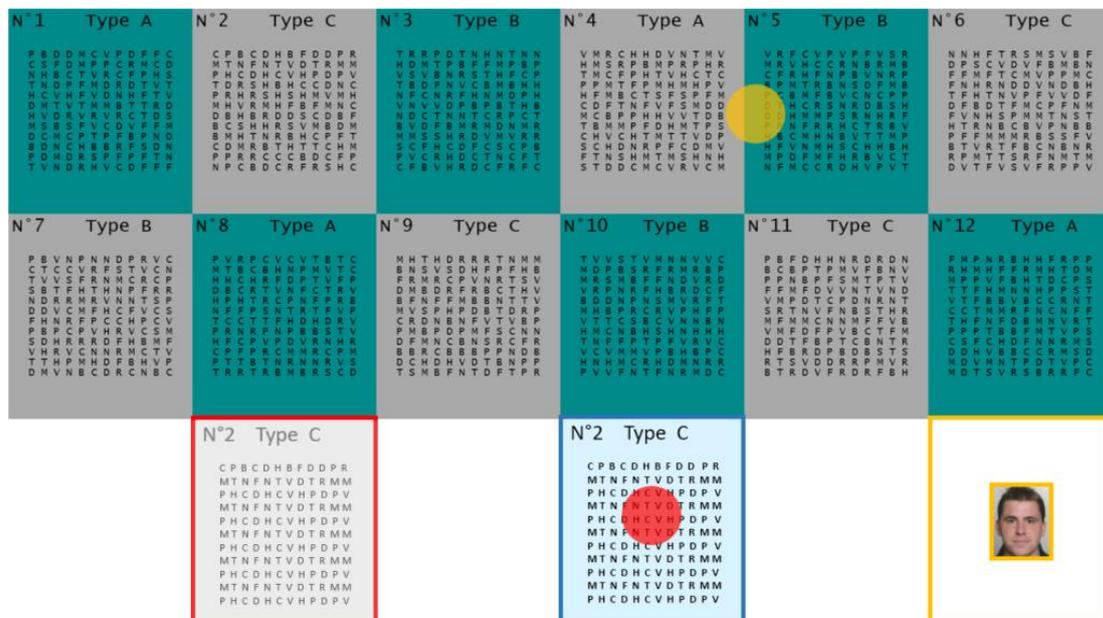


Figure 57 : CI-0-1-2 : Copie instantanée de l'affiche présente dans la Zone Utilisateur d'un autre, avec sa Zone d'Attention.

## CI-0-2-X : Cas interactifs se déroulant après réalisation de CI-0-1-X

Ces interactions ne peuvent se produire que si les utilisateurs ont réalisé les cas interactifs CI-0-1-1 ou CI-0-1-2 :

- CI-0-2-1 : si l'utilisateur déplace alors sa Zone d'Attention de manière à ce que celle-ci soit complètement comprise dans l'affiche ainsi copiée (CI-0-1-1 ou CI-0-1-2) dans sa Zone Utilisateur, l'affiche apparaît alors avec une opacité maximale comme présenté en Figure 58 ;

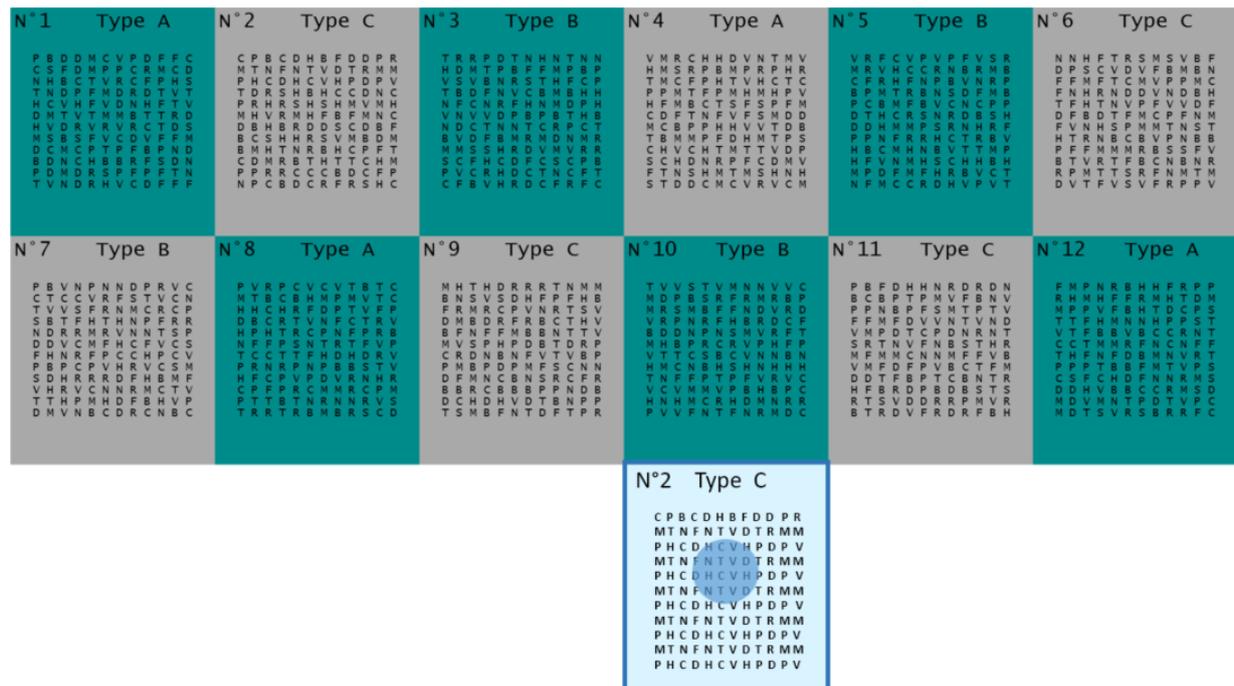


Figure 58 : CI-0-2-1 : Apparition de l'affiche avec dans la Zone Utilisateur à l'aide de la Zone d'Attention.

- CI-0-2-2 : si l'utilisateur déplace sa Zone d'attention au-delà de l'affiche copiée avec un contraste élevé dans sa Zone Utilisateur (CI-0-2-1) mais revient dessus avant sa disparition (4 secondes), il réactive le délai de 4 secondes avant sa disparition et ainsi de suite (CI-0-2-2). Si la Zone d'Attention de l'utilisateur n'est plus comprise dans l'affiche copiée avec un contraste élevé dans sa Zone Utilisateur (CI-0-2-1) pendant plus de 4 secondes, l'affiche disparaît de sa Zone Utilisateur (retour en PI-1 et PI-2).

## Principe de l'expérience

L'expérience repose sur la comparaison de deux modes d'affichage portant sur des informations identiques :

- un affichage dit statique (Modalité S) comportant les paradigmes d'affichage PI-1 et PI-2 et le cas interactif CI-0 ;

- un affichage dit dynamique (Modalité D) proposant les mêmes propriétés que l’affichage statique agrémenté de tous les cas interactifs (CI-0-X) présentés dans la partie *Interactions*.

L’affichage est découpé en 3 lignes et 6 colonnes d’égales dimensions comme illustré dans la *Figure 59*.

N°1 Type A P B B D D M C V P F F C C C S F C D M P F C F H S N B C T P Y C C F T V T T N D P P F D R R T T T K C H V R D N H T T R D D M T V T M M B T T R D H D R V R V R C T S M S S F V C D V F F M B D C C P F F F P N P D D C D R S P F F T N T V N D R V C D F F	N°2 Type C C P B C D H B T D D P R M F N T V H P H C M P R C D S H V H C C D N C T R R S H B H C C V H H H R V R H H F B F H N C D H B B D S C D B F B C S H H R S V B D H B H T R B C C F T C P R R R C C C D F F N P C B C C R F S C	N°3 Type B T R R P D T N H N T N N H D N T B F F H N P F P V L V B R B C T H B B H T D D F R V C C H B H H N C D F T F F H N D H V N V D P B P B T H B N D C T N T C F F C T B V D F B H A N D N R M S S R D H V R R S C F R C D F C S C P P C B V K D F R C	N°4 Type A V M R C H H D V N T H V T H C R P H H P R P C P F N B C T S F S M D C D F T N F V F S M D M C B P H H V T D B T B M H P F D H M T S C H V C H T T V P S C H D N R P F C D V V F T N S H M T S H H H S T D D C H C V R V C H	N°5 Type B V R F C V P V F F V S R C R V H C C R N B R B P E P H T F B S D F P B D T H C S N R D D S H D D H M F R N R F P P N F R R H C T R B V H B C N N B T R P H F V N H S C H H B H H P D C C R A H V V T	N°6 Type C N N S C T R V S H S B F D P F H T D V V F B N C F N H N D V F P D M C T F H T N V F V D M D F B D T F C P N M O T H C S N R D D S H F V N S F H T S T H T R N B C V P N B B P F F H M B S S F V S T V R T S F C N R H R P M T S F C N S T D V T V S F R P V
N°7 Type B P B V N P N D P R V C N C T V S C C F R N C R F S B T F H T H N F R S P N D R R V V H C T F P D D C H F C H F C S S F H R R P F C H F C V P B B C P P H R V C S M S D R R C N N R F H S P V T H P H H D R C N B D H N B C D R C N B	N°8 Type A P V R P C V C V T B T C M B C B H H D P T V F H C H F V N F T R V D E C E T C P F P R B N F P P S N T R T F V P T C T F P H D R V P R N R P P B B S T V H C F P P D N N R V S C P T T B T N R N R V S T A T B M R S C	N°9 Type C M H T H D R R R T N N H B S V R C P V R C T V D B B D R F R C T H V B N D F H S N T V M V S P H P D S T D R P C D N E N F T V P P H B P D P F S C N N D H N C B N F S C R B S R C B B B P N D B D C H B D V D T B N P P T S B F N T D F R	N°10 Type B T V V S T V M N N V V C H D P B S F F H R R D V B S F F H S R D D F B D D P N S V R D F T H B B P C R V P H P P U T T C S B C V N B N H M C N B S H S N H H V H F D P P V B C T F M V D T F P C S N R D D T B D P D B S T S H N H C H C H M N R R T S V F R D R F B	N°11 Type C P B F D H H N R D R N V F C B N P F S V T V N F F H P D V V N T V N V H P D T C D N R N T S R T N V F N S T H R M H F M C N M F B V H F D P P V B C T F M D D T B D P D B S T S H N H C H C H M N R R T S V F R D R F B	N°12 Type A F H P N R B H H F P P R H P F R M H D P S T T F H M N N H P S T T F F S V C R K T C C T M M R N C C P F T H F F B H H P P P P P P T B S H T V P S C S R C H N N R S D D V B S C R T P C H D V N T D T P C H D T V R B R F C

Figure 59 : Composition de l’affichage

Les deux lignes supérieures contiennent chacune six affiches composées d’un numéro (en haut à gauche), d’un titre et d’un tableau de caractères en partie centrale. L’expérience consiste, pour 3 utilisateurs en simultanée, à remplir des missions individualisées. Pour chaque mission individualisée, nous demandons aux utilisateurs de compter, dans des affiches désignées selon leur titre, un caractère spécifique présent dans les tableaux de caractères. Par exemple, nous pouvons demander de compter le nombre de caractère « T » dans les affiches dont le titre est « Type A ».

Les utilisateurs réalisent donc des tâches de consultation selon les deux modes d’affichage. En comparant la performance de consultation et le ressenti des utilisateurs dans les deux modes d’interactions, nous souhaitons démontrer l’intérêt de rendre les affichages interactifs à l’aide d’interaction basée sur les mouvements de tête afin d’améliorer l’expérience utilisateurs.

## Le SUS et l’AttrakDiff

Afin de comparer les deux modes d’affichage, nous avons choisi d’utiliser deux questionnaires post-expérience de recueil du ressenti des utilisateurs, servant à l’évaluation d’interface :

- le *System Usability Scale (SUS)* (Brooke 1996) ;
- l’*AttrakDiff* (Hassenzahl, Burmester et al. 2003).

L’avantage de ces deux questionnaires est qu’ils ne sont pas très long à remplir (5 minutes chacun environ).

Nous avons utilisé les versions de ces questionnaires en langue française (cf. *Annexe IV : Questionnaire System Usability Scale (SUS)* et *Annexe V : Questionnaire AttrakDiff*). L’usage

du SUS est très répandu dans la communauté pour évaluer les interfaces. Cependant, il est important de savoir que la version française du SUS n'a jamais été validée. Contrairement, à la version en langue française de l'*AttrakDiff* (Lallemant, Koenig et al. 2015).

Le *SUS* évalue exclusivement la composante pragmatique des interfaces. L'*AttrakDiff*, quant à lui, permet d'évaluer la composante pragmatique d'une interface mais également de mesurer la composante hédonique de celle-ci.

Ainsi, en utilisant conjointement le *SUS* et l'*AttrakDiff* nous pouvons comparer la composante pragmatique des deux questionnaires et également mesurer la composante hédonique apportée par l'*AttrakDiff*.

## Hypothèses

L'hypothèse principale, que nous explorons dans ce chapitre, est la suivante :

« Les interactions multi-utilisateur, non appareillées et sans calibrage basées sur le *head tracking*, améliorent la consultation des écrans de grand format et sont bien accueillies par les utilisateurs. »

De cette hypothèse générale, nous tirons les hypothèses suivantes :

- les 3 premières hypothèses sont basées sur l'amélioration de la consultation des écrans de grand format :
  - (H1) les utilisateurs commettent moins d'erreurs lorsqu'ils sont exposés à la modalité d'affichage D ;
  - (H2) en moyenne les utilisateurs donnent plus de bonnes réponses en modalité D ;
  - (H3) en moyenne les utilisateurs sont plus rapides avec la modalité S ;
- les six dernières hypothèses sont basées sur l'accueil réservé à ces nouvelles interactions par les utilisateurs :
  - (H4) les utilisateurs préfèrent utiliser la modalité d'affichage D ;
  - (H5) les utilisateurs pensent avoir été plus rapides avec la modalité S ;
  - (H6) les utilisateurs estiment être plus efficaces avec la modalité S ;
  - (H7) les utilisateurs utiliseraient la modalité D si elle était présente dans les lieux publics ;
  - (H8) les utilisateurs donnent de meilleures appréciations à la modalité D avec le questionnaire *AttrakDiff* ;
  - (H9) les utilisateurs donnent de meilleures appréciations à la modalité D avec le questionnaire *SUS*.



# Protocole expérimental

## Matériel

### *Configuration physique*

La configuration physique est la même que dans le chapitre précédent (*cf. Chapitre 4 : Matériel*).

## Méthode

### *Détails computationnels*

Les détails computationnels sont les mêmes que dans le chapitre précédent (*cf. Chapitre 4 : Matériel*). Nous avons cependant corrigé les données brutes fournies par le Kinect V2 à l'aide de la procédure complète (*cf. En résumé*) afin d'afficher à l'écran, en temps réel, la zone d'attention de 3 utilisateurs simultanément.

### *Milieu d'implantation*

Comme vu précédemment, les deux lignes supérieures de l'écran sont composées chacune de 6 affiches d'égales dimensions. La taille de la police (titres, numéros d'affiche et tableaux de caractères) correspond à 10% de la largeur de l'affiche. Sur la ligne écran supérieure, les affiches impaires ont un contraste (couleur de fond par rapport au texte) peu élevé, tandis que les affiches paires ont un contraste élevé et inversement pour la ligne écran centrale comme présenté sur la *Figure 59*. La ligne inférieure de l'écran est composée de 6 zones alignées et de mêmes dimensions que les affiches des deux autres lignes. Ces zones sont appelées « Zone Utilisateur ».

### *Description détaillée des deux modalités d'affichage*

L'expérience consiste à comparer deux styles d'affichage :

- un affichage dynamique (Modalité D, *cf. Tableau 31*) ;
- un affichage statique (Modalité S, *cf. Tableau 32*).

Comme vu précédemment, les affiches contiennent un titre. Les titres sont Type A, Type B ou Type C. Le *Tableau 31* et le *Tableau 32* décrivent leur répartition selon les deux modalités. Cette répartition des titres d'affiche selon chaque modalité a été choisie afin qu'aucune affiche avec le même titre ne soit côte à côte. Nous avons également réalisé une rotation circulaire des titres entre chaque modalité pour éviter un effet d'apprentissage. En effet, les groupes de 3 utilisateurs réalisent l'expérience d'abord avec une modalité d'affiche puis avec l'autre. L'ordre de présentation des modalités d'affichage a été contrebalancé pour éviter un effet d'ordre des présentations.

Chaque affiche possède en partie centrale un tableau de douze fois douze caractères. Chaque tableau de caractères est composé d'une répartition aléatoire des caractères parmi trois familles de caractère :

- famille 1 : S, N, R, T ;
- famille 2 : B, V, H, F ;
- famille 3 : D, C, M, P.

Tableau 31 : Répartition des titres d'affiche en modalité D.

B	A	C	B	C	A
C	B	A	C	A	B

Tableau 32 : Répartition des titres d'affiche en modalité S.

A	C	B	A	B	C
B	A	C	B	C	A

Les familles de caractères ont été constituées selon une étude sur la fréquence des caractères réalisée sur le corpus Wikipédia Français de 2008. Les tableaux composant chaque affiche dans une même modalité sont tous différents. D'une modalité à l'autre les tableaux de chaque affiche ont subi une permutation circulaire des caractères par famille. Ainsi le tableau de l'affiche numéro 1 en modalité D et le même que celui présent en modalité S à l'exception que les caractères ont été inversés par famille. Voici un exemple simple sur un tableau de quatre caractères :

- Modalité D : S N  
R T
- Modalité S : T S  
N R

Ces rotations circulaires ont pour but de conserver la difficulté de consultation d'une affiche d'une modalité à l'autre et permettent d'éviter un effet d'apprentissage.

Pour la modalité S, les paradigmes d'interface PI-1 et PI-2 sont présents et seules les interactions CI-0 sont accessibles (cf. *Interactions*). Pour la modalité D, toutes les interactions présentées dans la partie *Interactions* sont disponibles.

Par rapport à la modalité S, la modalité D offre la possibilité de dupliquer et déplacer des affiches dans sa Zone Utilisateur afin de les consulter dans des conditions de contraste et de position optimales.

Les interactions supplémentaires présentent seulement en modalité D conservent au maximum l'intégrité de l'affichage. C'est-à-dire qu'elles ne modifient que très peu l'affichage existant. En effet, les informations présentes en modalité S sont aussi présentes en modalité D. La modalité D permet seulement d'afficher des copies d'informations déjà présentes.

### *Déroulement de l'expérience*

Nous exposons donc chaque groupe de trois utilisateurs aux deux modalités d'affichage. Les groupes impairs ont démarré l'expérience avec la modalité D et inversement pour les groupes pairs.

L'expérience commence par la signature des formulaires de consentement libres, éclairés et exprès (cf. *Annexe VI : Formulaire de Consentement Libre, Eclairé et Exprès (deuxième et troisième expérience)*), puis nous récoltons des données personnelles (âge, sexe, œil directeur, etc.).

Ensuite, nous expliquons les consignes pour l'expérience. Chaque utilisateur apprend qu'il commence l'expérience dos à l'écran, face à une table placée sur l'arrière de la zone

d'interaction et sur laquelle sont disposées trois feuilles faces cachées (une par utilisateurs) contenant quatre missions à remplir par modalité (cf. *Annexe VII : Fiche mission*). Pour remplir leur mission les utilisateurs doivent se déplacer dans la zone d'interaction et consulter les informations présentes à l'écran. Ils sont chronométrés pour chaque mission. Les trois utilisateurs réalisent leurs missions en même temps. Nous leur présentons les missions à remplir les unes après les autres. Dès qu'un utilisateur a la réponse à sa mission, il doit venir noter la réponse sur sa fiche de mission. Il doit alors attendre de dos à l'écran, face à celle-ci que les deux autres aient terminé leur mission pour en démarrer une autre tous ensemble. Ils ont pour consignes de réaliser leur mission le plus rapidement possible et de veiller à faire le moins d'erreurs possibles. Ils n'ont pas le droit de communiquer entre eux de quelque manière que ce soit. Les trois premières missions consistent à compter un caractère sur les affiches désignées par leur titre. Par exemple, la mission numéro une de l'utilisateur numéro un en modalité D est : « Combien il y a de caractères « V » au total dans l'ensemble des affiches dont le titre est Type A ? ». Pour les trois premières missions, les utilisateurs doivent donc consulter quatre affiches à chaque fois pour obtenir le bon résultat. La quatrième mission de chaque modalité consiste à compter un caractère sur toutes les affiches présentes à l'écran. Le tableau contient l'ensemble des éléments concernant les missions.

	Modalité D								
	Utilisateur 1			Utilisateur 2			Utilisateur 3		
	Caractère à compter	Réponse	Titre affiche	Caractère à compter	Réponse	Titre affiche	Caractère à compter	Réponse	Titre affiche
Mission 1	V	39	A	S	34	B	R	36	C
Mission 2	R	36	C	P	36	C	F	34	B
Mission 3	S	34	B	V	39	A	D	39	A
Mission 4	T	82	Toutes	T	82	Toutes	T	82	Toutes

	Modalité S								
	Utilisateur 1			Utilisateur 2			Utilisateur 3		
	Caractère à compter	Réponse	Titre affiche	Caractère à compter	Réponse	Titre affiche	Caractère à compter	Réponse	Titre affiche
Mission 1	H	39	C	N	34	A	T	36	B
Mission 2	T	36	B	D	36	B	B	34	A
Mission 3	N	34	A	H	39	C	C	39	C
Mission 4	S	82	Toutes	S	82	Toutes	S	82	Toutes

Figure 60 : Tableaux récapitulatifs des missions pour chaque utilisateur dans chacune des modalités.

Nous remarquons dans la *Figure 60*, que pour la première mission, les utilisateurs n'ont aucune affiche en commun tandis que pour la mission numéro deux, l'utilisateur numéro un et deux ont les mêmes affiches à consulter mais pas le même caractère à compter. Le but étant de créer l'éventualité que deux utilisateurs aient la même affiche à consulter au même moment. Ainsi, les utilisateurs pourront avoir la possibilité de dupliquer une affiche présente dans la Zone Utilisateur de leur voisin. Pour la mission trois, c'est les utilisateurs deux et trois qui sont concernés par les mêmes affiches. Grâce à la permutation circulaire réalisée pour chaque famille de caractères, les utilisateurs cherchent un caractère différent dans les deux

modalités mais situé au même endroit dans les tableaux et dont la répartition et le nombre sont identiques. Ce qui permet de veiller à ce que les tâches réalisées dans une modalité ou l'autre soient comparables.

Ensuite, une démonstration des deux modalités d'affichage est réalisée. Pour la modalité D, chaque utilisateur est convié, séparément, pour essayer chaque interaction proposée.

L'expérience débute avec la première modalité. Puis à la fin de celle-ci, nous invitons le groupe à remplir deux questionnaires : L'*AttrakDiff* et le *SUS*. La consigne donnée pour remplir les questionnaires est de ne pas penser à la tâche (comptage de caractère) mais à la manière d'interagir avec l'affichage. Une fois les questionnaires terminés, nous exposons le groupe à la deuxième modalité. Puis, nous présentons à nouveau les deux mêmes questionnaires. Enfin, un dernier questionnaire de comparaison des deux modalités est proposé. Ce dernier questionnaire de comparaison est constitué de quatre questions réparties sur une échelle de Likert en cinq points et de deux questions ouvertes (cf. *Annexe VIII : Questionnaire Global*).

### *Enregistrement des données*

Durant l'expérience, nous enregistrons en continu, les coordonnées (x, y et z), dans le Plan Ecran et dans le Plan Kinect, des nœuds « *Head* » des 3 utilisateurs. Nous récupérons également les coordonnées FocusX et FocusY des Zones d'Attention des 3 utilisateurs ainsi que les angles de rotations de leurs têtes. De plus, nous notons le temps passé par chaque utilisateur à réaliser chaque mission. Nous récupérons les réponses de chacun sur leur fiche de mission. Enfin, grâce aux différents questionnaires, nous mesurons le ressenti et l'expérience utilisateur.

## Participants

Trente-cinq groupes de trois personnes ont participé à l'expérience. Les groupes impairs ont démarré l'expérience avec la modalité D soit dix-huit groupes et les dix-sept groupes impairs ont démarré avec la modalité S. Dans le *Tableau 33*, nous pouvons apprécier la répartition des utilisateurs dans chaque groupe selon leur âge, leur sexe, leur taille, s'ils sont droitiers ou gauchers de la main et selon leur œil directeur.

*Tableau 33 : Description de la population et de la répartition des groupes.*

	Age	Sexe	Taille	Main dominante	Œil directeur
<b>Population</b>	28 ± 11 ans	53H & 52F	1.72 ± 0.08 m	90D & 15G	75D & 30G
<b>Groupes Pairs</b>	25 ± 8 ans	27H & 27F	1.72 ± 0.08 m	46D & 8G	41D & 13G
<b>Groupes Impairs</b>	31 ± 13 ans	26H & 25F	1.72 ± 0.08 m	44D & 7G	34D & 17G

# Résultats

## Analyse des données

Les données ont été analysées avec le logiciel XLSTAT version 18.01.35929, logiciel développé par Addinsoft et s'intégrant dans Microsoft Excel. Nous prendrons comme seuil de significativité  $\alpha = 0,05$  pour l'étude statistique.

Dans un premier temps, nous vérifions l'absence d'un effet d'ordre des modalités sur chaque variable mesurée. Puis, nous présentons dans un second temps, les résultats de comparaison des modalités en termes de performance effective. Nous poursuivons l'analyse avec les résultats comparatifs des deux modalités concernant les aspects de ressenti des utilisateurs. Enfin, nous présentons la comparaison de l'évaluation de la qualité pragmatique des interfaces selon les questionnaires *SUS* et *AttrakDiff*.

## Comparaisons des modalités en termes de performance

### Comparaison du nombre d'erreurs commises (NbErr)

#### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Les données ne suivent pas une distribution normale. Nous avons donc effectué un test non paramétrique de Mann-Whitney. Le *Tableau 34* présente la description des données étudiées :

*Tableau 34 : Statistiques descriptives des données étudiées pour comparer le nombre d'erreurs commises en fonction de l'ordre de passation de l'expérience.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbErr   Gr Impairs	108	2,000	65,000	18,417	13,792
NbErr   Gr Pairs	102	3,000	72,000	18,961	13,020

Aucune différence significative n'est révélée avec une  $p$ -value = 0,456. Nous poursuivons donc l'analyse avec la comparaison des modalités quel que soit l'ordre de passation de l'expérience pour la variable NbErr.

#### *Comparaison Modalité S versus Modalité D : nombre d'erreurs commises*

Les données ne suivent pas une distribution normale. Nous avons donc effectué deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons appariés. Le *Tableau 35* présente la description des données étudiées :

*Tableau 35 : Statistiques descriptives des données étudiées pour comparer le nombre d'erreurs commises selon la modalité.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbErr   Mod D	105	2,000	72,000	16,181	12,870
NbErr   Mod S	105	4,000	78,000	21,181	13,499

Nous obtenons une différence significative,  $p$ -value  $< 0,0001$  pour les deux tests, entre la modalité D et la modalité S concernant le nombre d'erreurs commises. Les utilisateurs commettent significativement moins d'erreurs de comptage lorsqu'ils réalisent la tâche en Modalité D.

## Comparaison du nombre de bonnes réponses données (NbBonRep)

### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Les données ne suivent pas une distribution normale. Nous avons donc effectué un test non paramétrique de Mann-Whitney. Le *Tableau 36* présente la description des données étudiées :

*Tableau 36 : Statistiques descriptives des données étudiées pour comparer le nombre de bonnes réponses en fonction de l'ordre de passation de l'expérience.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>NbBonRep   Gr Pairs</b>	102	0,000	3,000	0,529	0,671
<b>NbBonRep   Gr Impairs</b>	108	0,000	3,000	0,611	0,874

Aucune différence significative n'est révélée pour une  $p$ -value = 0,503. Nous poursuivons donc l'analyse avec la comparaison des modalités quel que soit l'ordre de passation de l'expérience pour la variable NbBonRep.

### *Comparaison Modalité S versus Modalité D : nombre de bonnes réponses données*

Les données ne suivent pas une distribution normale. Nous avons donc effectué deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons appariés.

*Tableau 37 : Statistiques descriptives des données étudiées pour comparer le nombre de bonnes réponses données selon la modalité.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
<b>NbBonRep   Mod D</b>	105	0,000	3,000	0,600	0,804
<b>NbBonRep   Mod S</b>	105	0,000	3,000	0,543	0,760

Nous n'obtenons pas de différence significative,  $p$ -value = 0,301 pour le test du signe et  $p$ -value = 0,257 pour le test de Wilcoxon signé, entre la modalité D et la modalité S concernant le nombre de bonnes réponses données.

## Comparaison du temps écoulé pour réaliser les missions dans chaque modalité (TmpMiss)

### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Les données ne suivent pas une distribution normale. Nous avons donc effectué un test non paramétrique de Mann-Whitney. Le *Tableau 38* présente la description des données étudiées :

Tableau 38 : Statistiques descriptives des données étudiées pour comparer le temps écoulé (en minutes) pour réaliser les missions dans chaque modalité en fonction de l'ordre de passation de l'expérience.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbTmpMiss   Gr Pairs	102	8,183	27,600	13,023	3,410
NbTmpMiss   Gr Impairs	108	6,217	19,983	12,154	3,297

Nous obtenons une différence significative ( $p$ -value = 0,044). Les groupes exposés à l'ordre de passation des modalités S, puis D, sont significativement plus longs pour réaliser leur mission que ceux exposés à l'ordre de passation D puis S. Nous avons donc poursuivi l'analyse en comparant les modalités pour chaque ordre de passation.

*Comparaison Modalité S versus Modalité D des groupes pairs : temps écoulé pour réaliser les missions dans chaque modalité*

Les données suivent une distribution normale. Cependant, les tests de comparaison des variances (test F de Fisher et test de Levene) indiquent des différences significatives de variances entre les deux échantillons,  $p$ -value = 0,001 pour le test F et  $p$ -value = 0,002 pour le test de Levene. Nous avons donc effectué deux tests non paramétriques (test du signe et de test de Wilcoxon signé) de comparaison pour échantillons appariés. Le Tableau 39 présente la description des données étudiées :

Tableau 39 : Statistiques descriptives des données étudiées pour comparer le temps écoulé (en minutes) pour réaliser les missions dans chaque modalité en fonction des modalités pour l'ordre de passation S-D.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbTmpMiss   Mod S	51	8,183	17,167	11,559	1,998
NbTmpMiss   Mod D	51	8,433	27,600	14,486	3,889

Nous obtenons une différence significative,  $p$ -value < 0,0001 pour le test du signe et  $p$ -value < 0,0001 pour le test de Wilcoxon signé, entre la modalité D et la modalité S concernant le temps écoulé pour réaliser les missions dans chaque modalité pour les groupes exposés à l'ordre de passation S-D. Les groupes pairs sont significativement plus longs pour réaliser leur mission en Modalité D.

*Comparaison Modalité S versus Modalité D des groupes impairs : temps écoulé pour réaliser les missions dans chaque modalité*

Les données ne suivent pas une distribution normale. Nous avons donc effectué deux tests non paramétriques (test du signe et test de Wilcoxon signé) de comparaison pour échantillons appariés.

Tableau 40 : Statistiques descriptives des données étudiées pour comparer le temps écoulé (en minutes) pour réaliser les missions dans chaque modalité en fonction des modalités pour l'ordre de passation D-S.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
NbTmpMiss   Mod D	54	8,300	19,983	14,444	2,833
NbTmpMiss   Mod S	54	6,217	13,867	9,865	1,799

Nous obtenons une différence significative, p-value < 0,0001 pour le test du signe et p-value < 0,0001 pour le test de Wilcoxon signé, entre la modalité D et la modalité S concernant le temps écoulé pour réaliser les missions dans chaque modalité pour les groupes exposés à l'ordre de passation D-S. Les groupes impairs sont significativement plus longs pour réaliser leur mission en Modalité D.

## Comparaisons des modalités à l'aide du ressenti des utilisateurs

Pour réaliser ces comparaisons, nous nous basons sur les données extraites du questionnaire de fin d'expérience (cf. *Annexe VIII : Questionnaire Global*) et des réponses aux questionnaires *AttrakDiff* et *SUS* (cf. *Annexe IV : Questionnaire System Usability Scale (SUS)* et *Annexe V : Questionnaire AttrakDiff*). Pour le questionnaire de fin d'expérience, nous étudions donc les réponses données sur des échelles de Likert en 5 points (valeurs attribuées pour chaque point de l'échelle : -2, -1, 0, 1, 2).

### Comparaisons de la préférence d'utilisation d'une modalité par rapport à l'autre

#### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Nous avons utilisé un tableau de contingence dans le but de vérifier que l'ordre de passation n'influe pas sur les préférences des utilisateurs pour une modalité. Le *Tableau 41* présente la proportion des notes attribuées en fonction de l'ordre de passation (groupes Pairs ou Impairs). Les points -1 et -2 de l'échelle indiquent une préférence pour la modalité D, les points 1 et 2 pour la modalité S et le point 0 aucune préférence, symbolisée par « N ».

*Tableau 41 : Statistiques descriptives de la répartition de la préférence des utilisateurs pour chaque modalité.*

Modalités	Notes sur l'échelle de Likert	Notes données par Groupe Pairs	Notes données par Groupe Impairs	Total des notes
D	-2	16	23	39
	-1	18	16	34
N	0	4	1	5
S	1	5	10	15
	2	8	4	12

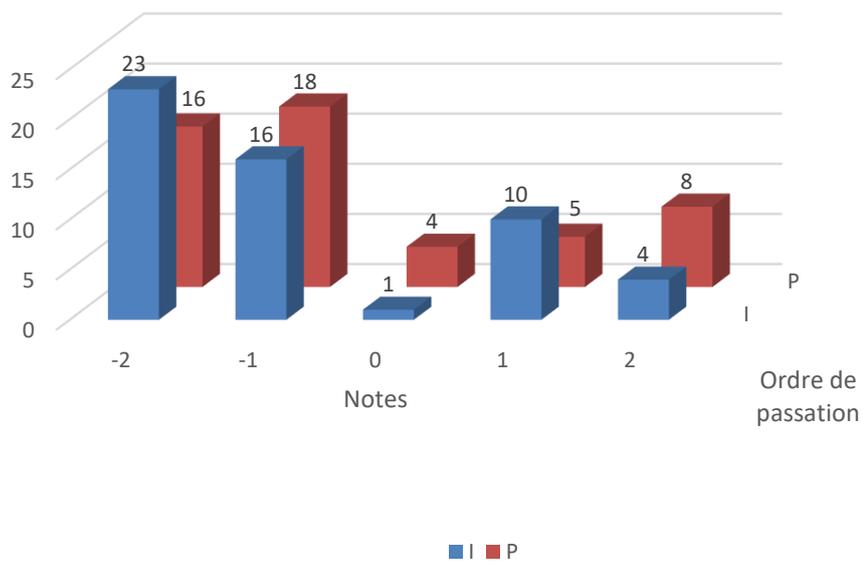


Figure 61 : Histogramme présentant la répartition des notes de préférences des utilisateurs envers une modalité (D, S ou N) en fonction de l'ordre de passation.

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value = 0,179, indique donc que la répartition des notes est indépendante de l'ordre de passation.

*Comparaison Modalité S versus Modalité D : préférences des utilisateurs*

Nous avons donc appliqué la même procédure pour comparer la répartition des notes de préférences envers une modalité sur l'ensemble des utilisateurs.

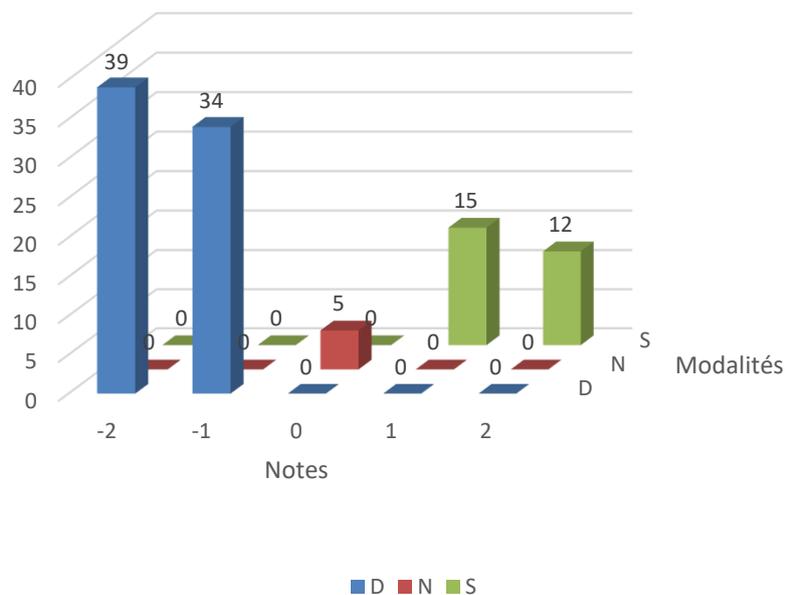


Figure 62 : Histogramme présentant la répartition des notes de préférences envers une modalité (D, S ou N) de l'ensemble des utilisateurs.

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value  $< 0,0001$ , indique donc que la répartition des notes est dépendante de la préférence envers une modalité. Ainsi, après application d'un test z de comparaison de proportions,  $p$ -value  $< 0,0001$ , nous pouvons conclure que nous avons une différence significative de préférence entre les modalités. Les utilisateurs préfèrent donc significativement la modalité D.

### Comparaisons du ressenti de rapidité d'une modalité par rapport à l'autre

#### Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs

Nous avons utilisé un tableau de contingence afin de vérifier que l'ordre de passation n'influe pas sur le ressenti de rapidité des utilisateurs envers une modalité. Le *Tableau 42* présente la proportion des notes attribuées en fonction de l'ordre de passation (groupes Pairs ou Impairs). Les points -1 et -2 de l'échelle indiquent une préférence pour la modalité D, les points 1 et 2 pour la modalité S et le point 0 aucune préférence, symbolisée par « N ».

Tableau 42 : Statistiques descriptives de la répartition du ressenti de rapidité des utilisateurs pour chaque modalité.

Modalités	Notes sur l'échelle de Likert	Notes données par Groupe Pairs	Notes données par Groupe Impairs	Total des notes
D	-2	10	8	18
	-1	8	11	19
N	0	5	8	13
S	1	10	13	23
	2	18	14	32

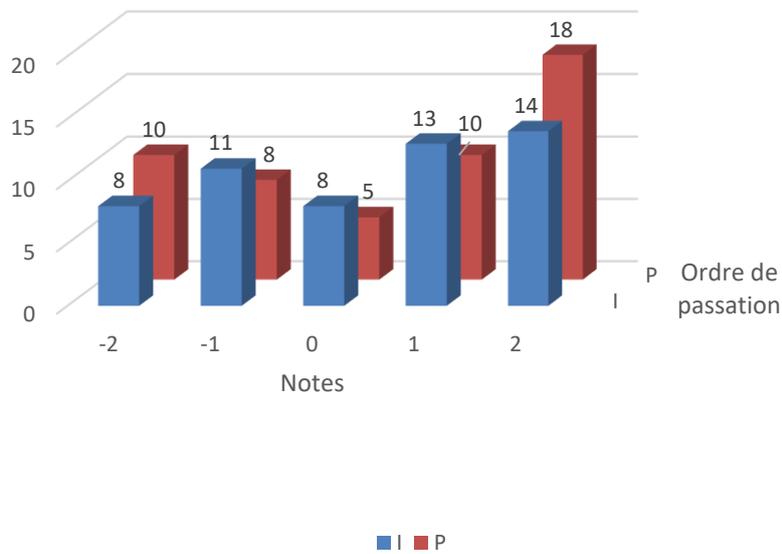


Figure 63 : Histogramme présentant la répartition des notes de ressenti de rapidité des utilisateurs vis-à-vis d'une modalité (D, S ou N) en fonction de l'ordre de passation.

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value = 0,698, indique donc que la répartition des notes est indépendante de l'ordre de passation.

#### Comparaison Modalité S versus Modalité D : ressenti de rapidité

Nous avons donc appliqué la même procédure pour comparer la répartition des notes de ressenti de rapidité vis-à-vis d'une modalité pour l'ensemble des utilisateurs.

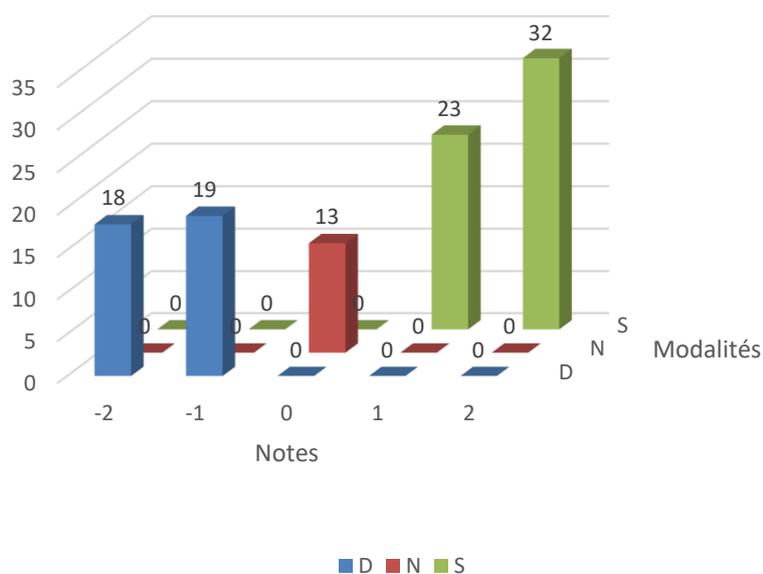


Figure 64 : Histogramme présentant la répartition des notes pour le ressenti de rapidité vis-à-vis une modalité (D, S ou N) de l'ensemble des utilisateurs.

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value  $< 0,0001$ , indique donc que la répartition des notes est dépendante du ressenti de rapidité des utilisateurs vis-à-vis d'une modalité. Ainsi, après application d'un test  $z$  de comparaison de proportions,  $p$ -value = 0,005, nous pouvons conclure que nous avons une différence significative de ressenti de rapidité entre les modalités. Les utilisateurs pensent de manière significative qu'ils sont plus rapides en modalité S.

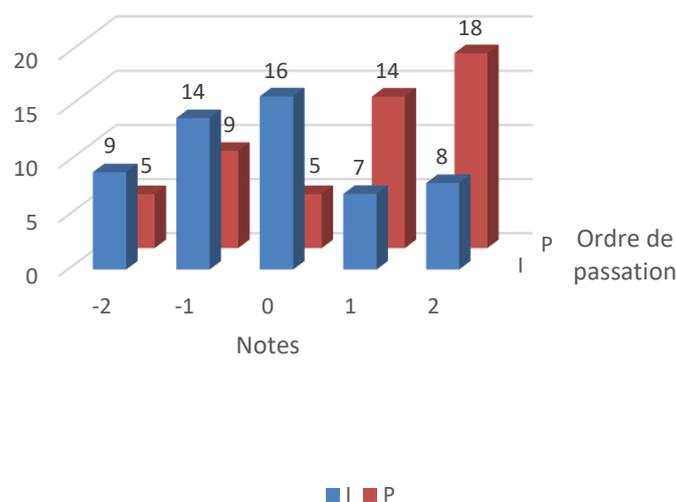
## Comparaisons du sentiment d'efficacité des utilisateurs d'une modalité par rapport à l'autre

### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Nous avons utilisé un tableau de contingence afin de vérifier que l'ordre de passation n'influe pas sur le sentiment d'efficacité des utilisateurs vis-à-vis une modalité. Le *Tableau 43* présente la proportion des notes attribuées en fonction de l'ordre de passation (groupes Pairs ou Impairs). Les points -1 et -2 de l'échelle indiquent une préférence pour la modalité D, les points 1 et 2 pour la modalité S et le point 0 aucune préférence, symbolisée par « N ».

*Tableau 43 : Statistiques descriptives de la répartition du sentiment d'efficacité des utilisateurs pour chaque modalité.*

Modalités	Notes sur l'échelle de Likert	Notes données par Groupe Pairs	Notes données par Groupe Impairs	Total des notes
<b>D</b>	-2	5	9	14
	-1	9	14	23
<b>N</b>	0	5	16	21
<b>S</b>	1	14	7	21
	2	18	8	26



*Figure 65 : Histogramme présentant la répartition des notes du sentiment d'efficacité des utilisateurs vis-à-vis d'une modalité (D, S ou N) en fonction de l'ordre de passation.*

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value = 0,006, indique donc que la répartition des notes est dépendante de l'ordre de passation. Nous avons donc comparé le sentiment d'efficacité envers une modalité pour chaque ordre de passation.

*Comparaison Modalité S versus Modalité D pour les groupes Pairs : sentiment d'efficacité*

Après application d'un test z de comparaison de proportions,  $p$ -value < 0,0001, nous pouvons conclure que nous avons une différence significative de sentiment d'efficacité des utilisateurs entre les modalités pour l'ordre de passation S-D (groupes Pairs). Les utilisateurs pensent de manière significative qu'ils sont plus efficaces en modalité S.

*Comparaison Modalité S versus Modalité D pour les groupes Impairs : sentiment d'efficacité*

Après application d'un test z de comparaison de proportions,  $p$ -value = 0,05, nous pouvons conclure que nous avons une différence significative de sentiment d'efficacité des utilisateurs entre les modalités pour l'ordre de passation D-S (groupes Impairs). Les utilisateurs pensent de manière significative qu'ils sont plus efficaces en modalité D.

**Comparaison de l'envie d'utilisation de la modalité D dans les lieux publics.**

*Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Nous avons utilisé un tableau de contingence afin de vérifier que l'ordre de passation n'influe pas sur l'envie d'utiliser la modalité D dans les lieux publics. Le *Tableau 42* présente la proportion des notes attribuées en fonction de l'ordre de passation (groupes Pairs ou Impairs). Les points 1 et 2 de l'échelle indiquent un désaccord pour l'utilisation de la modalité D dans les lieux publics, les points 4 et 5 indiquent au contraire un accord pour l'utilisation de la modalité D dans les lieux publics et le point 3 indiquent que les utilisateurs sont ni pour ni contre l'utilisation de la modalité D dans les lieux publics, symbolisé par « N ».

*Tableau 44 : Statistiques descriptives de la répartition de l'envie d'utilisation de la modalité D dans les lieux publics en fonction de l'ordre de passation.*

Modalités	Notes sur l'échelle de Likert	Notes données par Groupe Pairs	Notes données par Groupe Impairs	Total des notes
<b>contre</b>	1	4	3	7
	2	7	7	14
<b>N</b>	3	8	5	13
<b>pour</b>	4	18	27	45
	5	14	12	26

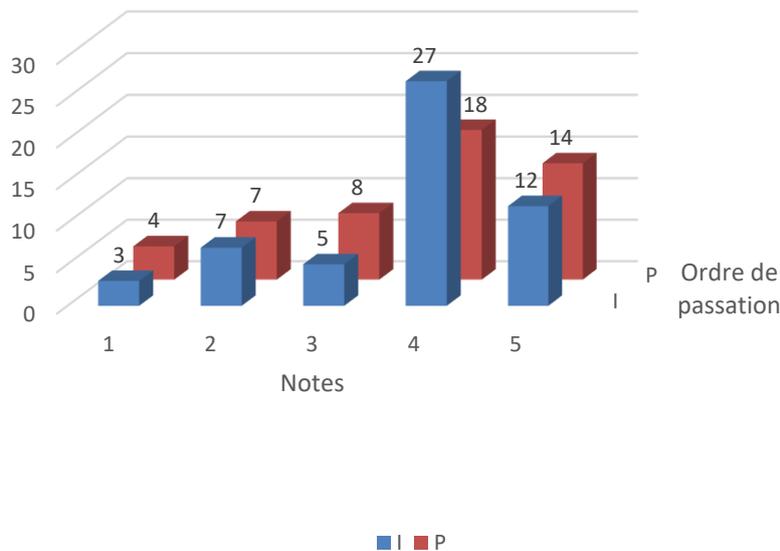


Figure 66 : Histogramme présentant la répartition des notes indiquant l'envie d'utilisation de la modalité D en fonction de l'ordre de passation.

Le résultat du test d'indépendance du  $G^2$  de Wilks,  $p$ -value = 0,608, indique donc que la répartition des notes est indépendante de l'ordre de passation.

#### Comparaison pour ou contre l'utilisation de la Modalité D dans les lieux publics

Après application d'un test z de comparaison de proportions,  $p$ -value < 0,0001, nous pouvons conclure que nous avons une différence significative d'envie d'utilisation de la modalité D dans les lieux publics. Les utilisateurs se prononcent significativement en faveur de l'utilisation de la modalité D dans les lieux publics.

#### Comparaison des notes attribuées à l'aide du questionnaire AttrakDiff

Les résultats du questionnaire *AttrakDiff* se divisent en quatre sous-échelles :

- l'échelle de Qualité Pragmatique (QP) qui décrit l'utilisabilité du produit et indique le niveau de facilité perçue des utilisateurs à atteindre leurs buts ;
- l'échelle de Qualité Hédonique – Stimulation (QHS) qui indique dans quelle mesure le produit soutient le besoin de stimulation en proposant des contenus, fonctionnalités, styles d'interaction nouveaux, intéressants et stimulants ;
- l'échelle de Qualité Hédonique – Identité (QHI) qui indique dans quelle mesure le produit soutient une fonction sociale et communique une certaine identité de l'utilisateur ;
- l'échelle d'Attractivité globale (ATT) qui décrit la valeur globale perçue du produit basée sur la perception des qualités pragmatiques et hédoniques.

Pour chaque item à compléter l'utilisateur doit se positionner sur une échelle de Likert en 7 points avec à chaque extrémité les mots servant à la comparaison comme le montre la Figure 67.



Figure 67 : Exemple d'item (paire de mots) à remplir utilisé dans l'AttrakDiff.

Les items sont donc regroupés par sous-échelles et placés autour d'un continuum avec au centre la valeur neutre 0, ce qui permet de distinguer très rapidement quels aspects sont perçus comme négatifs et quels aspects sont perçus comme positifs. Les valeurs extrêmes (entre -2 et -3 ou à l'inverse entre +2 et +3) sont particulièrement intéressantes. Elles montrent quelles dimensions sont critiques ou au contraire particulièrement positives, et appellent à des actions d'amélioration sur ces aspects.

#### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Afin de vérifier l'absence d'un effet d'ordre de passation de l'expérience sur l'évaluation de chaque modalité et pour chacun des scores des différentes sous-échelles, nous avons utilisé le test non paramétrique de Kruskal-Wallis. Le *Tableau 45* présente la description des données étudiées pour l'évaluation de la modalité D :

*Tableau 45 : Statistiques descriptives des données étudiées pour comparer les scores des sous-échelles de l'AttrakDiff pour la modalité D en fonction de l'ordre de passation (P et I).*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
QP   I	54	-1,857	1,857	0,114	0,794
QP   P	51	-2,429	2,000	-0,160	1,034
QHS   I	54	0,000	2,857	1,717	0,628
QHS   P	51	-0,857	2,857	1,580	0,765
QHI   I	54	-0,857	1,857	0,357	0,563
QHI   P	51	-1,286	2,571	0,454	0,774
ATT   I	54	-2,000	2,286	0,741	0,867
ATT   P	51	-1,286	2,571	0,605	1,118

Le *Tableau 46* présente les p-values associées aux comparaisons de chaque score des sous-échelles :

*Tableau 46 : Tableau récapitulatif des p-values associées aux comparaisons de chaque score des sous-échelles pour la modalité D en fonction de l'ordre de passation*

Variable\Test Kruskal-Wallis	p-value
QP	0,117
QHS	0,358
QHI	0,366
ATT	0,428

Suite à ces résultats, nous pouvons conclure que nous n'obtenons pas d'effet d'ordre de passation de l'expérience sur l'évaluation de la modalité D et ce quel que soit le score attribué

à chaque sous-échelles. Nous poursuivons avec l'analyse de l'effet d'ordre de passation sur l'évaluation de la modalité S.

Le *Tableau 47* présente la description des données étudiées pour l'évaluation de la modalité S :

*Tableau 47 : Statistiques descriptives des données étudiées pour comparer les scores des sous-échelles de l'AttrakDiff pour la modalité S en fonction de l'ordre de passation (P et I).*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
QP   I	54	-1,000	2,857	0,672	0,852
QP   P	51	-2,286	2,143	0,319	0,895
QHS   I	54	-2,857	2,143	-0,365	1,311
QHS   P	51	-2,571	2,857	0,420	1,340
QHI   I	54	-2,571	1,286	-0,257	0,715
QHI   P	51	-2,000	1,714	-0,204	0,903
ATT   I	54	-2,429	2,286	-0,085	1,042
ATT   P	51	-2,429	2,143	-0,011	1,015

Le *Tableau 48* présente les p-values associées aux comparaisons de chaque score des sous-échelles :

*Tableau 48 : Tableau récapitulatif des p-values associées aux comparaisons de chaque score des sous-échelles pour la modalité S en fonction de l'ordre de passation.*

Variable\Test Kruskal-Wallis	p-value
QP	0,066
QHS	<b>0,004</b>
QHI	0,913
ATT	0,588

Suite à ces résultats, nous pouvons conclure que nous obtenons un effet d'ordre de passation de l'expérience sur l'évaluation de la modalité S pour le score de la sous-échelle QHS. Nous ne pouvons donc pas comparer l'ensemble des scores des sous-échelles pour chaque modalité sur l'ensemble de la population. Nous poursuivons donc l'analyse en comparant les scores obtenus par chaque modalité selon chaque ordre de passation.

#### *Comparaison des scores AttrakDiff obtenus par chaque modalité pour l'ordre de passation D-S*

Dans le but de comparer l'ensemble des scores obtenus pour chaque sous-échelle pour les deux modalités (D et S), nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour échantillons appariés. Le *Tableau 49* présente la description des données étudiées pour la comparaison des évaluations de la modalité S et D :

Tableau 49 : Statistiques descriptives des données étudiées pour comparer les scores des sous-échelles de l'AttrakDiff entre la modalité D et la modalité S pour l'ordre de passation D-S.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
QP   D	54	-1,857	1,857	0,114	0,794
QP   S	54	-1,000	2,857	0,672	0,852
QHS   D	54	0,000	2,857	1,717	0,628
QHS   S	54	-2,857	2,143	-0,365	1,311
QHI   D	54	-0,857	1,857	0,357	0,563
QHI   S	54	-2,571	1,286	-0,257	0,715
ATT   D	54	-2,000	2,286	0,741	0,867
ATT   S	54	-2,429	2,286	-0,085	1,042

Le Tableau 50 présente les p-values associées aux comparaisons de chaque score des sous-échelles :

Tableau 50 : Tableau récapitulatif des p-values associées aux comparaisons de chaque score des sous-échelles entre la modalité S et la modalité D pour l'ordre de passation D-S.

Variable	p-value test du signe	p-value test de Wilcoxon signé
QP	< 0,0001	0,0002
QHS	< 0,0001	< 0,0001
QHI	< 0,0001	< 0,0001
ATT	0,0001	< 0,0001

Suite à ces résultats, nous pouvons conclure que nous obtenons une différence significative pour chaque score des sous-échelles entre la modalité S et la modalité D. Pour la sous-échelle QP, la modalité S obtient un score significativement plus élevé que la modalité D. Pour les autres sous-échelles (QHS, QHI et ATT), la modalité D obtient des scores significativement plus élevés que la modalité S.

#### Comparaison des scores AttrakDiff obtenus par chaque modalité pour l'ordre de passation S-D

Dans le but de comparer l'ensemble des scores obtenus pour chaque sous-échelle pour les deux modalités (D et S), nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour échantillons appariés. Le Tableau 51 présente la description des données étudiées pour la comparaison des évaluations de la modalité S et D :

Tableau 51 : Statistiques descriptives des données étudiées pour comparer les scores des sous-échelles de l'AttrakDiff entre la modalité D et la modalité S pour l'ordre de passation S-D.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
QP   D	51	-2,429	2,000	-0,160	1,034
QP   S	51	-2,286	2,143	0,319	0,895
QHS   D	51	-0,857	2,857	1,580	0,765
QHS   S	51	-2,571	2,857	0,420	1,340
QHI   D	51	-1,286	2,571	0,454	0,774
QHI   S	51	-2,000	1,714	-0,204	0,903
ATT   D	51	-1,286	2,571	0,605	1,118
ATT   S	51	-2,429	2,143	-0,011	1,015

Le Tableau 52 présente les p-values associées aux comparaisons de chaque score des sous-échelles :

Tableau 52 : Tableau récapitulatif des p-values associées aux comparaisons de chaque score des sous-échelles entre la modalité S et la modalité D pour l'ordre de passation S-D.

Variable	p-value test du signe	p-value test de Wilcoxon signé
QP	<b>0,006</b>	<b>0,005</b>
QHS	< <b>0,0001</b>	< <b>0,0001</b>
QHI	<b>0,0001</b>	< <b>0,0001</b>
ATT	<b>0,013</b>	<b>0,001</b>

Suite à ces résultats, nous pouvons conclure que nous obtenons une différence significative pour chaque score des sous-échelles entre la modalité S et la modalité D. Pour la sous-échelle QP, la modalité S obtient un score significativement plus élevé que la modalité D. Pour les autres sous-échelles (QHS, QHI et ATT), la modalité D obtient des scores significativement plus élevés que la modalité S.

## Comparaison des notes attribuées à l'aide du questionnaire SUS

Pour chaque item à compléter l'utilisateur doit se positionner sur une échelle de Likert en 5 points avec à l'extrémité gauche la notion « Pas du tout d'accord » et à l'extrémité droite la notion « Tout à fait d'accord ». L'ensemble des scores des dix items donne une note sur 100.

### *Effet d'ordre des modalités : Groupes Pairs versus Groupes Impairs*

Afin de vérifier l'absence d'un effet d'ordre de passation de l'expérience sur le score global obtenu lors de l'évaluation de chaque modalité, nous avons utilisé le test non paramétrique de Mann-Whitney. Le Tableau 53 présente la description des données étudiées pour l'évaluation de la modalité D :

Tableau 53 : Statistiques descriptives des données étudiées pour comparer le score global du SUS pour la modalité D en fonction de l'ordre de passation (P et I).

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score SUS   P	51	25,000	95,000	59,510	16,140
Score SUS   I	54	30,000	90,000	67,407	14,033

Nous obtenons une différence significative de score globale attribué à la modalité D en fonction de l'ordre de passation, p-value = 0,009. Nous poursuivons avec la même analyse pour la modalité S.

Le Tableau 54 présente la description des données étudiées pour l'évaluation de la modalité S :

Tableau 54 : Statistiques descriptives des données étudiées pour comparer le score global du SUS pour la modalité S en fonction de l'ordre de passation (P et I).

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score SUS   I	54	27,500	100,000	71,574	16,399
Score SUS   P	51	32,500	92,500	66,961	15,028

Nous n'observons pas de différence significative de score global attribué à la modalité S en fonction de l'ordre de passation, p-value = 0,098.

Suite à ces résultats, nous pouvons conclure que nous obtenons un effet d'ordre de passation de l'expérience sur l'évaluation de la modalité D pour le score global du SUS. Nous ne pouvons donc pas comparer l'ensemble des scores globaux du SUS pour chaque modalité sur l'ensemble de la population. Nous poursuivons donc l'analyse en comparant les scores obtenus par chaque modalité selon chaque ordre de passation.

#### *Comparaison des scores SUS obtenus par chaque modalité pour l'ordre de passation D-S*

Dans le but de comparer les scores globaux obtenus pour les deux modalités (D et S), nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour échantillons appariés. Le Tableau 55 présente la description des données étudiées pour la comparaison des évaluations de la modalité S et D :

Tableau 55 : Statistiques descriptives des données étudiées pour comparer les scores globaux du SUS entre la modalité D et la modalité S pour l'ordre de passation D-S.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score SUS   D	54	30,000	90,000	67,407	14,033
Score SUS   S	54	27,500	100,000	71,574	16,399

Nous n'observons pas de différence significative de score global entre la modalité S et la modalité D pour l'ordre de passation de l'expérience D-S, p-value = 0,284 pour le test du signe et une p-value = 0,066 pour le test de Wilcoxon signé.

### Comparaison des scores SUS obtenus par chaque modalité pour l'ordre de passation S-D

Afin de comparer les scores globaux obtenus pour les deux modalités (D et S), nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour échantillons appariés. Le *Tableau 56* présente la description des données étudiées pour la comparaison des évaluations de la modalité S et D :

*Tableau 56 : Statistiques descriptives des données étudiées pour comparer les scores globaux du SUS entre la modalité D et la modalité S pour l'ordre de passation S-D.*

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score SUS   D	51	25,000	95,000	59,510	16,140
Score SUS   S	51	32,500	92,500	66,961	15,028

Nous observons, une différence significative de score global entre la modalité S et la modalité D pour l'ordre de passation de l'expérience S-D, p-value = 0,001 pour le test du signe et une p-value = 0,001 pour le test de Wilcoxon signé. Les utilisateurs attribuent de manière significative une note plus élevée à la modalité S pour l'ordre de passation de l'expérience S-D.

## Comparaisons des scores du SUS et de l'AttrakDiff

Pour réaliser cette comparaison des scores entre les deux questionnaires, nous avons comparé le score de la sous-échelle Qualité Pragmatique (QP) de l'*AttrakDiff* et le score global du *SUS* pour chaque modalité. En effet, le *SUS* permet seulement une évaluation de la dimension pragmatique d'un système.

Pour comparer ces deux scores nous avons ramené le score obtenu à la sous-échelle QP de l'*AttrakDiff* à un score sur 100 comme celui de son homologue. Les items de l'*AttrakDiff* sont basés sur une échelle de Likert en 7 points, pour ramener le score sur 100, nous avons procédé à la normalisation suivante :

$$Y = 16,6667 * x + 50 ;$$

où :

- y : est le score sur 100 obtenu après normalisation ;
- x : la valeur du score QP à normaliser.

En normalisant ainsi les scores *SUS* et *AttrakDiff* attribués à chaque modalité (D et S) pour chaque ordre de passation, nous pouvons procéder à leur comparaison.

## Comparaison des scores SUS et AttrakDiff de la modalité D

### Pour l'ordre de passation D-S

Nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour réaliser la comparaison. Le *Tableau 57* présente la description des données étudiées pour la comparaison des évaluations de la modalité D selon les deux questionnaires :

Tableau 57 : Statistiques descriptives de la comparaison des scores *SUS* et *AttrakDiff* obtenus pour la modalité D pour l'ordre de passation D-S.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score <i>SUS</i>	54	19,048	80,952	51,896	13,231
Score <i>AttrakDiff</i>	54	30,000	90,000	67,407	14,033

Nous observons une différence significative des scores *SUS* et *AttrakDiff* attribués à la modalité D pour l'ordre de passation de l'expérience D-S, p-value < 0,0001 pour le test du signe et une p-value < 0,0001 pour le test de Wilcoxon signé.

#### *Pour l'ordre de passation S-D*

Nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour réaliser la comparaison. Le *Tableau 58* présente la description des données étudiées pour la comparaison des évaluations de la modalité D selon les deux questionnaires :

Tableau 58 : Statistiques descriptives de la comparaison des scores *SUS* et *AttrakDiff* obtenus pour la modalité D pour l'ordre de passation S-D.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score <i>SUS</i>	51	25,000	95,000	59,510	16,140
Score <i>AttrakDiff</i>	51	9,524	83,333	47,339	17,239

Nous observons une différence significative des scores *SUS* et *AttrakDiff* attribués à la modalité D pour l'ordre de passation de l'expérience S-D, p-value < 0,0001 pour le test du signe et une p-value < 0,0001 pour le test de Wilcoxon signé.

## Comparaison des scores *SUS* et *AttrakDiff* de la modalité S

#### *Pour l'ordre de passation D-S*

Nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour réaliser la comparaison. Le *Tableau 59* présente la description des données étudiées pour la comparaison des évaluations de la modalité S selon les deux questionnaires :

Tableau 59 : Statistiques descriptives de la comparaison des scores *SUS* et *AttrakDiff* obtenus pour la modalité S pour l'ordre de passation D-S.

Variable	Observations	Minimum	Maximum	Moyenne	Ecart-type
Score <i>SUS</i>	54	27,500	100,000	71,574	16,399
Score <i>AttrakDiff</i>	54	33,333	97,619	61,199	14,208

Nous observons une différence significative des scores *SUS* et *AttrakDiff* attribués à la modalité S pour l'ordre de passation de l'expérience D-S, p-value < 0,0001 pour le test du signe et une p-value < 0,0001 pour le test de Wilcoxon signé.

*Pour l'ordre de passation S-D*

Nous avons utilisé les tests non paramétriques du signe et de Wilcoxon signé pour réaliser la comparaison. Le *Tableau 60* présente la description des données étudiées pour la comparaison des évaluations de la modalité S selon les deux questionnaires :

*Tableau 60 : Statistiques descriptives de la comparaison des scores SUS et AttrakDiff obtenus pour la modalité S pour l'ordre de passation S-D.*

<b>Variable</b>	<b>Observations</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Moyenne</b>	<b>Ecart-type</b>
<b>Score SUS</b>	51	32,500	92,500	66,961	15,028
<b>Score AttrakDiff</b>	51	11,905	85,714	55,322	14,912

Nous observons une différence significative des scores *SUS* et *AttrakDiff* attribués à la modalité S pour l'ordre de passation de l'expérience S-D, p-value < 0,0001 pour le test du signe et une p-value < 0,0001 pour le test de Wilcoxon signé.

# Discussion

## Effet des modalités sur la performance de consultation

Dans un premier temps, nous avons présenté les résultats de performance concernant le nombre d'erreurs commises par les utilisateurs dans leur tâche de comptage de caractères et donc de consultation de l'écran. Nous avons mis en évidence que nous n'avons pas d'effet d'ordre de passation de l'expérience sur cette variable. Les résultats montrent que les utilisateurs font significativement moins d'erreurs en modalité D. La qualité de consultation est donc améliorée par cette modalité d'affichage. En effet, offrir la possibilité aux utilisateurs de dupliquer l'information voulue et de la déplacer dans une zone de confort optimal favorise la qualité de prise d'information.

Dans un second temps, nous avons observé les résultats des utilisateurs ayant trouvé le nombre exact de caractères. Ceux-ci ne sont pas dépendant d'un effet d'ordre de passage de l'expérience. Cependant, contrairement à ce que nous observions précédemment, ils ne présentent pas de différence significative entre les deux modalités. Malgré, une moyenne légèrement plus élevée en modalité D nous ne pouvons pas conclure de manière indiscutable sur cette variable.

Enfin, dans un troisième temps, nous avons présenté les résultats concernant le temps passé pour réaliser l'ensemble des missions. Pour cette variable, nous avons mis en évidence un effet d'ordre de passation de l'expérience. Les groupes exposés à l'ordre de passation des modalités S puis D (groupe pair) sont significativement plus longs pour réaliser leur mission que ceux exposés à l'ordre de passation D puis S. La tâche de comptage est particulièrement fatigante. Ainsi, la différence significative de rapidité entre les deux ordres de passation s'explique par l'addition de la fatigue engendrée par la tâche et l'arrivée de l'aspect novateur des interactions apportées par la modalité D dans un second temps. Quel que soit l'ordre de passation de l'expérience, nous avons montré que les utilisateurs sont significativement plus rapides pour réaliser le comptage de caractères en modalité S. Cette différence significative en temps s'explique notamment par le choix que nous avons fait concernant le cas interactif : CI-0-2-2. En effet, si l'utilisateur a déjà une affiche dans sa zone utilisateur, il doit attendre 4 secondes que celle-ci disparaisse avant de pouvoir en dupliquer une nouvelle. De plus, si par malchance sa zone d'attention n'est pas positionnée sur l'affiche qu'il souhaite déplacer au moment où sa zone utilisateur ne contient plus d'affiche il duplique l'affiche sur laquelle sa zone d'attention est positionnée et doit donc attendre à nouveau 4 secondes pour récupérer l'affiche désirée. Afin de limiter cette perte de temps générée par nos choix de conception, nous pensons tester une alternative prochainement. Celle-ci consiste à implémenter une interaction permettant de supprimer instantanément l'affiche présente dans sa zone utilisateur (via la détection d'un mouvement de rotation important de la tête indiquant un changement de point d'intérêt de l'utilisateur).

## Ressenti des utilisateurs envers les modalités

Dans un premier temps, nous avons analysé la préférence des utilisateurs envers une modalité. Nous avons pu mettre en évidence que nous n'avons pas d'effet d'ordre de passation de l'expérience influant la préférence des utilisateurs pour une modalité ou l'autre. Les résultats montrent que de manière significative les utilisateurs préfèrent la modalité D à la modalité S. En poursuivant l'analyse, nous avons pu mettre en évidence que le sentiment de rapidité des utilisateurs n'était pas influencé par l'effet d'ordre de passation de l'expérience. Les résultats montrent que les utilisateurs sont très conscients qu'ils sont plus rapides en modalité S mais

préfèrent quand même la modalité D. La rapidité de prise d'information ne semble donc pas primer sur la qualité et le confort de consultation lorsque les utilisateurs doivent passer du temps devant un affichage.

Le sentiment d'efficacité, quant à lui, est influencé par un effet d'ordre de passation de l'expérience. Les groupes ayant réalisé l'expérience dans l'ordre modalité S puis modalité D (S-D) estiment être plus efficaces en modalité S contrairement aux utilisateurs ayant passé l'expérience dans l'ordre D-S qui pensent être plus efficaces en modalité D. Il apparaît donc que les utilisateurs pensent être plus efficaces selon la première modalité d'affichage à laquelle ils sont exposés. Ce constat nous laisse penser que les utilisateurs ont fini par fatiguer sur la deuxième partie de l'expérience notamment à cause de la tâche de comptage particulièrement fatigante et rébarbative.

Nous avons ensuite étudié l'éventualité d'une utilisation de la modalité D dans les lieux publics, si elle venait à être disponible. L'analyse de la projection d'utilisation de la modalité D dans les lieux publics n'a pas mis en avant d'effet d'ordre de passation de l'expérience. Les résultats montrent que les utilisateurs sont significativement favorables à l'utilisation de cette modalité dans les lieux publics.

L'ensemble des résultats concernant le ressenti des utilisateurs à l'égard de la modalité D est très encourageant pour la suite de nos travaux. En effet, les utilisateurs semblent accueillir très favorablement ces nouvelles interactions et se disent même prêts à les utiliser dans la vie de tous les jours.

## Résultats de l'évaluation de la modalité D et de la modalité S

Les évaluations des modalités d'affichage réalisées par les utilisateurs via le questionnaire *AttrakDiff* ont été comparées selon chaque ordre de passation de l'expérience.

Pour l'ensemble des utilisateurs, nous obtenons un score pour la sous-échelle QP significativement plus élevé pour la modalité S. Pour les autres sous-échelles (QHS, QHI et ATT), la modalité D obtient des scores significativement plus élevés que la modalité S.

Ces résultats s'expliquent en grande partie par le stade de développement de notre dispositif qui est loin d'être prêt pour un véritable usage en condition écologique. La précision et la robustesse de la détection et du calcul de la zone d'attention ne sont pas encore suffisantes pour proposer des interactions suffisamment précises et pour ne pas frustrer les utilisateurs. C'est d'ailleurs ce point particulier qu'ils mettent en avant dans l'évaluation de l'interface D, avec un score plus faible pour la sous-échelle Qualité pragmatique qui décrit l'utilisabilité du produit et indique le niveau de facilité perçue des utilisateurs à atteindre leurs buts. De plus, comme vu dans la partie *Effet des modalités sur la performance de consultation*, nos premiers choix de conception doivent être améliorés. Néanmoins, si nous observons l'évaluation moyenne globale de chaque modalité, présenté dans la *Figure 68*, nous pouvons rester optimistes et confiants concernant l'intérêt et l'attrait envers ces nouvelles interactions via *head tracking*.

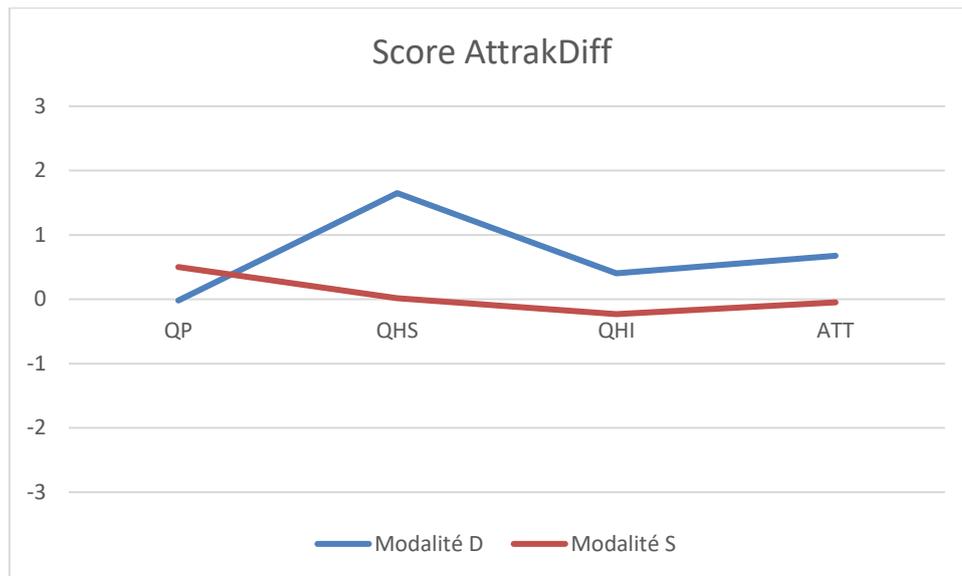


Figure 68 : Histogramme des valeurs moyennes obtenues pour chaque sous-échelle par chaque modalité

En effet, les valeurs proches de la moyenne (zone entre 0 et 1) sont standards : elles ne sont pas négatives et signifie que le produit remplit son objectif sans avoir d'impact négatif. Toutefois, des améliorations sont possibles sur ces aspects pour créer une expérience utilisateur ou une attractivité très positive. Les valeurs en dehors de cette zone neutre sont considérées comme des points positifs (de 1 à 3) ou négatifs (de -1 à -3) du système.

Nous avons ensuite présenté les résultats de comparaison à l'aide du questionnaire *SUS*. Ces résultats sont à interpréter avec prudence car, comme évoqué précédemment, la version française utilisée dans cette expérience n'a jamais été validée. Les résultats du *SUS* semblent amener au même constat que ceux de l'*AttrakDiff*. En effet, la modalité S est jugée meilleure que la modalité D sur la dimension pragmatique. Cependant, nous obtenons une différence significative de score que pour les groupes ayant passé l'expérience dans l'ordre S-D. De manière globale, il est d'usage de considérer un score *SUS* en dessous de 75 comme indicateur que le système n'est pas assez mature. Nous obtenons ici, un score moyen de 69,3 pour la modalité S et un score moyen de 63,6 pour la modalité D qui mettent en avant les besoins d'amélioration du dispositif.

## Comparaison *SUS* et *AttrakDiff*

Ces deux questionnaires servent à l'évaluation d'interface. L'un permet seulement d'intercepter la dimension pragmatique d'une interface, l'autre permet d'en identifier les dimensions hédonique et pragmatique. A notre connaissance, il n'existe pas d'étude comparative des scores obtenus via ces deux questionnaires pour une même interface. L'analyse de la comparaison des scores obtenus via le *SUS* et l'*AttrakDiff* pour la dimension pragmatique de nos deux modalités montre de manière significative une distribution des scores différente. Pour la modalité S, quel que soit l'ordre de passation de l'expérience, le *SUS* donne un score global significativement plus élevé que l'*AttrakDiff*. Pour la modalité D, selon l'ordre de passation de l'expérience, nous observons une inversion. Ainsi, pour l'ordre D-S, l'*AttrakDiff* évalue de manière significativement plus élevée l'interface, alors que pour l'ordre S-D c'est le *SUS*. Ces résultats ne sont pas alarmants. En effet, si nous reprenons l'interprétation globale donnée par les scores de ces deux questionnaires, nous arrivons aux mêmes conclusions. Les deux questionnaires indiquent bien un manque de maturité de notre modalité D. De ce fait, même si la version française du *SUS* n'est à notre connaissance pas

validée à ce jour, elle permet d'arriver dans notre étude aux mêmes conclusions que la version française de l'*AttrakDiff* validée malgré des différences significatives de distribution des scores. Il convient donc d'être prudent, nous recommandons d'utiliser l'*AttrakDiff* qui permet d'identifier une dimension supplémentaire et qui donne une indication plus saillante de la qualité pragmatique d'une interface.

## Conclusions

Cette expérience nous a permis de tester en situation d'usage un dispositif permettant des interactions multi-utilisateurs, sans appareillage ni calibrage préalable via *head tracking* avec un écran de grand format. Nous avons mis en place un protocole expérimental afin de comparer une interface dite dynamique (modalité D), autorisant des duplications et des déplacements de l'information, à une interface dite simple, comparable à une interface classique de consultation sur grand écran n'autorisant pas d'interactions directes. L'analyse de l'expérience montre qu'en termes de qualité de consultation l'interface dynamique est plus adaptée que l'interface statique, même si elle pêche en termes de rapidité. Néanmoins, l'analyse du ressenti des utilisateurs a mis en lumière que ceux-ci préfèrent largement la modalité dynamique, même si ils sont conscients d'être moins rapides et que leur sentiment d'efficacité dans la tâche n'est pas optimal. Les évaluations des deux interfaces réalisées à l'aide des questionnaires *AttrakDiff* et *SUS*, nous permettent de conclure en faveur de la modalité dynamique même si beaucoup de travail reste encore à accomplir pour rendre ce dispositif pleinement utile et utilisable.

La comparaison des résultats obtenus via les deux questionnaires concernant l'évaluation des interfaces va en faveur de l'utilisation du questionnaire *AttrakDiff* qui intercepte plusieurs dimensions pour évaluer les interfaces. Même si cette comparaison nous a permis de mettre en évidence que la version française non validée du *SUS* autorise les mêmes interprétations concernant les deux interfaces évaluées.

## Résumé

Encore aujourd'hui, les écrans de grand format restent le parent pauvre des interactions. Nous les utilisons dans la plupart des cas comme de simples murs d'affiches ou d'informations statiques. Dans ce chapitre, nous présentons une nouvelle manière d'interagir à plusieurs avec les écrans de grand format, sans appareillage et sans calibration individuelle. Ce système basé sur les mouvements de tête des individus permet, à plusieurs utilisateurs en même temps, de consulter dans des conditions de confort optimales, les informations affichées à l'écran. En effet, cet écran dynamique déplace et duplique les informations afin que chaque utilisateur les consulte dans les meilleures conditions possibles. Les résultats montrent que la dynamisation de l'écran augmente significativement la qualité de consultation des utilisateurs. De plus, les utilisateurs présentent un fort attrait pour cette dynamisation des affichages de grand format même si notre dispositif requiert encore un travail conséquent en termes de précision de détection et d'interaction.



# PARTIE III – SYNTHÈSE & PERSPECTIVES

---

*“There is nothing noble in being superior to your fellow man; true nobility is being superior to your former self.”*

*- Ernest Hemingway -*

---



# Chapitre 6 : Synthèse générale, limites et perspectives.

## Synthèse générale

L'objectif de notre travail de thèse était de proposer une nouvelle manière de partager un affichage en situation de groupe ouvert d'utilisateurs, dans un contexte social hétérogène. Nous avons décidé de nous intéresser plus particulièrement aux affichages de grand format verticaux composés d'un seul écran. En partant du constat que ces écrans ne sont que trop souvent de simples supports d'informations statiques. Ainsi, nous avons proposé une nouvelle manière d'interagir avec ces écrans, basée sur la détection et le suivi de la position et de l'orientation de la tête des utilisateurs. Cette nouvelle manière d'interagir avec les écrans de grand format permet à plusieurs utilisateurs non appareillés et sans calibrage de consulter les informations à l'écran dans des conditions de confort optimal. Les principaux enjeux étaient :

- 1) de montrer la faisabilité d'une interaction en groupe ouvert, sans instrumentation des utilisateurs et sans calibrage avec les écrans de grand format basée sur une technique d'*head tracking* ;
- 2) de s'assurer que ce mode d'interaction est effectivement viable dans un contexte multi-utilisateur ;
- 3) de trouver les paradigmes d'interfaces adaptés pour soutenir l'usage de ces interactions.

Le but étant de créer des situations interactionnelles adaptées à différents contextes. En effet, dans le contexte d'étude que nous avons choisi, nous ne pouvions pas imaginer une technique d'interaction nécessitant le port d'artefact, obligeant un calibrage individuel du système ou n'autorisant qu'un seul utilisateur à interagir. C'est pourquoi, le travail présenté dans le Chapitre 3 a permis de mettre en évidence la faisabilité et l'intérêt de ce mode d'interaction et du paradigme d'interface matérialisant la zone d'attention à l'écran. Dans le Chapitre 4, nous nous sommes affranchis des contraintes de calibration individuelle et de port d'artefact pour interagir avec un écran de grand format. Enfin, c'est dans le Chapitre 5 que nous avons pu explorer les prémisses d'interaction en situation d'usage. Notre volonté étant de guider l'adaptation de technologies existantes ou d'aider à la spécification de systèmes innovants.

## La zone d'attention un paradigme d'interface au service de la vision flottante

L'idée développée dans le cadre du Chapitre 3 était donc d'utiliser un dispositif de *tracking* instrumenté ("motion tracking" par infrarouge) pour émuler le dispositif de *head tracking* "idéal" et de déterminer les éventuels freins à la mise en œuvre de ce mode d'interaction multi-utilisateur mais sans calibration ni appareillage.

Nous avons donc proposé d'afficher à l'écran la zone visée par la tête des utilisateurs que nous avons nommée zone d'attention. Grâce à ce paradigme d'interface, nous avons étudié l'influence de la connaissance de la zone regardée par ses pairs sur la performance lors d'une tâche de recherche collaborative d'informations sur écran de grand format.

Suite à une première expérimentation, nous avons confirmé notre idée que la connaissance de l'endroit regardé par ses partenaires a un effet bénéfique sur la collaboration implicite et la performance du groupe. Principalement dans le cas où les groupes ont cette connaissance de l'endroit regardé dans un premier temps puis ne l'ont plus après. Ces premiers résultats ont confirmé l'intérêt et la faisabilité d'une interaction via *head tracking* avec les écrans de grand format.

Néanmoins, nous n'avons malheureusement pas eu le temps d'analyser la distribution des zones d'attention à l'écran en fonction des deux modalités. Cette analyse est en cours et elle devrait nous permettre de mettre en évidence si effectivement les utilisateurs répartissent mieux leurs recherches visuelles lorsque la zone d'attention est affichée. Ainsi, nous pourrions confirmer l'établissement de stratégies de recherches collaboratives implicites liées à la mise en place du mécanisme de la vision flottante par les utilisateurs. De plus, avec le recul, une étude des verbalisations des utilisateurs du type entretien d'explicitation, questionnaire ou entretien d'auto-confrontation concernant les stratégies de recherches conscientes mises en place lors de l'expérience aurait été bénéfique à l'analyse bien que très chronophage.

## **Vers une interaction multi-utilisateur, sans appareillage et sans calibrage avec les écrans de grand format.**

Suite à notre première expérience, nous avons exploré une technique permettant de s'affranchir des contraintes de calibration et d'appareillage des utilisateurs afin d'interagir avec l'écran via *head tracking*. C'est tout naturellement que nous nous sommes donc tournés vers le capteur Kinect V2 et son SDK autorisant des interactions non instrumentées et sans calibrage préalable, basées sur la détection et le suivi de la position et de l'orientation de points caractéristiques sur l'ensemble du corps des utilisateurs. Malheureusement, nous nous sommes rapidement heurtés à l'impossibilité de mettre en œuvre des interactions utiles, utilisables et acceptables via les données brutes fournies par le capteur et son SDK.

Nous avons alors établi un protocole expérimental permettant de caractériser les mouvements de la tête des utilisateurs lors d'une tâche de consultation d'un affichage de grand format. Cette caractérisation nous a permis de nous affranchir d'une calibration personnalisée pour chaque utilisateur utilisant le système.

La récolte de ces données de consultation nous a également servi à établir une procédure de filtrage des données brutes fournies par le capteur et son SDK. Cette procédure de filtrage nous a permis de trouver le meilleur compromis entre latence et stabilité de la zone d'attention à des fins d'interactions viables avec l'écran. De plus, cette expérience nous a permis d'identifier la taille minimale des zones informationnelles déplaçables, dans notre configuration physique et avec le Kinect V2 et ainsi de s'assurer que le capteur est assez précis pour discriminer des rotations de têtes différentes selon la zone informationnelle consultée à l'écran.

Avec le recul, nous pensons que l'utilisation de l'intelligence artificielle couplée à notre procédure devrait autoriser des interactions encore plus précises et sûrement nous permettre de nous affranchir de l'affichage de la zone d'attention. En effet, dans nos futurs travaux nous aimerions comparer deux interfaces :

- la première afficherait la zone d'attention sur l'écran et l'écran réorganiserait ses informations en fonction de la zone d'attention et de la position des utilisateurs ;

- la deuxième n'afficherait plus cette zone d'attention mais l'écran réorganiserait également ses informations en fonction de la zone d'attention et de la position des utilisateurs.

## **Evaluation d'une interaction multi-utilisateur, sans appareillage et sans calibrage via *head tracking* avec les écrans de grand format.**

Grâce aux deux premières expérimentations, nous avons été en mesure de présenter un dispositif autorisant une interaction multi-utilisateur, sans appareillage et sans calibrage basée sur du *head tracking* avec les écrans de grand format. Il ne nous restait plus qu'à proposer et évaluer de nouveaux paradigmes d'interaction qui soient à la fois efficaces, intuitifs et confortables pour les utilisateurs, malgré l'absence de phase de calibrage et d'apprentissage du système. Nous avons présenté une étude comparative entre deux modes d'affichage. Dans un des deux modes, les utilisateurs peuvent dupliquer et déplacer de l'information à l'aide d'interactions basées sur le calcul de la projection de leur zone d'attention sur l'écran tandis que dans l'autre, aucune interaction n'est possible.

Nous avons alors comparé la performance de consultation et le ressenti des utilisateurs par rapport à ces deux modes d'affichage à l'aide de questionnaire d'évaluation d'interface.

Les résultats montrent qu'en termes de qualité de consultation l'interface dynamique est meilleure que l'interface statique même si elle pêche en termes de rapidité. De plus, l'analyse du ressenti des utilisateurs a mis en lumière que ceux-ci préfèrent largement la modalité dynamique même si ils sont conscients d'être moins rapides. Les évaluations des deux interfaces réalisées à l'aide des questionnaires *AttrakDiff* et *SUS*, nous permettent de conclure en faveur de la modalité dynamique même si beaucoup de travail reste encore à accomplir pour rendre ce dispositif complètement utile et utilisable. Nous sommes encore en cours d'analyse des déplacements physiques des utilisateurs afin de mesurer s'ils se déplacent plus ou moins en fonction du mode d'affichage.

A plus long terme, nous imaginons développer un dispositif basé sur des caméras stéréoscopiques et des algorithmes de détection et de suivi de la position et de l'orientation de la tête afin de s'affranchir de des contraintes d'usage en intérieur et de la limitation à 6 utilisateurs simultanés liées au capteur Kinect V2.

Enfin, nous souhaitons également tester rapidement des interactions de réorganisation et d'enrichissement des informations sans forcément de duplication de celles-ci pour simuler davantage des contextes réels d'usages dans des lieux publics. Pour cela, nous devons trouver des cas d'application réels afin de concevoir des interactions adaptées aux contextes sociaux et environnementaux dans lesquels le dispositif sera placé.

Enfin, l'ensemble de notre démarche de développement des technologies et métaphores d'interactions nouvelles pour le partage collaboratif d'affichage en groupe ouvert s'est entièrement reposé sur l'analyse du comportement des utilisateurs. Cette démarche d'ingénierie cognitive nous a permis de ne pas nous enfermer dans une recherche technocentrée. Ainsi, nous avons pu mettre en lumière que la dynamisation des écrans de grand format via une interaction par *head tracking* est une solution bien accueillie par le public et avec un potentiel d'application considérable.



# Bibliographie

## - A -

- Abras, C., D. Maloney-Krichmar and J. Preece (2004). "User-centered design." Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications **37**(4): 445-456.
- AForge.NET, F. (2008-2012). "<http://www.aforgenet.com/framework/>."
- Agamanolis, S. (2003). Designing displays for human connectedness. Public and Situated Displays, Springer: 309-334.
- Ahlvers, U., R. Rajagopalan and U. Zölzer (2005). Model-free face detection and head tracking with morphological hole mapping. Signal Processing Conference, 2005 13th European, IEEE.
- Alt, F., S. Schneegaß, A. Schmidt, J. Müller and N. Memarovic (2012). How to evaluate public displays. Proceedings of the 2012 International Symposium on Pervasive Displays, ACM.
- Alt, F., A. S. Shirazi, T. Kubitzka and A. Schmidt (2013). Interaction techniques for creating and exchanging content with public displays. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Arhippainen, L. (2013). A Tutorial of Ten User Experience Heuristics. Proceedings of International Conference on Making Sense of Converging Media, ACM.
- Ariz, M., J. J. Bengoechea, A. Villanueva and R. Cabeza (2016). "A novel 2D/3D database with automatic face annotation for head tracking and pose estimation." Computer Vision and Image Understanding **148**: 201-210.
- Ashdown, M., K. Oka and Y. Sato (2005). Combining head tracking and mouse input for a GUI on multiple monitors. CHI'05 extended abstracts on Human factors in computing systems, ACM.
- AW, A. and J.-C. ROUHET (2004). "Ergonomie. Système homme-machine." Techniques de l'ingénieur. Informatique industrielle **2**(S7610): S7610. 7611-S7610. 7612.

## - B -

- Ba, S. O. and J.-M. Odobez (2004). A probabilistic framework for joint head tracking and pose estimation. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, IEEE.
- Ba, S. O. and J.-M. Odobez (2011). "Multiperson visual focus of attention from head pose and meeting contextual cues." IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(1): 101-116.
- Ball, R. and C. North (2005). "Analysis of user behavior on high-resolution tiled displays." Human-Computer Interaction-INTERACT 2005: 350-363.
- Ball, R. and C. North (2005). Effects of tiled high-resolution display on basic visualization and navigation tasks. CHI'05 extended abstracts on Human factors in computing systems, ACM.

- Ball, R., C. North and D. A. Bowman (2007). Move to improve: promoting physical navigation to increase user performance with large displays. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Bastien, J. C. and D. L. Scapin (1992). "A validation of ergonomic criteria for the evaluation of human-computer interfaces." International Journal of Human-Computer Interaction 4(2): 183-196.
- Bastien, J. C. and D. L. Scapin (1993). Ergonomic criteria for the evaluation of human-computer interfaces, Inria.
- Basu, S., I. Essa and A. Pentland (1996). Motion regularization for model-based head tracking. Pattern Recognition, 1996., Proceedings of the 13th International Conference on, IEEE.
- Baudisch, P., E. Cutrell, D. Robbins, M. Czerwinski, P. Tandler, B. Bederson and A. Zierlinger (2003). Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch-and pen-operated systems. Proceedings of INTERACT.
- Baudisch, P., E. Cutrell and G. Robertson (2003). High-density cursor: a visualization technique that helps users keep track of fast-moving mouse cursors. Proc. Interact'03.
- Baudisch, P., N. Good and P. Stewart (2001). Focus plus context screens: combining display technology with visualization techniques. Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM.
- Baudisch, P., D. Tan, M. Collomb, D. Robbins, K. Hinckley, M. Agrawala, S. Zhao and G. Ramos (2006). Phosphor: explaining transitions in the user interface using afterglow effects. Proceedings of the 19th annual ACM symposium on User interface software and technology, ACM.
- Beaudouin-Lafon, M. (1992). "Interaction homme-machine." CNRS-Cahiers IMABIO.
- Bednarik, R., H. Vrzakova and M. Hradis (2012). What do you want to do next: a novel approach for intent prediction in gaze-based interaction. Proceedings of the symposium on eye tracking research and applications, ACM.
- Benson, A. and P. Jobson (1973). "Body sway induced by a low frequency alternating current." International journal of equilibrium research 3(1): 55-61.
- Berthoz, A. (1992). "Effets des vibrations sur l'homme." Scherrer J: Précis de physiologie du travail, 2ème édition, Masson édit.
- Bey, C. (2016). Gestion des ressources cognitives et stratégies d'adaptation court terme chez les pilotes d'aéronefs, Université de Bordeaux.
- Beyer, H. and K. Holtzblatt (1997). Contextual design: defining customer-centered systems, Elsevier.
- Bierz, T. (2006). "Interaction technologies for large displays-an overview." Visualization of Large and Unstructured Data Sets, GI-Edition Lecture Notes in Informatics (LNI) 4.
- Bingham, G. P. and M. G. Stassen (1994). "Monocular egocentric distance information generated by head movement." Ecological psychology 6(3): 219-238.
- Birchfield, S. (1997). An elliptical head tracker. Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on, IEEE.
- Birchfield, S. (1998). Elliptical head tracking using intensity gradients and color histograms. Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on, IEEE.
- Bishop, G. and G. Welch (2001). "An introduction to the kalman filter." Proc of SIGGRAPH, Course 8(27599-23175): 41.
- Bona, A. (2016). Modélisation de l'affordance dans le domaine des constructions durables, Université de Bordeaux.
- Borji, A. and L. Itti (2013). "State-of-the-art in visual attention modeling." IEEE transactions on pattern analysis and machine intelligence 35(1): 185-207.

- Borovoy, R., F. Martin, S. Vemuri, M. Resnick, B. Silverman and C. Hancock (1998). Meme tags and community mirrors: moving from conferences to collaboration. Proceedings of the 1998 ACM conference on Computer supported cooperative work, ACM.
- Boy, G. (2003). Ingénierie cognitive: IHM et cognition, Hermès science publications.
- Brignull, H. and Y. Rogers (2003). Enticing people to interact with large public displays in public spaces. Proceedings of INTERACT.
- Broadbent, D. E. (1958). Perception and communication, New York: Oxford University Press.
- Broadbent, D. E. (2013). Perception and communication, Elsevier.
- Brooke, J. (1996). "SUS-A quick and dirty usability scale." Usability evaluation in industry **189**(194): 4-7.
- Brown, C. and D. Schaum (1980). "User-adjusted VDU parameters." Ergonomic aspects of visual display terminals: 195-200.
- Buckley, P. and J. Duhig (2005). Multi-user display system, Google Patents.
- Burke, M., A. Hornof, E. Nilsen and N. Gorman (2005). "High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten." ACM Transactions on Computer-Human Interaction (TOCHI) **12**(4): 423-445.

## - C -

- Carranza, J., C. Theobalt, M. A. Magnor and H.-P. Seidel (2003). Free-viewpoint video of human actors. ACM transactions on graphics (TOG), ACM.
- Carrasco, M. (2011). "Visual attention: The past 25 years." Vision research **51**(13): 1484-1525.
- Casiez, G., N. Roussel and D. Vogel (2012). 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears." The Journal of the acoustical society of America **25**(5): 975-979.
- Cheverst, K., N. Davies, K. Mitchell, A. Friday and C. Efstratiou (2000). Developing a context-aware electronic tourist guide: some issues and experiences. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM.
- Cheverst, K., D. Fitton and A. Dix (2003). "Exploring the evolution of office door displays." Public and Situated Displays. Kluwer: 141-169.
- Chouvardas, V. G., A. N. Miliou and M. K. Hatalis (2008). "Tactile displays: Overview and recent advances." Displays **29**(3): 185-194.
- Churchill, E., L. Nelson, L. Denoue, P. Murphy and J. Helfman (2003). The plasma poster network. Public and Situated Displays, Springer: 233-260.
- Churchill, E. F., L. Nelson, L. Denoue, J. Helfman and P. Murphy (2004). Sharing multimedia content with interactive public displays: a case study. Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques, ACM.
- Claverie, B. (2005). Cognitique: science et pratique des relations à la machine à penser, Editions L'Harmattan.
- Colley, A., M. Pakanen, S. Koskinen, K. Mikkonen and J. Häkkinä (2016). Smart Handbag as a Wearable Public Display-Exploring Concepts and User Perceptions. Proceedings of the 7th Augmented Human International Conference 2016, ACM.

- Collomb, M., M. Hascoët, P. Baudisch and B. Lee (2005). Improving drag-and-drop on wall-size displays. Proceedings of Graphics interface 2005, Canadian Human-Computer Communications Society.
- Colombo, L. and M. Pasch (2012). "Heuristics for an Optimal User Experience." Proc. CHI2012 Altchi. ACM Press.
- Cooper, A. (1995). "The myth of metaphor." Visual Basic Programmer's Journal **3**: 127-128.
- Czerwinski, M., G. Smith, T. Regan, B. Meyers, G. G. Robertson and G. K. Starkweather (2003). Toward characterizing the productivity benefits of very large displays. Interact.

## - D -

- Darby, J., M. B. Sánchez, P. B. Butler and I. D. Loram (2016). "An evaluation of 3D head pose estimation using the Microsoft Kinect v2." Gait & Posture **48**: 83-88.
- Davis, J. and X. Chen (2002). "Lumipoint: Multi-user laser-based interaction on large tiled displays." Displays **23**(5): 205-211.
- Deutsch, J. A. and D. Deutsch (1963). "Attention: Some theoretical considerations." Psychological review **70**(1): 80.
- Dietz, P. and D. Leigh (2001). DiamondTouch: a multi-user touch technology. Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM.
- Dillon, A. and M. Morris (1999). Power, perception and performance: from usability engineering to technology acceptance with the P3 model of user response. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications Sage CA: Los Angeles, CA.
- Dillon, A. and M. G. Morris (1996). User acceptance of new information technology: theories and models. Annual review of information science and technology, Medford, NJ: Information Today.
- DIS, I. (2002). "16982 : Ergonomics of human system interaction - Usability methods supporting human-centred design." International Standardization Organization (ISO). Switzerland.
- DIS, I. (2006). "9241-110: 2006. Ergonomics of human system interaction-Part 110: Dialogue principles." International Standardization Organization (ISO). Switzerland.
- DIS, I. (2009). "9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems." International Standardization Organization (ISO). Switzerland.
- Drewitz, U. and S. Brandenburg (2010). From design to experience: Towards a process model of user experience. Proceedings of the 9th Pan-Pacific Conference on Ergonomics.
- Drouard, V., R. Horaud, A. Deleforge, S. Ba and G. Evangelidis (2017). "Robust head-pose estimation based on partially-latent mixture of linear regressions." IEEE Transactions on Image Processing **26**(3): 1428-1440.
- Duchowski, A. T. (2007). "Eye tracking methodology." Theory and practice **328**.

## - E -

- Eichenlaub, J., B. (2017). Single Person and Multiple Person Full Resolution 2D/3D Head Tracking Display, Google Patents.
- Encaoua, D. and P. Thomet (2017). Method for estimating a reference posture, Google Patents.

- Endert, A., C. Andrews, Y. H. Lee and C. North (2011). Visual encodings that support physical navigation on large displays. Proceedings of Graphics Interface 2011, Canadian Human-Computer Communications Society.
- Endsley, M. R. (1995). "Toward a theory of situation awareness in dynamic systems." Human factors **37**(1): 32-64.
- Eppler, M. J. and J. Mengis (2004). "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines." The information society **20**(5): 325-344.
- Essilor, E. (2017). "<https://www.essilor.fr/Produits/Autres-solutions/technologie-verres-eyecode>."

## - F -

- Fanelli, G., J. Gall and L. Van Gool (2011). Real time head pose estimation with random regression forests. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE.
- Fanelli, G., J. Gall and L. Van Gool (2012). Real time 3d head pose estimation: Recent achievements and future challenges. Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on, IEEE.
- Fitzmaurice, G. W. and W. Buxton (1997). Graspable user interfaces, University of Toronto.
- Foucambert, D. (2003). Syntaxe, vision parafovéale et processus de lecture: contribution du modèle structural à la pédagogie, Université Pierre Mendès France (Grenoble).
- Frenkel, K. A. (1989). "The next generation of interactive technologies." Communications of the ACM **32**(7): 872-881.
- Funkhouser, T. and K. Li (2000). "Large-format displays." IEEE Computer Graphics and Applications **20**(4): 20-21.

## - G -

- Gaver, W. W. (1991). Technology affordances. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Gibson, J. J. (1977). "Perceiving, acting, and knowing: Toward an ecological psychology." The Theory of Affordances: 67-82.
- Gibson, J. J. (2014). The ecological approach to visual perception: classic edition, Psychology Press.
- Grandjean, E., K. Nishiyama, W. Hünting and M. Piderman (1982). "A laboratory study on preferred and imposed settings of a VDT workstation." Behaviour & Information Technology **1**(3): 289-304.
- Grasso, A., M. Mühlenbrock, F. Roulland and D. Snowdon (2003). Supporting communities of practice with large screen displays. Public and Situated Displays, Springer: 261-282.
- Grosz, B. J. (1996). "Collaborative Systems (AAAI-94 Presidential Address)." AI magazine **17**(2): 67.
- Grudin, J. (1988). Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. Proceedings of the 1988 ACM conference on Computer-supported cooperative work, ACM.
- Grudin, J. (2001). Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.

- Guimbretière, F., M. Stone and T. Winograd (2001). Fluid interaction with high-resolution wall-size displays. Proceedings of the 14th annual ACM symposium on User interface software and technology, ACM.
- Guimbretière, F. and T. Winograd (2000). FlowMenu: combining command, text, and data entry. Proceedings of the 13th annual ACM symposium on User interface software and technology, ACM.

## - H -

- Hager, G. D. and P. N. Belhumeur (1998). "Efficient region tracking with parametric models of geometry and illumination." IEEE transactions on pattern analysis and machine intelligence **20**(10): 1025-1039.
- Handford, M. (1987). Où est Charlie ?, Grund.
- Hannun, A., C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta and A. Coates (2014). "Deep speech: Scaling up end-to-end speech recognition." arXiv preprint arXiv:1412.5567.
- Hansen, D. W. (2003). Committing eye tracking, IT University of Copenhagen.
- Hansen, D. W. and Q. Ji (2010). "In the eye of the beholder: A survey of models for eyes and gaze." IEEE transactions on pattern analysis and machine intelligence **32**(3): 478-500.
- Hansen, J. P., D. W. Hansen and A. S. Johansen (2001). Bringing gaze-based interaction back to basics. HCI.
- Hassenzahl, M., M. Burmester and F. Koller (2003). AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality. Mensch & Computer.
- Hécaen, H. (1972). Neuropsychologie de la perception visuelle, Masson.
- Hilliges, O., D. Kim, S. Izadi, M. Weiss and A. Wilson (2012). HoloDesk: direct 3d interactions with a situated see-through display. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Hoffmann, R., P. Baudisch and D. S. Weld (2008). Evaluating visual cues for window switching on large screens. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Huang, E. M. (2007). When does the public look at public displays. In Companion Proceedings of UbiComp 2007, Citeseer.
- Huang, E. M., E. D. Mynatt, D. M. Russell and A. E. Sue (2006). "Secrets to success and fatal flaws: The design of large-display groupware." IEEE Computer Graphics and Applications **26**(1): 37-45.
- Hutchinson, T. E., K. P. White, W. N. Martin, K. C. Reichert and L. A. Frey (1989). "Human-computer interaction using eye-gaze input." IEEE Transactions on systems, man, and cybernetics **19**(6): 1527-1534.
- Huteau, M. (1975). "Un style cognitif: la dépendance-indépendance à l'égard du champ." L'année psychologique **75**(1): 197-262.
- Hyrskykari, A., H. Istance and S. Vickers (2012). Gaze gestures or dwell-based interaction? Proceedings of the Symposium on Eye Tracking Research and Applications, ACM.

## - I -

- Ishii, H. (2007). Tangible user interfaces, CRC Press.
- Ishii, H. and B. Ullmer (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, ACM.

## - J -

- Jacob, R. and K. S. Karn (2003). "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises." Mind **2**(3): 4.
- Jacob, T., G. Bailly, E. Lecolinet, G. Casiez and M. Teyssier (2016). Desktop Orbital Camera Motions Using Rotational Head Movements. Proceedings of the 2016 Symposium on Spatial User Interaction, ACM.
- James, W. (1890). "The principles of psychology (Vol. 1)." New York: Holt **474**.
- Jeni, L. A., J. F. Cohn and T. Kanade (2017). "Dense 3D face alignment from 2D video for real-time use." Image and Vision Computing **58**: 13-24.
- Jordan, B. and A. Henderson (1995). "Interaction analysis: Foundations and practice." The journal of the learning sciences **4**(1): 39-103.
- José, R., N. Otero, S. Izadi and R. Harper (2008). "Instant places: Using bluetooth for situated interaction in public displays." IEEE Pervasive Computing **7**(4).

## - K -

- Kahneman, D. (1973). Attention and effort, Prentice-Hall Englewood Cliffs, NJ.
- Karwowski, W. and J. B. Peacock (1993). Automotive Ergonomics, Taylor & Francis.
- Kennedy, K. (2008). Multi-user display, Google Patents.
- Khalidov, V. and J.-M. Odobez (2017). Real-time multiple head tracking using texture and colour cues, Idiap.
- Khan, A., J. Matejka, G. Fitzmaurice and G. Kurtenbach (2005). Spotlight: directing users' attention on large displays. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Kimball, R. and M. Ross (2011). The data warehouse toolkit: the complete guide to dimensional modeling, John Wiley & Sons.
- Kinect, M. (2017). "<https://developer.microsoft.com/fr-fr/windows/kinect>."
- Kowalski, M., J. Naruniec and M. Daniluk (2015). Live Scan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors. 3D Vision (3DV), 2015 International Conference on, IEEE.
- Krumbholz, C., J. Leigh, A. Johnson, L. Renambot and R. Kooima (2005). Lambda table: high resolution tiled display table for interacting with large visualizations. Proceedings of 5th Workshop on Advanced Collaborative Environments, Redmond, Washington.
- Kuhn, T. S. (1972). "La structure des révolutions scientifiques."

## - L -

- Laberge, D. (1995). "Computational and anatomical models of selective attention in object identification."
- Lachat, E., H. Macher, M.-A. Mittet, T. Landes and P. Grussenmeyer (2015). "First experiences with Kinect v2 sensor for close range 3D modelling." The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **40**(5): 93.
- Lallemant, C. and G. Gronier (2016). "Méthodes de design UX." Paris, Eyrolles.

- Lallemand, C., G. Gronier and V. Koenig (2013). "L'expérience utilisateur: un concept sans consensus? Enquête sur le point de vue des professionnels." Activités humaines, Technologies et Bien-être.
- Lallemand, C., V. Koenig, G. Gronier and R. Martin (2015). "Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs." Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology **65**(5): 239-252.
- Le Boulch, G. (2001). Approche systémique de la proximité: définitions et discussion. 3ème Journées de la Proximité, Ministère de la Recherche.
- LeCun, Y. (1985). "A learning scheme for asymmetric threshold networks." Proceedings of Cognitiva **85**: 599-604.
- Lécuyer, T. (1976). "Psychosociologie de l'espace.—II. Rapports spatiaux interpersonnels et la notion d'«espace personnel»." L'Année psychologique **76**(2): 563-596.
- Lee, M., R. Green and M. Billinghurst (2008). 3D natural hand interaction for AR applications. Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference, IEEE.
- Lee, T. and T. Hollerer (2007). Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. Wearable Computers, 2007 11th IEEE International Symposium on, IEEE.
- Lespinet-Najib, V. (2013). De la neuropsychologie cognitive à la cognitive: vers une recherche transdisciplinaire, université bordeaux 1.
- Lin, J.-W., H. B.-L. Duh, D. E. Parker, H. Abi-Rached and T. A. Furness (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. Virtual Reality, 2002. Proceedings. IEEE, IEEE.
- Lini, S. (2013). L'anticipation et sa représentation dans les interfaces homme-système en aéronautique: L'anticipation et sa représentation dans les interfaces homme-système en aéronautique, Université Sciences et Technologies-Bordeaux I.
- Linn, M. C. and A. C. Petersen (1985). "Emergence and characterization of sex differences in spatial ability: A meta-analysis." Child development: 1479-1498.
- Liu, C., O. Chapuis, M. Beaudouin-Lafon and E. Lecolinet (2016). Shared interaction on a wall-sized display in a data manipulation task. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM.
- Luo, H. and A. Eleftheriadis (2003). "Model-based segmentation and tracking of head-and-shoulder video objects for real time multimedia services." IEEE Transactions on multimedia **5**(3): 379-389.

## - M -

- Mackinlay, J. D. and J. Heer (2004). Wideband displays: mitigating multiple monitor seams. CHI'04 extended abstracts on Human factors in computing systems, ACM.
- Maisonneuve, J. (2014). La dynamique des groupes:«Que sais-je?» n° 1306, Presses universitaires de France.
- Majaranta, P. and K.-J. Rähä (2002). Twenty years of eye typing: systems and design issues. Proceedings of the 2002 symposium on Eye tracking research & applications, ACM.
- Mankoff, J., A. K. Dey, G. Hsieh, J. Kientz, S. Lederer and M. Ames (2003). Heuristic evaluation of ambient displays. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Marc, E. and D. Picard (2002). Interaction. Vocabulaire de psychosociologie, ERES: 189-196.

- Marion, D., P.-A. Favier and J.-M. André (2017). Caractérisation des mouvements de la tête lors d'une tâche de consultation d'un affichage de grand format à l'aide d'un dispositif de head tracking. 29ème conférence francophone sur l'Interaction Homme-Machine, ACM.
- Marion, D., P.-A. Favier, C. Bey, S. Lini and J.-M. André (2015). Etude outillée de l'attention visuelle flottante dans la recherche collaborative d'informations sur grands écrans. EPIQUE 2015, Aix-en-Provence, France, Arpege Science Publishing.
- Markussen, A., M. R. Jakobsen and K. Hornbæk (2014). Vulture: a mid-air word-gesture keyboard. Proceedings of the 32nd annual ACM conference on Human factors in computing systems, ACM.
- Mayhew, D. J. (1999). The usability engineering lifecycle. CHI'99 Extended Abstracts on Human Factors in Computing Systems, ACM.
- McCarthy, J. (1960). Programs with common sense, RLE and MIT Computation Center.
- McCarthy, J. (1960). "Recursive functions of symbolic expressions and their computation by machine, Part I." Communications of the ACM **3**(4): 184-195.
- McCarthy, J. F. (2003). Promoting a sense of community with ubiquitous peripheral displays. Public and Situated Displays, Springer: 283-308.
- Milgram, S. (1970). "The experience of living in cities." Science **167**(3924): 1461-1468.
- Mistry, P. and P. Maes (2009). SixthSense: a wearable gestural interface. ACM SIGGRAPH ASIA 2009 Sketches, ACM.
- Moray, N. (1959). "Attention in dichotic listening: Affective cues and the influence of instructions." Quarterly journal of experimental psychology **11**(1): 56-60.
- Morimoto, C. H. and M. R. Mimica (2005). "Eye gaze tracking techniques for interactive applications." Computer vision and image understanding **98**(1): 4-24.
- Moustakides, G. (1989). "Stabilizing the fast Kalman algorithms." IEEE Transactions on Acoustics, Speech, and Signal Processing **37**(9): 1342-1348.
- Müller, H. J. and P. M. Rabbitt (1989). "Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption." Journal of Experimental psychology: Human perception and performance **15**(2): 315.
- Müller, J., J. Exeler, M. Buzeck and A. Krüger (2009). Reflectivesigns: Digital signs that adapt to audience attention. International Conference on Pervasive Computing, Springer.
- Müller, J. and A. Krüger (2007). How much to bid in Digital Signage Advertising Auctions?, na.
- Müller, J. and A. Krüger (2007). "User profiling for generating bids in digital signage advertising auctions." UBIDEUM 2007: 46.
- Müller, J., O. Paczkowski and A. Krüger (2007). Situated public news and reminder displays. European Conference on Ambient Intelligence, Springer.
- Müller, J., R. Walter, G. Bailly, M. Nischt and F. Alt (2012). Looking glass: a field study on noticing interactivity of a shop window. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Müller, J., D. Wilmsmann, J. Exeler, M. Buzeck, A. Schmidt, T. Jay and A. Krüger (2009). "Display blindness: The effect of expectations on attention towards digital signage." Pervasive Computing: 1-8.
- Murphy-Chutorian, E. and M. M. Trivedi (2009). "Head pose estimation in computer vision: A survey." IEEE transactions on pattern analysis and machine intelligence **31**(4): 607-626.
- Myers, B. A., R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller and A. C. Long (2002). Interacting at a distance: measuring the performance of laser pointers and

other devices. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.

Mynski, M. and S. Papert (1969). "Perceptrons: An introduction to Computational Geometry." MA: MIT Press, Cambridge.

## - N -

Nancel, M., O. Chapuis, E. Pietriga, X.-D. Yang, P. P. Irani and M. Beaudouin-Lafon (2013). High-precision pointing on large wall displays using small handheld devices. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.

NaturalPoint, I. D. O. (2017). "<https://optitrack.com/>."

Navon, D. and D. Gopher (1979). "On the economy of the human-processing system." Psychological review **86**(3): 214.

Ni, T., G. S. Schmidt, O. G. Staadt, M. A. Livingston, R. Ball and R. May (2006). A survey of large high-resolution display technologies, techniques, and applications. Virtual Reality Conference, 2006, IEEE.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM.

Nielsen, J. (1994). "Heuristic evaluation." Usability inspection methods **17**(1): 25-62.

Nielsen, J. and R. Molich (1990). Heuristic evaluation of user interfaces. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.

Nonnecke, B. and J. Preece (2003). "Silent participants: Getting to know lurkers better." From Usenet to CoWebs: Interacting with social information spaces: 110-132.

Norman, D., J. Miller and A. Henderson (1995). What you see, some of what's in the future, and how we go about doing it: HI at Apple Computer. Conference companion on Human factors in computing systems, ACM.

Norman, D. A. (1968). "Toward a theory of memory and attention." Psychological review **75**(6): 522.

Norman, D. A. (1999). "Affordance, conventions, and design." interactions **6**(3): 38-43.

Norman, D. A. (2002). "The design of every things." Basic Books. New York, NY. Pp. Xi-10 ISBN 13: 978-970.

Norman, D. A. and D. G. Bobrow (1975). "On data-limited and resource-limited processes." Cognitive psychology **7**(1): 44-64.

Norman, D. A. and S. W. Draper (1986). "User centered system design." Hillsdale, NJ: 1-2.

Norman, D. A. and J. Nielsen (2010). "Gestural interfaces: a step backward in usability." interactions **17**(5): 46-49.

## - O -

O'Hara, K., M. Perry, E. Churchill and D. Russell (2013). Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies, Springer Netherlands.

O'Hara, K., M. Perry and S. Lewis (2003). "Situated web signs and the ordering of social action." Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies. Kluwer: 105-140.

Olsen Jr, D. R. and T. Nielsen (2001). Laser pointer interaction. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.

Owens, J. W., B. S. Chaparro and E. M. Palmer (2011). "Text advertising blindness: the new banner blindness?" Journal of Usability Studies **6**(3): 172-197.

## - P -

- Peltonen, P., A. Salovaara, G. Jacucci, T. Ilmonen, C. Ardito, P. Saarikko and V. Batra (2007). Extending large-scale event participation with user-created mobile media on a public display. Proceedings of the 6th international conference on Mobile and ubiquitous multimedia, ACM.
- Perreira Da Silva, M., V. Courboulay, A. Prigent and P. Estraillier (2009). "Fast, low resource, head detection and tracking for interactive applications." PsychNology Journal **7**(3): 243-264.
- Peterson, E. R. and I. J. Deary (2006). "Examining wholistic–analytic style using preferences in early information processing." Personality and Individual Differences **41**(1): 3-14.
- Peterson, E. R., I. J. Deary and E. J. Austin (2003). "The reliability of Riding's cognitive style analysis test." Personality and individual differences **34**(5): 881-891.
- Peterson, E. R., I. J. Deary and E. J. Austin (2005). "Are intelligence and personality related to verbal-imagery and wholistic-analytic cognitive styles?" Personality and Individual Differences **39**(1): 201-213.
- Pietriga, E. and C. Appert (2008). Sigma Lenses: Focus-Context Transitions Combining Space, Time and Translucence. CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, ACM.
- Pittman, C. and J. J. LaViola Jr (2014). Exploring head tracked head mounted displays for first person robot teleoperation. Proceedings of the 19th international conference on Intelligent User Interfaces, ACM.
- Piumsomboon, T., D. Altimira, H. Kim, A. Clark, G. Lee and M. Billinghamurst (2014). Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on, IEEE.
- Poole, A. and L. J. Ball (2006). "Eye tracking in HCI and usability research." Encyclopedia of human computer interaction **1**: 211-219.
- Posner, M. I., M. J. Nissen and W. C. Ogden (1978). "Attended and unattended processing modes: The role of set for spatial location." Modes of perceiving and processing information **137**: 158.
- Pousman, Z. and J. Stasko (2006). A taxonomy of ambient information systems: four patterns of design. Proceedings of the working conference on Advanced visual interfaces, ACM.
- Prêteux, F. and M. Malciu (1998). Model-based head tracking and 3d pose estimation. Proceedings of SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision.

## - R -

- Ram, S. (1987). "A model of innovation resistance." ACR North American Advances.
- Ranjan, R., V. M. Patel and R. Chellappa (2016). "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition." arXiv preprint arXiv:1603.01249.

- Rasmussen, J. (1983). "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models." IEEE transactions on systems, man, and cybernetics(3): 257-266.
- Relieu, M., P. Salembier and J. Theureau (2004). "Introduction au numéro spécial «activité et action/cognition située»." Activités **1**(1-2).
- Riding, R. J. and E. Sadler-Smith (1997). "Cognitive style and learning strategies: Some implications for training design." International Journal of Training and Development **1**(3): 199-208.
- Robertson, G., M. Czerwinski, P. Baudisch, B. Meyers, D. Robbins, G. Smith and D. Tan (2005). "The large-display user experience." IEEE computer graphics and applications **25**(4): 44-51.
- Roche, A. (2015). Proposition d'une méthode de conception systémique d'interface homme-système adaptée aux situations de multihandicap, Université de Bordeaux.
- Rogers Everett, M. (1995). "Diffusion of innovations." New York **12**.
- Rogers, Y. and S. Lindley (2004). "Collaborating around vertical and horizontal large interactive displays: which way is best?" Interacting with Computers **16**(6): 1133-1152.
- Rogers, Y. and T. Rodden (2003). Configuring spaces and surfaces to support collaborative interactions. Public and Situated Displays, Springer: 45-79.
- Rogers, Y., H. Sharp and J. Preece (2011). Interaction design: beyond human-computer interaction, John Wiley & Sons.
- Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization in the brain." Psychological review **65**(6): 386.
- Rossi, S., E. Leone and M. Staffa (2016). Using Random Forests for the Estimation of Multiple Users' Visual Focus of Attention from Head Pose. AI\*IA 2016 Advances in Artificial Intelligence: XVth International Conference of the Italian Association for Artificial Intelligence, Genova, Italy, November 29 – December 1, 2016, Proceedings. G. Adorni, S. Cagnoni, M. Gori and M. Maratea. Cham, Springer International Publishing: 89-102.
- Roventa-Frumusani, D. (2015). "Manuel Maria CARRILHO (dir.)(2012), La rhétorique. Paris, CNRS Éditions, Coll.«Les Essentiels d'Hermès»." Communication. Information médias théories pratiques **33**(1).
- Russakoff, D. B. and M. Herman (2002). "Head tracking using stereo." Machine Vision and Applications **13**(3): 164-173.
- Russell, D. M. and A. Sue (2003). Large interactive public displays: Use patterns, support patterns, community patterns. Public and Situated Displays, Springer: 3-17.

## - S -

- Sadri, F. (2011). "Ambient intelligence: A survey." ACM Computing Surveys (CSUR) **43**(4): 36.
- Salas, E., C. Prince, D. P. Baker and L. Shrestha (1995). "Situation awareness in team performance: Implications for measurement and training." Human factors **37**(1): 123-136.
- Saner, L. D., C. A. Bolstad, C. Gonzalez and H. M. Cuevas (2009). "Measuring and predicting shared situation awareness in teams." Journal of Cognitive Engineering and Decision Making **3**(3): 280-308.

- Scheible, J. and T. Ojala (2005). MobiLenin combining a multi-track music video, personal mobile phones and a public display into multi-user interactive entertainment. Proceedings of the 13th annual ACM international conference on Multimedia, ACM.
- Schiavo, G., E. Mencarini, K. Vovard and M. Zancanaro (2013). Sensing and reacting to users' interest: an adaptive public display. CHI'13 Extended Abstracts on Human Factors in Computing Systems, ACM.
- Schneider, W. and R. M. Shiffrin (1977). "Controlled and automatic human information processing: I. Detection, search, and attention." Psychological review **84**(1): 1.
- Schnipke, S. K. and M. W. Todd (2000). Trials and tribulations of using an eye-tracking system. CHI'00 Extended Abstracts on Human Factors in Computing Systems, ACM.
- Schodl, A., A. Haro and I. A. Essa (1998). Head tracking using a textured polygonal model, Georgia Institute of Technology.
- Schwerdtner, A. (2008). Sweet Spot Unit For a Multi-User Display Device With an Expanded Viewing Zone, Google Patents.
- Shaer, O. and E. Hornecker (2010). "Tangible user interfaces: past, present, and future directions." Foundations and Trends in Human-Computer Interaction **3**(1–2): 1-137.
- Shen, C., K. Everitt and K. Ryall (2003). UbiTable: Impromptu face-to-face collaboration on horizontal interactive surfaces. UbiComp 2003: Ubiquitous Computing, Springer.
- Shen, C., F. D. Vernier, C. Forlines and M. Ringel (2004). DiamondSpin: an extensible toolkit for around-the-table interaction. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Shneiderman, B., C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist and N. Diakopoulos (2016). Designing the user interface: strategies for effective human-computer interaction, Pearson.
- Shotton, J., T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook and R. Moore (2013). "Real-time human pose recognition in parts from single depth images." Communications of the ACM **56**(1): 116-124.
- Simmons, T. (2001). "What's the optimum computer display size?" Ergonomics in Design **9**(4): 19-25.
- Somervell, J. P., S. Wahid and D. S. McCrickard (2003). Usability Heuristics for Large Screen Information Exhibits. INTERACT.
- Spector, R. H. (1990). Clinical Methods: The History, Physical, and Laboratory Examinations. Clinical Methods: The History, Physical, and Laboratory Examinations. H. K. Walker, W. D. Hall and J. W. Hurst. Boston, Butterworth Publishers, a division of Reed Publishing.
- Standard, U. M. (1999). "United States Military Standard: Department of Defense Design Criteria Standard." Mil-Std-1472F, Edited by Boff, K & Lincoln, J.
- Stanton, N., P. Salmon, G. Walker, E. Salas and P. Hancock (2017). "State-of-science: situation awareness in individuals, teams and systems." Ergonomics **60**(4): 449-466.
- Storz, O., A. Friday, N. Davies, J. Finney, C. Sas and J. Sheridan (2006). "Public ubiquitous computing systems: Lessons from the e-campus display deployments." IEEE Pervasive Computing **5**(3): 40-47.
- Suchman, L. (2007). Human-machine reconfigurations: Plans and situated actions, Cambridge University Press.
- Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication, Cambridge university press.
- Swaminathan, K. and S. Sato (1997). "Interaction design for large displays." interactions **4**(1): 15-24.

## - T -

- Tan, D. S. and M. Czerwinski (2003). Effects of visual separation and physical discontinuities when distributing information across multiple displays. Proc. Interact.
- Tan, D. S., D. Gergle, P. Scupelli and R. Pausch (2003). With similar visual angles, larger displays improve spatial performance. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM.
- Tan, D. S., J. K. Stefanucci, D. R. Proffitt and R. Pausch (2001). The Infocockpit: Providing location and place to aid human memory. Proceedings of the 2001 workshop on Perceptive user interfaces, ACM.
- Terrenghi, L. (2007). Designing hybrid interaction through an understanding of the affordances of physical and digital technologies. Creativity and Cognition: Proceedings of the 6 th ACM SIGCHI conference on Creativity & cognition.
- Terrenghi, L., A. Quigley and A. Dix (2009). "A taxonomy for and analysis of multi-person-display ecosystems." Personal and Ubiquitous Computing **13**(8): 583.
- Thüring, M. and S. Mahlke (2007). "Usability, aesthetics and emotions in human–technology interaction." International Journal of Psychology **42**(4): 253-264.
- Tilley, A. R. (2002). The measure of man and woman: human factors in design, John Wiley & Sons.
- Touzet, C. (1992). les réseaux de neurones artificiels, introduction au connexionnisme, EC2.
- Toyama, K. and E. Horvitz (2000). Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for head tracking, ACCV.
- Treisman, A. (1964). "Monitoring and storage of irrelevant messages in selective attention." Journal of Verbal Learning and Verbal Behavior **3**(6): 449-459.
- Trimble, J., R. Wales and R. Gossweiler (2003). NASA's MERBoard. Public and Situated Displays, Springer: 18-44.
- Turing, A. M. (1950). "Computing machinery and intelligence." Mind **59**(236): 433-460.
- Turner, J., J. Alexander, A. Bulling and H. Gellersen (2015). Gaze+ rst: Integrating gaze and multitouch for remote rotate-scale-translate tasks. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM.
- Tyndiuk, F. (2005). Référentiels Spatiaux des Tâches d'Interaction et Caractéristiques de l'Utilisateur influençant la Performance en Réalité Virtuelle, Université Victor Segalen-Bordeaux II.

## - U -

- Urey, H., K. V. Chellappan, E. Erden and P. Surman (2011). "State of the art in stereoscopic and autostereoscopic displays." Proceedings of the IEEE **99**(4): 540-555.

## - V -

- Vallerand, R. J. (1989). "Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française." Canadian Psychology/Psychologie Canadienne **30**(4): 662.
- Van der Zanden, A. (2016). "Readability in classrooms." Delft University of Technology. [http://homepage.tudelft.nl/9c41c/Readability\\_in\\_classrooms.pdf](http://homepage.tudelft.nl/9c41c/Readability_in_classrooms.pdf). Accessed December.

- Vidal, M., A. Bulling and H. Gellersen (2013). Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, ACM.
- Villame, T. (2004). "Conception de systèmes d'assistance au conducteur: comment prendre en compte le caractère complexe, dynamique et situé de la conduite automobile?. Cognition située et conception de systèmes d'assistance au conducteur." Activités 1(1-2).
- Vogel, D. and R. Balakrishnan (2004). Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. Proceedings of the 17th annual ACM symposium on User interface software and technology, ACM.
- Vogel, D. and R. Balakrishnan (2005). Distant freehand pointing and clicking on very large, high resolution displays. Proceedings of the 18th annual ACM symposium on User interface software and technology, ACM.

## - W -

- Walter, R., G. Bailly and J. Müller (2013). StrikeAPose: revealing mid-air gestures on public displays. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Wang, S., X. Xiong, Y. Xu, C. Wang, W. Zhang, X. Dai and D. Zhang (2006). Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM.
- Wang, X., K. Liu and X. Qian (2015). A survey on gaze estimation. Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on, IEEE.
- Ware, C. and H. H. Mikaelian (1987). An evaluation of an eye tracker as a device for computer input. ACM Sigchi Bulletin, ACM.
- Weiser, M. (1991). "The Computer for the 21 st Century." Scientific american 265(3): 94-105.
- Wellner, P. (1993). "Interacting with paper on the DigitalDesk." Communications of the ACM 36(7): 87-96.
- Whyte, W. H. (1980). The social life of small urban spaces.
- Wichary, M., L. Gunawan, N. Van den Ende, Q. Hjortzberg-Nordlund, A. Matysiak, R. Janssen and X. Sun (2005). Vista: interactive coffee-corner display. CHI'05 Extended Abstracts on Human Factors in Computing Systems, ACM.
- Wickens, C. D. (1991). "Processing resources and attention." Multiple-task performance 1991: 3-34.
- Wiener, N. (1949). Extrapolation, interpolation, and smoothing of stationary time series, MIT press Cambridge, MA.
- Wilmore, C. E. (2004). Interactive multi-user display arrangement for displaying goods and services, Google Patents.
- Witkin, H. A. (1965). "Psychological differentiation and forms of pathology." Journal of abnormal psychology 70(5): 317.

## - Y -

- Yan, C., Y. Wang and Z. Zhang (2011). "Robust real-time multi-user pupil detection and tracking under various illumination and large-scale head motion." Computer Vision and Image Understanding 115(8): 1223-1238.

- Yan, Y., E. Ricci, R. Subramanian, G. Liu, O. Lanz and N. Sebe (2016). "A multi-task learning framework for head pose estimation under target motion." IEEE transactions on pattern analysis and machine intelligence **38**(6): 1070-1083.
- Yang, R. and Z. Zhang (2002). Model-based head pose tracking with stereovision. Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, IEEE.
- Yerkes, R. M. and J. D. Dodson (1908). "The relation of strength of stimulus to rapidity of habit-formation." Journal of comparative neurology **18**(5): 459-482.

- Z -

- Zhai, S., C. Morimoto and S. Ihde (1999). Manual and gaze input cascaded (MAGIC) pointing. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM.
- Zhang, X. and I. S. MacKenzie (2007). Evaluating eye tracking with ISO 9241-part 9. International Conference on Human-Computer Interaction, Springer.
- Zhao, W., R. Chellappa, P. J. Phillips and A. Rosenfeld (2003). "Face recognition: A literature survey." ACM computing surveys (CSUR) **35**(4): 399-458.
- Zhong, X., J. Xue and N. Zheng (2006). Graphical Model based Cue Integration Strategy for Head Tracking. BMVC.
- Zivkovic, Z. and F. Van der Heijden (2001). A stabilized adaptive appearance changes model for 3D head tracking. Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on, IEEE.

# Annexes

# Annexe I : Formulaire de Consentement libre, éclairé et exprès (première expérimentation)

*Thèse : ShaBiDi : Développement des technologies et métaphores d'interactions nouvelles pour le partage collaboratif d'affichage en groupe ouvert.*

## ShaBiDi

Equipe CIH - "Cognitive et Ingénierie Humaine" - IMS CNRS 5218  
ENSC - Ecole Nationale Supérieure de Cognitive - 109 Avenue Roul, 33400 Talence

Je certifie avoir donné mon accord pour participer à l'expérience **ShaBiDi**. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données sur mes réponses. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité de recherche. J'ai bien noté que le droit d'accès prévu par la loi « informatique et libertés » s'exerce à tout moment auprès de l'unité de recherche.

De plus, j'accepte que les organisateurs de la recherche utilisent et diffusent à titre gratuit et non exclusif mon image par des photographies et vidéos. Ces dernières pourront être exploitées et utilisées directement par les organisateurs, sous toutes formes et tous supports connus et inconnus à ce jour, dans le monde entier, sans limitation de durée, intégralement ou par extraits lors :

- de publications dans une revue, ouvrage ou journal scientifique portant sur l'objet de la recherche
- de diffusion interne et externe au sein de l'ENSC, Ecole Nationale Supérieure de Cognitive.

Les bénéficiaires de l'autorisation s'interdisent expressément de procéder à une exploitation des images susceptible de porter atteinte à la vie privée ou à la réputation, ni de les utiliser, dans tout support à caractère pornographique, raciste, xénophobe ou toute autre exploitation préjudiciable.

Je me reconnais être entièrement rempli de mes droits et je ne pourrai prétendre à aucune rémunération pour l'exploitation des droits visés aux présentes.

Date : .....

Nom du volontaire : .....

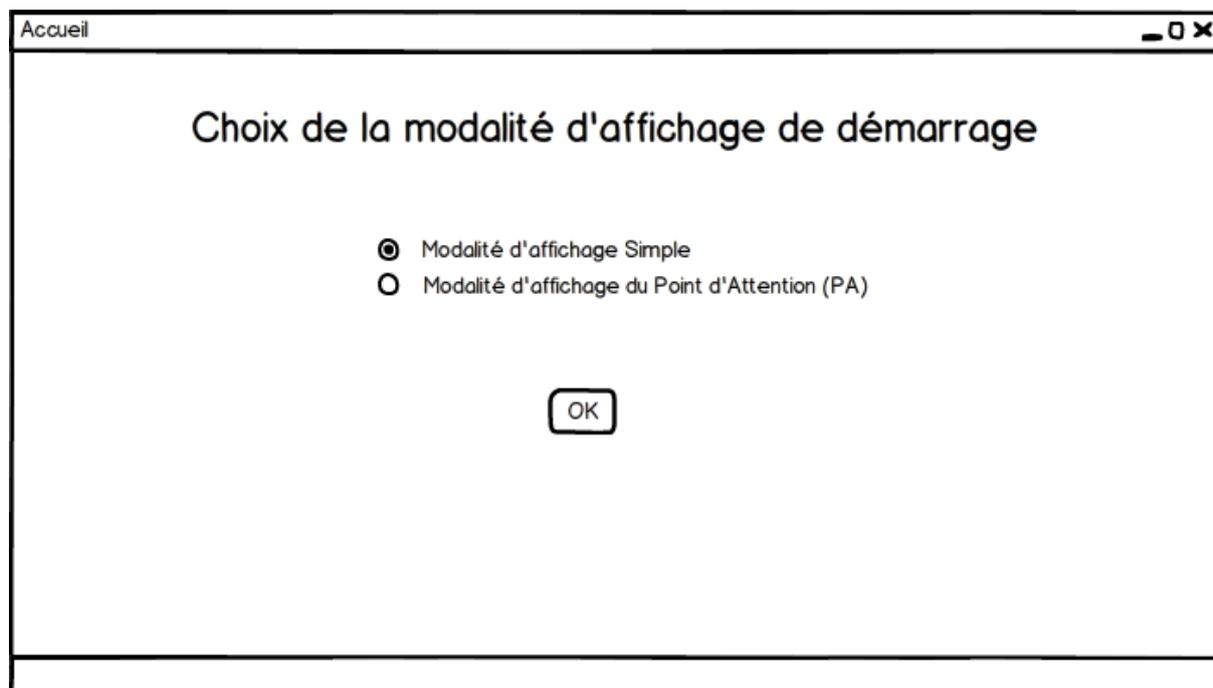
Signature du volontaire (précédée de la mention « lu et approuvé ») :

Nom de l'expérimentateur: .....

Signature de l'expérimentateur :

## Annexe II : Spécification fonctionnelle du logiciel de passation (première expérience)

### Ecran N°1 :



RG1E1 : Au lancement du programme la fenêtre ci-dessus s'ouvre.

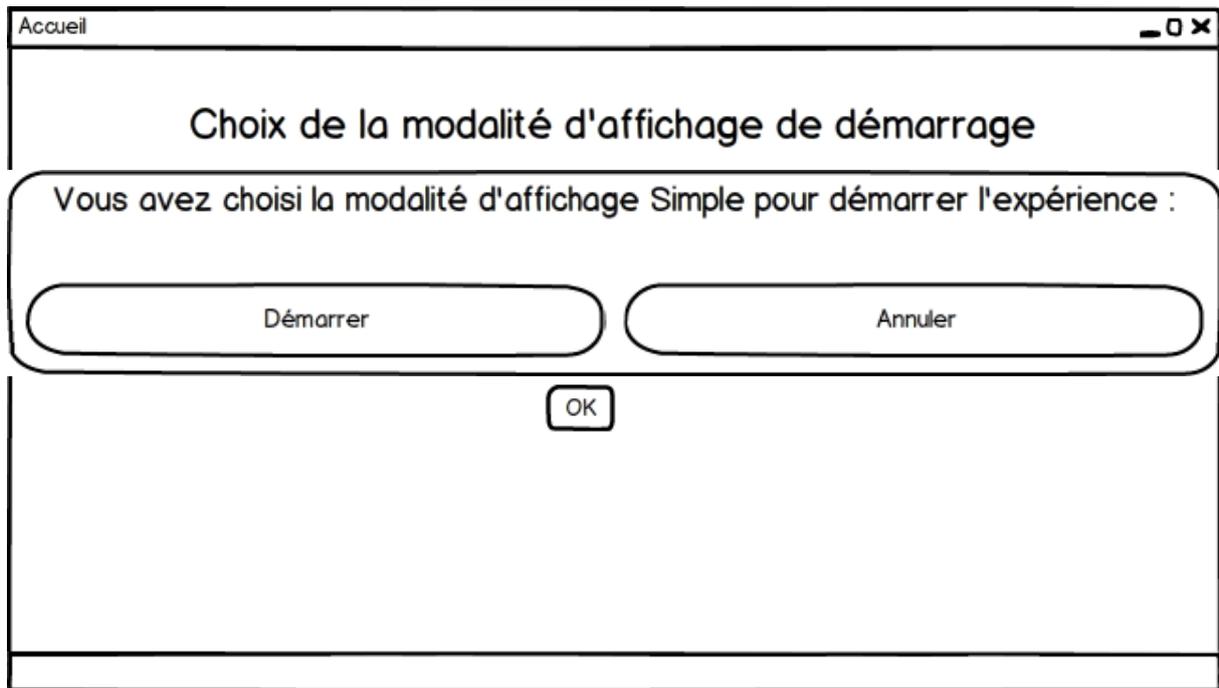
RG2E1 : La fenêtre s'ouvre avec la propriété « Maximized ».

RG3E1 : Par défaut la « Modalité d'affichage Simple » est cochée.

RG4E1 : Au clic sur le bouton « OK », ouverture de la Pop-Up :

- [Ecran N°2A](#), si Radio bouton « Modalité d'affichage Simple » est coché
- [Ecran N°2B](#), si Radio bouton « Modalité d'affichage du Point d'Attention (PA) » est coché

## Ecran N°2A :



RG1E2A : Au clic sur le bouton « Démarrer » affichage de l'écran : [Ecran N°3](#)

RG2E2A : Au clic sur le bouton « Annuler » affichage de l'écran : [Ecran N°1](#)

## Ecran N°2B :

Accueil

### Choix de la modalité d'affichage de démarrage

Vous avez choisi la modalité d'affichage du PA pour démarrer l'expérience :

Démarrer Annuler

RG1E2B : Au clic sur le bouton « Démarrer » affichage de l'écran : [Ecran N°3](#)

RG2E2B : Au clic sur le bouton « Annuler » affichage de l'écran : [Ecran N°1](#)

### Ecran N°3 :

Entrer le numéro du groupe :

Groupe N° 3

Démarrer Annuler

RG1E3 : Au clic sur le bouton « Démarrer » affichage de l'écran : [Ecran N°4](#)

RG2E3 : Au clic sur le bouton « Annuler » affichage de l'écran : [Ecran N°1](#)

RG3E3 : Le « Numeric Stepper » :

- Peut être renseigné seulement avec des caractères numériques par l'utilisateur.
- On ne peut pas effacer le texte à l'intérieur, seulement le remplacer en tapant un chiffre ou un nombre.
- Au clic sur la flèche vers le haut ▲, on incrémente le chiffre affiché.
- Au clic sur la flèche vers le bas ▼, on décrémente le chiffre affiché.

## Ecran N°4 :

\_ 0 x

**Consignes :** Groupe n°3

A partir de maintenant, vous ne devez en aucun cas interagir entre vous peu importe la manière (paroles, gestes, regards, ect ...).  
Vous pouvez poser toutes les questions que vous voulez mais seulement à l'expérimentateur et ceci seulement jusqu'au démarrage de l'expérience. Après le démarrage de l'expérience, indiquez par « Début de l'expérience », les questions sont à barrir.

Équipez-vous des casquettes.  
Nous allons vous présenter deux séries d'une quinzaine d'images dans lesquelles vous devez retrouver des éléments. En d'autre terme vous aller jouer à un jeu type « Où est Charlie ».

Le but est de trouver le plus d'éléments possibles en 40 secondes. Les éléments à retrouver seront placés en haut de l'écran.  
Plus vous trouvez les éléments rapidement plus votre score sera élevé.

La souris placées devant vous, vous permettra de cliquer sur l'élément trouvé dans l'image.  
Plus vous cliquez sur des éléments de l'image qui ne sont pas à trouver, plus vous perdez de points.

Pour une des deux séries, vous observerez sur l'écran la matérialisation de votre point d'attention ainsi que celui de vos partenaires. Vous serez prévenus avant le passage de la modalité d'affichage sans les points d'attention affichés à celle où ils sont affichés.

Nous allons maintenant procédez à une présentation en situation, celle-ci ne comptera pas dans le décompte de votre score final.  
Vous commencerez avec la modalité d'affichage sans la matérialisation du point d'attention sur la première image.  
Vous avez 40 secondes pour trouver le plus d'éléments possibles.

Une fois, les 40 secondes écoulées, un écran avec un compte à rebours de 5 secondes s'affichera.  
Puis s'affichera une seconde image avec vos points d'attention matérialisés.  
Vous aurez à nouveau 40 secondes pour trouver tous les éléments.

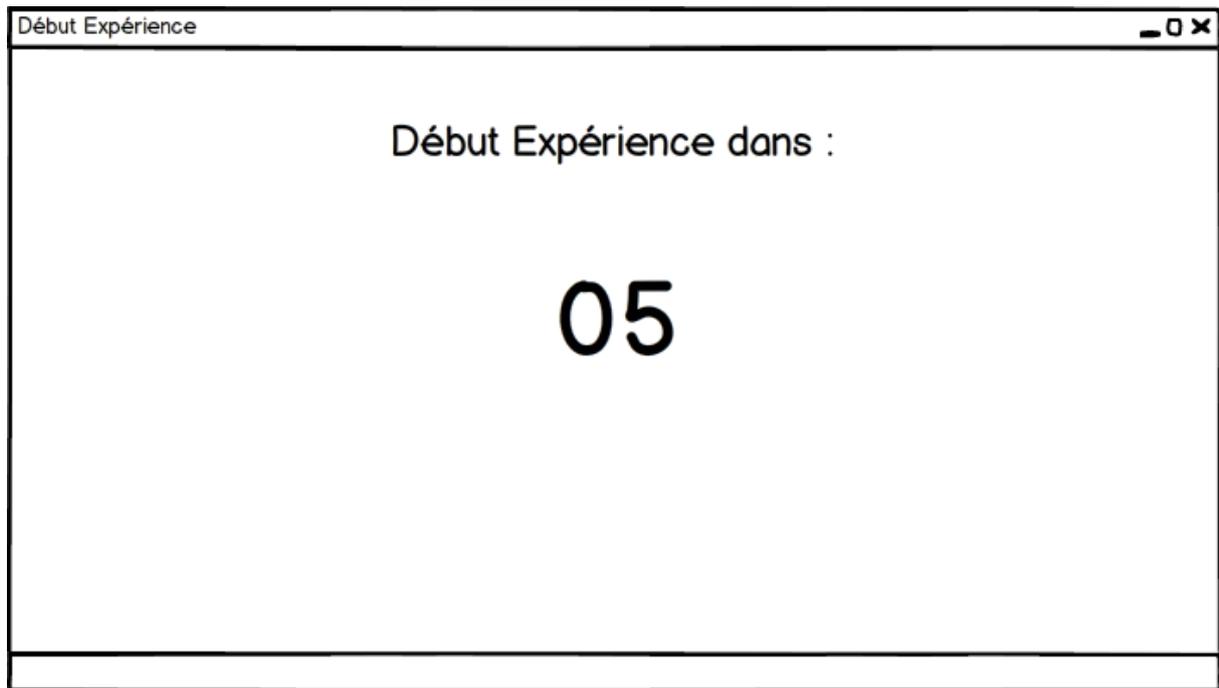
Une fois les 40 secondes écoulées, vous aurez effectué la présentation test, l'écran "Début de l'expérience" s'affiche.

Etes-vous prêt ? **Commencer**

RG1E4 : Au clic sur le bouton « Commencer » affichage de l'écran : [Ecran N°5](#)

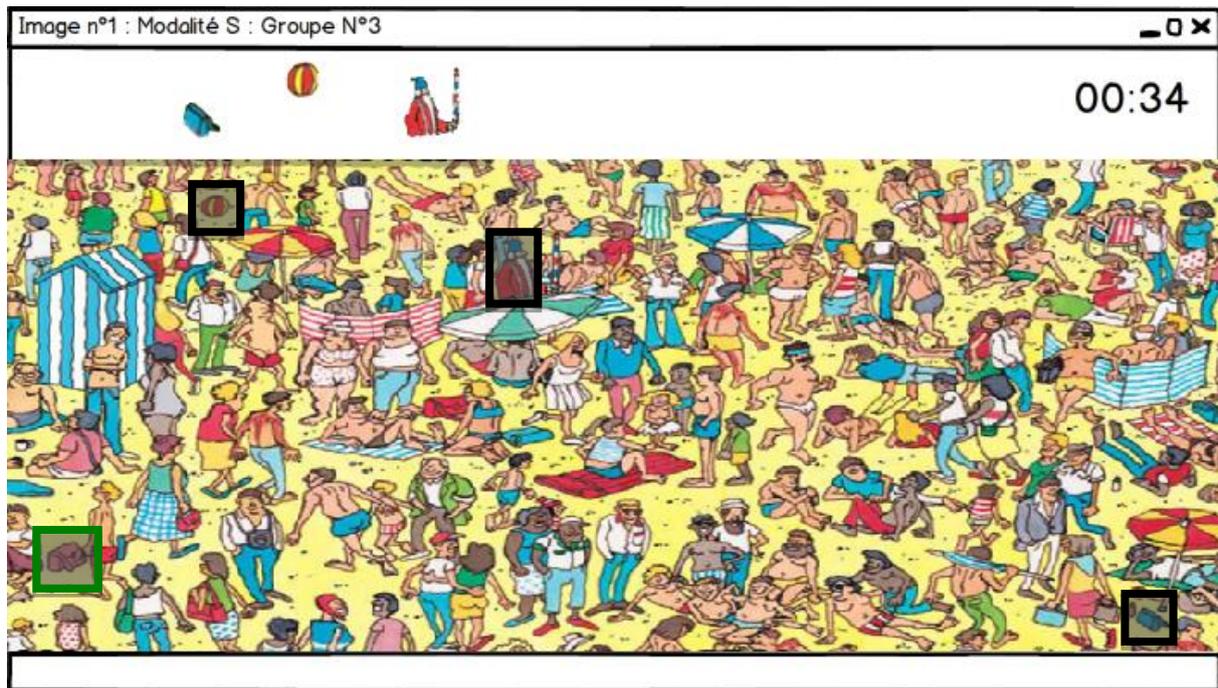
RG2E4 : Au clic sur le lien « Groupe n°3 » affichage de l'écran : [Ecran N°3](#)

## Ecran N°5 :



RG1E5 : Le chiffre « 05 » représente un compte à rebours démarrant à 5, à la fin de celui-ci affichage de l'écran : [Ecran N°6](#)

## Ecran N°6 :



RG1E6 : En haut de la fenêtre apparaissent le n° de l'image, la modalité d'affichage utilisée et le n° du groupe.

RG2E6 : En haut à droite, le compte à rebours indiquant le temps restant avant le passage à une autre image est affiché.

RG3E6 : La liste des éléments à trouver contient un élément non présent dans l'image.

RG4E6 : La liste des éléments à trouver est disposée en nuage.

RG5E7 :  : ce sont les zones cliquables invisibles pour l'utilisateur, qui définissent les éléments à trouver.

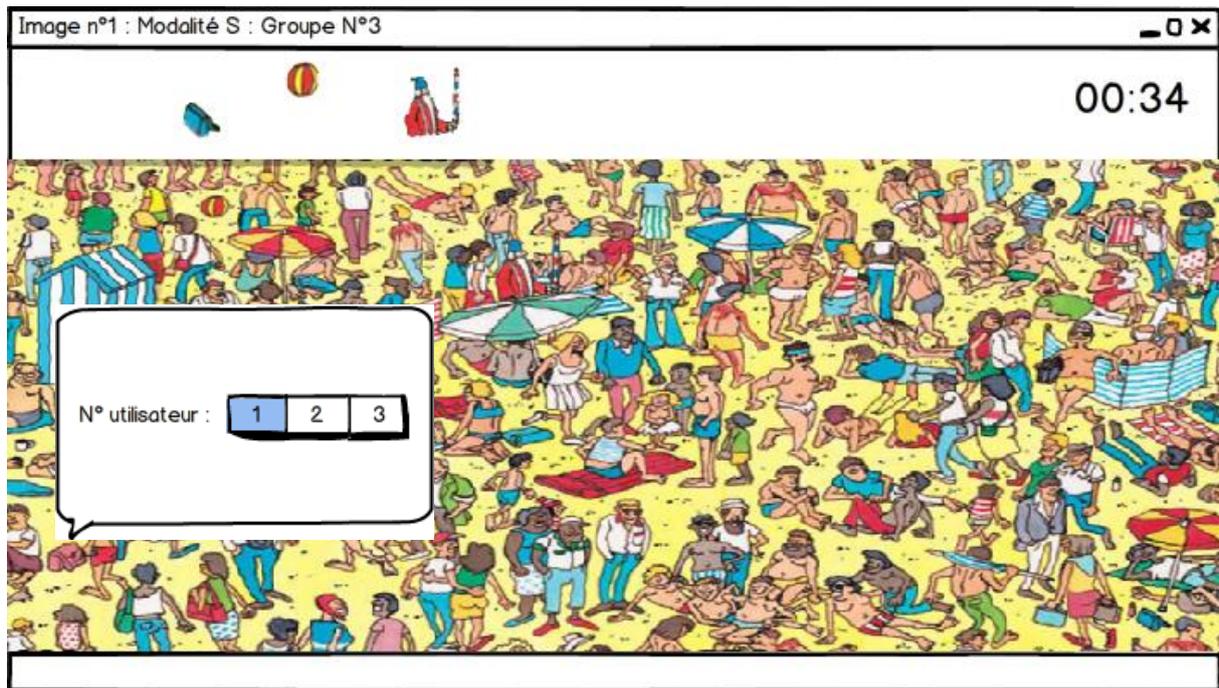
Si un utilisateur clique :

- sur une partie de l'image ne correspondant pas à ces zones cliquables, il ne passe rien.
- sur une de ces zones cliquables, l'élément correspondant à la zone cliquée disparaît de la liste des éléments à trouver et l'élément trouvé est entouré sur l'image. La zone cliquable est alors désactivée et une Pop-Up apparaît sur l'écran de l'expérimentateur afin de renseigner le n° de l'utilisateur qui a identifié l'élément trouvé (voir [Ecran N°7](#))

Exemple : ci-dessus l'utilisateur a cliqué sur le sac rose :



## Ecran N°7 :

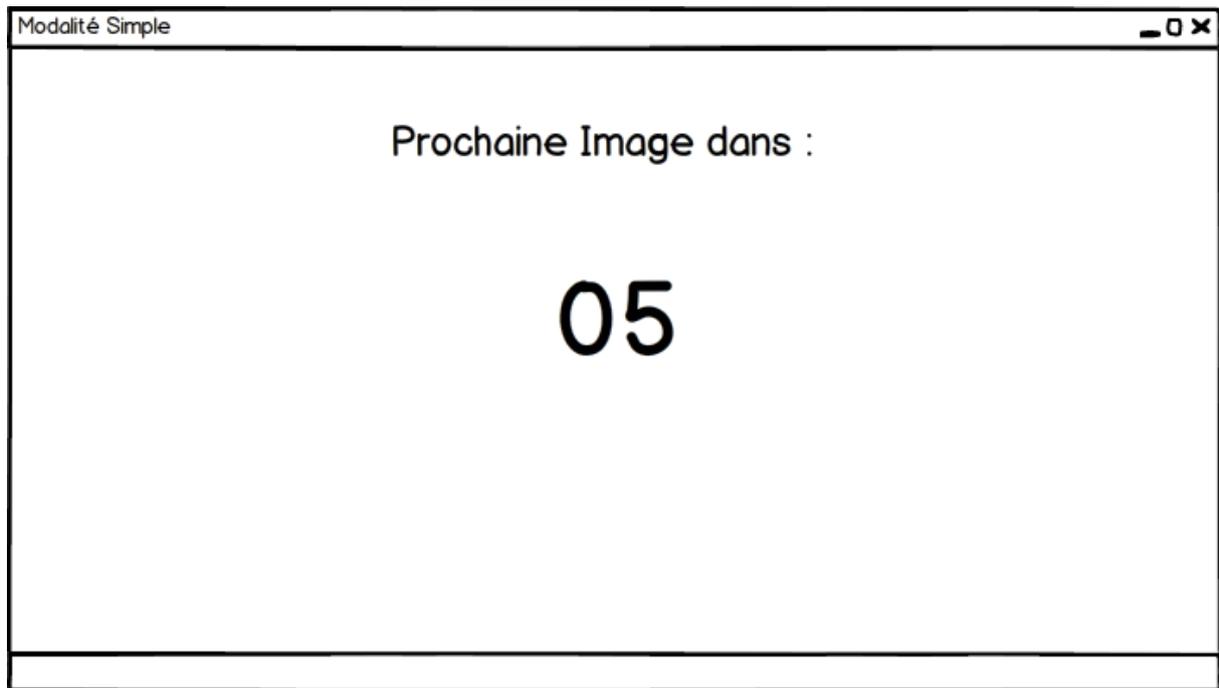


RG1E7 : l'expérimentateur clique sur la zone 1, 2 ou 3 dans la Pop-Up correspondant au numéro de l'utilisateur ayant trouvé l'élément. La pop-up disparaît.

RG2E7 : En cas d'erreur sur le N° de l'utilisateur, l'examineur peut appuyer sur la touche « Espace » du clavier, la dernière Pop-up est alors réaffichée, affichage écran :

[Ecran N°7](#)

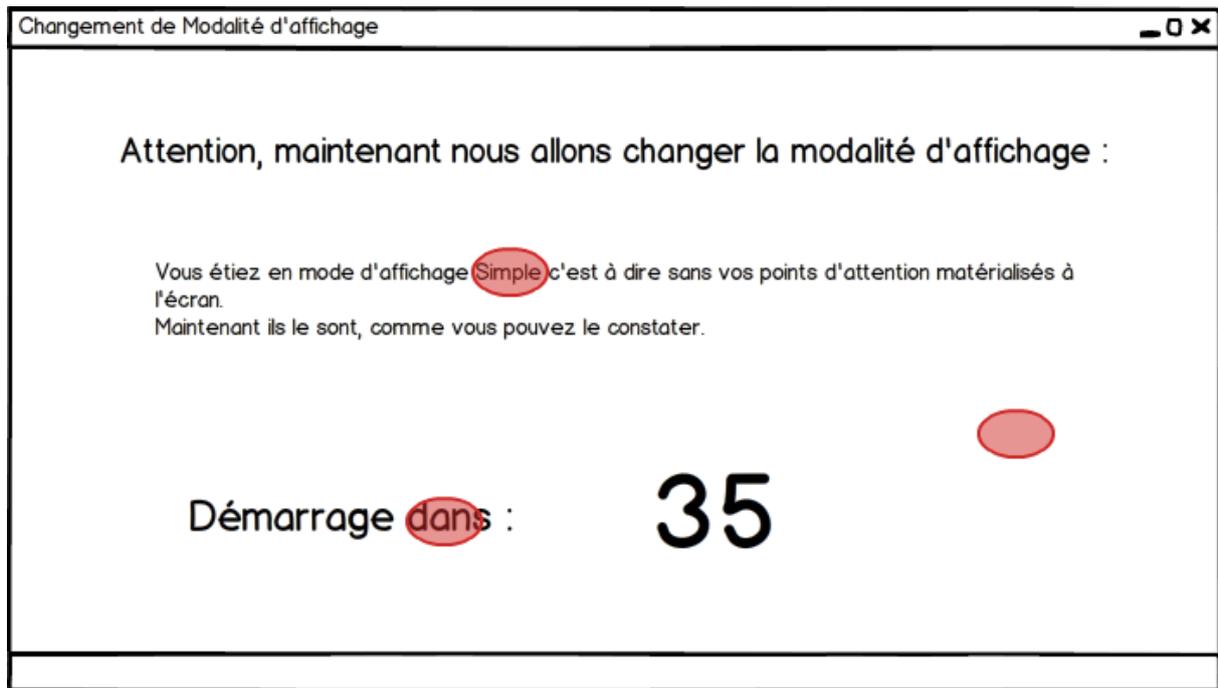
## Ecran N°8 :



RG1E8 : En haut à gauche de l'écran est affiché le nom de la modalité utilisée pour l'image suivante

RG2E8 : Au milieu de l'écran est affiché un compte à rebours de 5 secondes. A la fin de celui-ci affichage de l'écran : [Ecran N°6](#)

## Ecran N°9 :



RG1E9 : Cet écran apparaît après une série de 15 images pour annoncer le changement de modalité :

- Si la modalité d'affichage précédente était la modalité Simple alors on obtient l'écran ci-dessus.  
Les cercles rouges représentent les PA des utilisateurs.
- Si la modalité d'affichage précédente était la modalité avec matérialisation du PA, les PA disparaissent et le texte central devient celui-ci :  
« Vous étiez en mode d'affichage avec matérialisation de vos points d'attention sur l'écran.  
Maintenant, ils ne sont plus affichés comme vous pouvez le constater. »

RG2E9 : Le compte à rebours dure 40 secondes puis affichage de l'écran : [Ecran N°6](#)

## Ecran N°10 :



RG1E10 : Au clic sur le Bouton « Affichage des Scores » affichage de l'écran : Ecran N°11

## Ecran N°11 :

The screenshot shows a window titled "Fin de l'expérimentation" with a standard Windows-style title bar (minimize, maximize, close buttons). The main content area is titled "Tableau des scores" and contains a table with three columns: "Position", "N° Groupe", and "Score". The table has three rows of data. Below the table is a button labeled "Terminer".

Position	N° Groupe	Score
1	Groupe N°3	100000
2	Groupe N°1	80000
3	Groupe N°2	15000

Terminer

RG1E11 : Au clic sur le bouton « Terminer », fermeture de l'application.

## Autres Fonctionnalités

Le score du groupe :

Le score du groupe est calculé de la façon suivante :

- Quand un membre du groupe clique sur une zone de l'image n'étant pas une zone cliquable et donc par extension un élément à trouver, il ne se passe rien (cependant, l'examineur aura volontairement annoncé en début d'expérience que si ce cas se produisait, le groupe perdrait des points).
- Quand un membre du groupe clique sur un élément à trouver dans l'image, il obtient le score suivant :

$$\text{Nouveau Score du groupe} = \text{Score du groupe} + \text{Temps restant} * 1000$$

## Annexe III : Consignes affichées à l'écran (première expérience)

A partir de maintenant, vous ne devez en aucun cas interagir entre vous peu importe la manière (paroles, gestes, regards, *etc.*).

Vous pouvez poser toutes les questions que vous voulez mais seulement à l'expérimentateur et ceci seulement jusqu'au démarrage de l'expérience.

Après le lancement de l'expérience, indiqué par « Début de l'expérience », les questions sont à bannir.

Nous allons vous présenter deux séries de 15 images dans lesquelles vous devez retrouver des éléments. En d'autres termes, vous allez jouer à un jeu type « Où est Charlie ».

Le but est de trouver le plus d'éléments possibles en 40 secondes. Les éléments à retrouver seront placés en haut de l'écran.

Plus vous trouvez les éléments rapidement plus le score de votre équipe sera élevé.

La souris placée devant vous, vous permettra de cliquer sur l'élément trouvé dans l'image.

Plus vous cliquez sur des éléments de l'image qui ne sont pas à trouver, plus votre équipe perd des points.

Pour une des deux séries, vous observerez sur l'écran la matérialisation de votre point d'attention ainsi que ceux de vos partenaires. Vous serez avertis des changements de modalité d'affichage.

Nous allons maintenant procéder à une présentation en situation, celle-ci ne comptera pas dans le décompte de votre score final.

Vous avez 40 secondes pour trouver le maximum d'éléments possibles.

Une fois les 40 secondes écoulées, un écran avec un compte à rebours de 3 secondes s'affichera indiquant le passage à une nouvelle image.

Vous aurez à nouveau 40 secondes pour trouver un maximum d'éléments. Une fois ce temps écoulé, vous aurez effectué la présentation test. L'écran affichera alors « Début de l'expérience ».

A partir de ce moment, le score de l'équipe sera comptabilisé.

Etes-vous prêt ?

# Annexe IV : Questionnaire *System Usability Scale* (SUS)

	Pas du tout d'accord						Tout à fait d'accord
1. Je pense que je vais utiliser ce service fréquemment	<input type="checkbox"/>						
	1	2	3	4	5		
2. Je trouve ce service inutilement complexe	<input type="checkbox"/>						
	1	2	3	4	5		
3. Je pense que ce service est facile à utiliser	<input type="checkbox"/>						
	1	2	3	4	5		
4. Je pense que j'aurai besoin de l'aide d'un technicien pour être capable d'utiliser ce service	<input type="checkbox"/>						
	1	2	3	4	5		
5. J'ai trouvé que les différentes fonctions de ce service ont été bien intégrées	<input type="checkbox"/>						
	1	2	3	4	5		
6. Je pense qu'il y a trop d'incohérence dans ce service.	<input type="checkbox"/>						
	1	2	3	4	5		
7. J'imagine que la plupart des gens seraient capable d'apprendre à utiliser ce service très rapidement.	<input type="checkbox"/>						
	1	2	3	4	5		
8 J'ai trouvé ce service très lourd à utiliser.	<input type="checkbox"/>						
	1	2	3	4	5		
9. Je me sentais très en confiance en utilisant ce service.	<input type="checkbox"/>						
	1	2	3	4	5		
10. J'ai besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce service.	<input type="checkbox"/>						
	1	2	3	4	5		

## Annexe V : Questionnaire *AttrakDiff*

### Consignes Questionnaire

Dans le cadre de ma thèse sur l'expérience utilisateur, nous souhaiterions évaluer vos impressions sur l'interface/service utilisé(e).

Ce questionnaire se présente sous forme de paires de mots pour vous assister dans l'évaluation du système. Chaque paire représente des contrastes. Les échelons entre les deux extrémités vous permettent de décrire l'intensité de la qualité choisie.

Répondez de manière spontanée même si vous avez l'impression que certains termes ne décrivent pas correctement l'interface/service utilisé(e). Dans ce cas, assurez-vous de donner tout de même une réponse. Gardez à l'esprit qu'il n'y a pas de bonne ou mauvaise réponse.

Seul votre avis compte !

Humain	<input type="radio"/>	Technique						
M'isole	<input type="radio"/>	Me sociabilise						
Plaisant	<input type="radio"/>	Déplaisant						
Original	<input type="radio"/>	Conventionnel						
Simple	<input type="radio"/>	Compliqué						
Professionnel	<input type="radio"/>	Amateur						
Laid	<input type="radio"/>	Beau						
Pratique	<input type="radio"/>	Pas pratique						
Agréable	<input type="radio"/>	Désagréable						
Fastidieux	<input type="radio"/>	Efficace						
De bon goût	<input type="radio"/>	De mauvais goût						
Prévisible	<input type="radio"/>	Imprévisible						
Bas de gamme	<input type="radio"/>	Haut de gamme						
M'exclut	<input type="radio"/>	M'intègre						
Me rapproche des autres	<input type="radio"/>	Me sépare des autres						
Non présentable	<input type="radio"/>	Présentable						
Rebutant	<input type="radio"/>	Attirant						
Sans imagination	<input type="radio"/>	Créatif						
Bon	<input type="radio"/>	Mauvais						
Confus	<input type="radio"/>	Clair						
Repoussant	<input type="radio"/>	Attrayant						
Audacieux	<input type="radio"/>	Prudent						
Novateur	<input type="radio"/>	Conservateur						
Ennuyeux	<input type="radio"/>	Captivant						
Peu exigeant	<input type="radio"/>	Challenging						
Motivant	<input type="radio"/>	Décourageant						
Nouveau	<input type="radio"/>	Commun						
Incontrôlable	<input type="radio"/>	Maîtrisable						

# Annexe VI : Formulaire de Consentement Libre, Eclairé et Exprès (deuxième et troisième expérience)

*Thèse : ShaBiDi : Développement des technologies et métaphores d'interactions nouvelles pour le partage collaboratif d'affichage en groupe ouvert.*

## ShaBiDi 2

Equipe CIH - "Cognitique et Ingénierie Humaine" - IMS CNRS 5218  
ENSC - Ecole Nationale Supérieure de Cognitique - 109 Avenue Roul, 33400 Talence

Je certifie avoir donné mon accord pour participer à l'expérience **ShaBiDi 2**. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données sur mes réponses. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité de recherche. J'ai bien noté que le droit d'accès prévu par la loi « informatique et libertés » s'exerce à tout moment auprès de l'unité de recherche.

De plus, j'accepte que les organisateurs de la recherche utilisent et diffusent à titre gratuit et non exclusif mon image par des photographies et vidéos. Ces dernières pourront être exploitées et utilisées directement par les organisateurs, sous toutes formes et tous supports connus et inconnus à ce jour, dans le monde entier, sans limitation de durée, intégralement ou par extraits lors :

- de publications dans une revue, ouvrage ou journal scientifique portant sur l'objet de la recherche
- de diffusion interne et externe au sein de l'ENSC, Ecole Nationale Supérieure de Cognitique.

Les bénéficiaires de l'autorisation s'interdisent expressément de procéder à une exploitation des images susceptible de porter atteinte à la vie privée ou à la réputation, ni de les utiliser, dans tout support à caractère pornographique, raciste, xénophobe ou toute autre exploitation préjudiciable.

Je me reconnais être entièrement rempli de mes droits et je ne pourrai prétendre à aucune rémunération pour l'exploitation des droits visés aux présentes.

Date : .....

Nom du volontaire : .....

Signature du volontaire (précédée de la mention « lu et approuvé ») :

Nom de l'expérimentateur: .....

Signature de l'expérimentateur :

## Annexe VII : Fiche mission

Combien il y a de caractères « T » au total sur toutes les affiches ?

---

Combien il y a de caractères « S » au total dans l'ensemble des affiches de type B ?

---

Combien il y a de caractères « R » au total dans l'ensemble des affiches de type C ?

---

Combien il y a de caractères « V » au total dans l'ensemble des affiches de type A ?

## Annexe VIII : Questionnaire Global

Quelle interface avez-vous préféré ?

Interface D						Interface S
-------------	--	--	--	--	--	-------------

Avec quelle interface vous êtes-vous senti(e) le/la plus rapide ?

Interface D						Interface S
-------------	--	--	--	--	--	-------------

Avec quelle interface pensez-vous avoir fait le moins d'erreur dans vos missions ?

Interface D						Interface S
-------------	--	--	--	--	--	-------------

Si l'interface D était présente dans des lieux publics, l'utiliserez-vous ?

Pas d'accord						Tout à fait d'accord
--------------	--	--	--	--	--	----------------------

Si vous deviez changer ou améliorer des interactions sur l'interface D en vue de l'utiliser dans des lieux publics, quelles seraient-elles ?

--

Selon vous l'interface D, pourrait servir dans quel(s) cas d'usage ?

--