



**HAL**  
open science

## 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

Damien Marion, Pierre-Alexandre Favier, Jean-Marc André

### ► To cite this version:

Damien Marion, Pierre-Alexandre Favier, Jean-Marc André. 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, AFIHM, Aug 2017, Poitiers, France. 8 p., 10.1145/3132129.3132141 . hal-01578622

**HAL Id: hal-01578622**

**<https://hal.science/hal-01578622>**

Submitted on 29 Aug 2017

**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.

# 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.

## *Head pose characterization with Face Tracking during information consultation on a large display.*

Damien Marion

Univ. Bordeaux, CNRS, IMS, UMR 5218  
33405 Talence

France

damien.marion@ensc.fr

Pierre-Alexandre Favier

Univ. Bordeaux, CNRS, IMS, UMR 5218  
33405 Talence

France

pierre-alexandre.favier@ensc.fr

Jean-Marc André

Univ. Bordeaux, CNRS, IMS, UMR 5218  
33405 Talence

France

jean-marc.andre@ensc.fr



Figure 1: Photo de l'installation

### ABSTRACT

Wherever you look at, you can see a screen. Large displays are increasingly present in our life. So, we design more and more innovative interaction. This article presents the characterization of human average head pose, with a Head Tracking system based on Kinect V2. Results revealed that we target information at a given distance in the same way. This characterization allows us to design new natural intuitive interaction for large display without individual calibration.

### CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction device; HCI design and evaluation methods* • **Human-centered computing** → Interaction design

### KEYWORDS

Human-Computer Interaction, Head Pose, Head Tracking, Large display, Kinect V2.

### RÉSUMÉ

L'omniprésence des écrans dans nos vies, et plus particulièrement ceux de grand format, nous amène à constamment inventer de nouvelles métaphores d'interactions. Dans cet article, nous présentons 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 dans le but de lire une information à une distance donnée. Cette caractérisation des mouvements de tête via la Kinect V2, nous a permis de développer un système d'*head tracking* permettant une interaction intuitive, sans appareillage ni calibrage individuel. Ce travail permet, pour plusieurs utilisateurs simultanément, la consultation d'informations individualisées sur un affichage de grand format.

### MOTS-CLEFS

Interaction Homme-Machine, Orientation de la tête, Head Tracking, Ecran de grand format, Kinect V2.

---

© ACM, 2017. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 29ème conférence francophone sur l'Interaction Homme- Machine, IHM'17, August 28–September 1, 2017, Poitiers, France  
<https://doi.org/10.1145/3132129.3132141>

## 1 INTRODUCTION

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 [1]. 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 une expérience antérieure [2], nous utilisons la technologie Optitrack afin de mesurer avec précision les mouvements de la tête. Grâce à ces mesures nous extrapolons 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.

L'avènement récent de capteurs de profondeur [3] à bas prix, comme la Kinect V2, permet l'accès, de manière relativement simple et précise [4], 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 nos travaux 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 multiutilisateurs lors de la consultation d'un affichage de grand format en milieu ouvert.

De nombreuses études ont été menées sur les différents types de capteurs et méthodes [5-11] permettant d'estimer l'orientation de la tête des utilisateurs [12, 13]. Selon nous, de façon idéale, le système proposé devrait permettre :

- Une détection précise, proche de 1° ;
- Une détection non intrusive/invasive ;
- Une détection multiutilisateur ;
- Une détection précise sans calibrage préalable ;
- Un système à bas prix.

Dans notre contexte d'étude, 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° [14]. 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° va permettre d'englober plus d'information 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 la Kinect V2 et son SDK pour réaliser le *head tracking*.

L'objectif est double :

- Dans un premier temps, qualifier la précision et les erreurs de détection issues de la Kinect V2. Cet objectif se rapproche de celui de l'étude [15].
- 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.

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 que nous mettons en évidence dans cet article. Par exemple, une information affichée à 15° sur la droite et 10° vers le haut par rapport à une position neutre<sup>1</sup> de la tête de l'utilisateur ne nécessitera qu'une rotation moyenne de la tête de 12° vers la droite et de 5° 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 système de système masse-ressort [16, 17]. 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. Mais nous essayons ici de qualifier une plage de valeurs moyennes de ces rotations de la tête permettant de caractériser au plus près le plus grand nombre d'individus. En l'état actuel, la Kinect V2 nous permet de calculer la zone de focalisation sur l'écran de six individus simultanément en se basant sur la détection de l'orientation de la tête. Cependant, les calculs basés sur les données brutes issues de la 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.

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.

<sup>1</sup> Dans notre cas, la position dite neutre de la tête correspond à une rotation de la tête :

- dans le plan frontal, appelée rotation de tangage (rotation droite-gauche) de 0° ;
- dans le plan sagittal, appelée rotation de lacet (rotation haut-bas) de 0° ;
- dans le plan transverse ou horizontal, appelée rotation de roulis de 0°.

## 2 PROTOCOLE EXPÉRIMENTAL

### 2.1 Matériel

Nous utilisons la Kinect V2 qui est une caméra infrarouge et la version 2.0.1410.19000 du SDK. Celle-ci est placée 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° vers le bas par rapport au sol. Cette position et cette inclinaison ont été choisies afin de maximiser la couverture de la zone d'interaction possible devant l'écran.

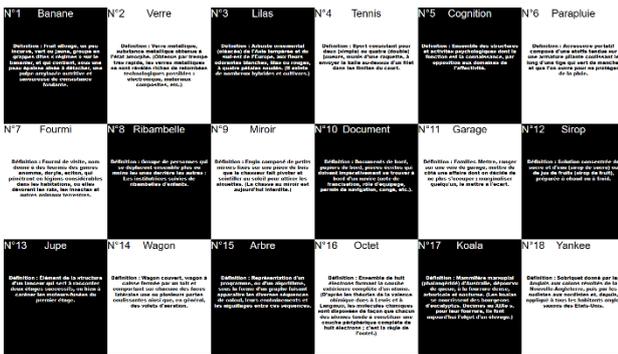


Figure 2: Ensemble des affiches présentes à l'écran.

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 est 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.

### 2.2 Méthode

2.2.1 *Détails computationnels.* Techniquement, le SDK donne accès aux coordonnées 3D d'une multitude de points 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 Fig. 3.

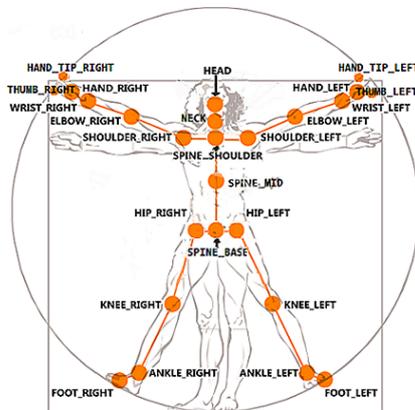


Figure 3: Body Joints issus du SDK de la Kinect V2.

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 possibles en degrés de la tête de l'utilisateur dans le système de coordonnées de la 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.

2.2.2 *Milieu d'implémentation.* Une zone d'interaction de 2,81 mètres de largeur par 2,02 mètres de longueur est matérialisée au sol à 1,60 mètre devant l'écran. Cette zone est divisée en 18 parties identiques, 6 sur la longueur répartie sur 3 lignes comme le montre la Fig. 1. La zone d'affichage sur l'écran est divisée de la même manière comme le montre la Fig. 2. La taille de la zone au sol a été déterminée à l'aide de tests de détection d'head tracking pour maximiser la qualité de détection. Le découpage en 6 zones sur la longueur a été établi par les contraintes de détection du nombre d'utilisateurs de la Kinect V2, soit 6 utilisateurs maximum.

2.2.3 *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 présentes 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 automatiques et ainsi biaiser nos données. Les affiches sont donc volontairement données 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 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 Fig. 4. Ces trois positions correspondent à 3 distances différentes par rapport à l'écran.

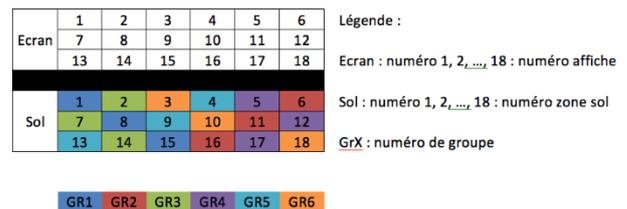


Figure 4: Répartition de la position des groupes par rapport à l'écran.

Nous observons ainsi que chaque utilisateur réalise la lecture de 18 affiches selon 3 distances différentes, soit un total de 54 cas couverts sur les 324 cas possibles par utilisateur. Pour couvrir un jeu de données complet, le nombre d'utilisateur doit être un multiple de 6. 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.

**2.2.4 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 Fig. 2. 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. La Kinect V2 possédant un taux de rafraîchissement de trente images par seconde (30 FPS), nous obtenons un minimum 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.

**2.2.5 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 de la 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 degrés de rotations de la tête et de la distance à l'écran de l'utilisateur mesurés par la Kinect V2.

**2.2.6 Analyse des données.** A la vue des données brutes récoltées pour chaque lecture d'une affiche par chaque utilisateur, nous avons décidé de présenter 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.

**2.2.7 Analyse statistique.** A l'aide du test de comparaison de moyenne par paire de Tukey, nous comparons les différents groupes entre eux en fonction des trois distances à l'écran (LS1, LS2 et LS3) et en fonction des cas de consultation. En effet, les cas de consultation sont les suivants pour l'analyse de la rotation de tangage :

- Cas 0 : Le groupe regarde en face de lui ;
- Cas 1 : Le groupe regarde à plus une colonne écran par rapport à sa position ;
- Cas 2 : Le groupe regarde à plus deux colonnes écran par rapport à sa position ;
- Cas 3 : Le groupe regarde à plus trois colonnes écran par rapport à sa position ;
- Cas 4 : Le groupe regarde à plus quatre colonnes écran par rapport à sa position ;
- Cas 5 : Le groupe regarde à plus cinq colonnes écran par rapport à sa position.

Nous effectuons la même démarche pour la rotation de lacet :

- Cas 1 : Le groupe regarde la ligne supérieure de l'écran ;
- Cas 2 : Le groupe regarde la ligne centrale de l'écran ;
- Cas 3 : Le groupe le regarde la ligne inférieure de l'écran ;

En analysant les données d'orientation de la tête de chaque groupe en fonction de la distance des individus à l'écran, nous testons les hypothèses suivantes :

(H1) la moyenne en degré de la rotation de tangage n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation ;

(H2) la moyenne en degré de la rotation de lacet n'est significativement pas différente d'un groupe à l'autre pour un même cas de consultation ;

Dans un premier temps, afin de vérifier que la Kinect V2 permet de discriminer les angles de rotation de la tête (tangage et lacet) en fonction des différents cas de consultation présentés ci-dessus, nous vérifions que pour chaque cas nous obtenons des différences significatives des angles de rotations de la tête selon les 3 distances possibles. 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.

Dans un second temps, nous réalisons la même démarche afin de montrer que 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 de tangage et pour les trois cas de la rotation de lacet et ceci en fonction des trois distances possibles à l'écran (LS1, LS2 et LS3).

## 2.3 Participants

Quarante-deux personnes ont participé à l'expérience, réparties en 6 groupes. Dans la Fig. 5, on peut 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.

	Age	Sexe	Taille	Latéralité	Œil Directeur	Lunettes
Population	35 ± 14 ans	26H & 16F	1.71 ± 0.08 m	40D & 2G	34D & 8G	23
GR1	39 ± 14 ans	4H & 3F	1.74 ± 0.10 m	6D & 1G	5D & 2G	4
GR2	39 ± 14 ans	4H & 3F	1.71 ± 0.08 m	7D & 0G	5D & 2G	4
GR3	34 ± 11 ans	3H & 4F	1.69 ± 0.10 m	7D & 0G	6D & 1G	4
GR4	27 ± 6 ans	6H & 1F	1.75 ± 0.05 m	7D & 0G	6D & 1G	4
GR5	28 ± 12 ans	5H & 2F	1.71 ± 0.05 m	6D & 1G	7D & 0G	3
GR6	42 ± 16 ans	4H & 3F	1.67 ± 0.07 m	7D & 0G	5D & 2G	4

Figure 5: Tableau des groupes d'utilisateurs.

## 3 RÉSULTATS

Dans la première partie des résultats, nous exposons les angles de rotation théoriques et les angles effectifs mesurés par la Kinect V2 pour chaque cas de consultation. Les angles de rotation théoriques ont été calculés pour l'ensemble des individus de notre population et avec pour origine la tête des

individus. Les angles de rotation mesurés sont les angles donnés par la Kinect V2 mais ramenés à la même origine que les angles de rotation théoriques. Le Tableau 1 présente les données pour la rotation de tangage et le Tableau 2 pour la rotation de lacet. 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. Par construction le test repose donc sur un nombre de données théoriques différent pour chaque cas de consultation. Dans le Tableau 1 et le Tableau 2, le dernier nombre dans la parenthèse indique ce maximum théorique, la différence s'expliquant par les défauts de détections.

Les données sont présentées comme suit : Moyenne angle théorique / moyenne angle mesuré (écart entre moyenne angle théorique et mesuré – nombre de mesures réelles / nombre de mesures théoriques).

**Tableau 1: Comparaison angles de rotation de tangage théoriques VS mesurés en degrés**

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

**Tableau 2: Comparaison angles de rotation de lacet théoriques VS mesurés en degrés**

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

### 3.1 Comparaison des groupes selon chaque cas de consultation pour la rotation de tangage en fonction de chaque distance

**3.1.1 Distance la plus proche de l'écran (LS1).** Ici, nous voulons confirmer qu'il existe bien des résultats significativement différents entre les groupes exposés à des conditions expérimentales différentes.

Les données suivent une loi normale. Mais les échantillons n'ont pas tous des variances égales. 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 3: P-values après procédure de Steel-Dwass-Critchlow-Fligner**

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

Le Tableau 3 met en évidence que le cas 4 et le cas 5 ne sont pas significativement différents au seuil alpha 0.05.

Cependant, suite à un test de Fisher ces deux cas ont des variances égales (p-value = 0,391). Nous avons donc pu appliquer un test t. Nous observons ainsi une différence significative entre le cas 4 et le cas 5 avec une p-value = 0,002.

**3.1.2 Distance intermédiaire (LS2).** Application de la même procédure que précédemment.

**Tableau 4: P-values après procédure de Steel-Dwass-Critchlow-Fligner**

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

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

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

**3.1.3 Distance la plus loin de l'écran (LS3).** Application de la même procédure que précédemment.

**Tableau 5: P-values après procédure de Steel-Dwass-Critchlow-Fligner**

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

Le **Tableau 5** 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.

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

Les données suivent une loi normale. Nous avons donc effectué un test paramétrique de comparaisons des moyennes de Tukey. Nous obtenons une différence significative entre les groupes qui regardent la ligne supérieure de l'écran (LE1), la ligne écran centrale (LE2) ou la ligne écran inférieure (LE3) sur la rotation de lacet (haut-bas) lorsqu'ils sont positionnés en LS1. Nous observons cette différence significative sur la rotation de lacet en fonction des deux autres distances (LS2 et LS3) également.

### 3.3 Comparaison des groupes cas par cas pour la rotation de tangage en fonction de chaque distance

**3.3.1 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 lorsque les groupes d'utilisateurs regardent à plus une, plus deux, plus trois, plus quatre et plus cinq colonnes écran par rapport à leur position et ceci quelle que soit la ligne écran consultée (trois hauteurs possibles, LE1, LE2 et LE3). Nous présenterons les 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, dans un premier temps, nous comparons les résultats des utilisateurs regardant à plus une, plus deux, plus trois, plus quatre et plus cinq colonnes 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 colonnes vers la gauche.

**3.3.2 Cas 1 :** les groupes regardent à plus une colonne par rapport à leur position. Le **Tableau 6** montre que, pour la distance la plus proche de l'écran, il y a une différence significative en moyenne pour les groupes 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.

**Tableau 6: Résultats après test de Tukey comparant les groupes regardant à plus une colonne écran vers la droite**

		Cas 1 vers la droite		
		LS1	LS2	LS3
2 vs 6	0,007	Pas de Différence		2 vs 6 0,036
4 vs 6	0,034	Significative (PDS)		

Pour les mêmes 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.

**3.3.3 Cas 2 :** les groupes regardent à plus deux colonnes 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 ces mêmes groupes regardent à plus deux colonnes écran vers la gauche.

**3.3.4 Cas 3 :** les groupes regardent à plus trois colonnes 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 ces mêmes groupes regardent à plus trois colonnes écran vers la gauche.

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

**Tableau 7: Résultats après test de Tukey comparant les groupes regardant à plus quatre colonnes écran vers la droite**

		Cas 4 vers la droite		
		LS1	LS2	LS3
5 vs 6	0,024	PDS	PDS	PDS

Le **Tableau 7** montre que, pour la distance la plus proche de l'écran, il y a une différence significative 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 ces mêmes groupes regardant à plus quatre colonnes écran vers la gauche.

**3.3.6 Cas 5 :** les groupes regardent à plus cinq colonnes 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 cinq colonnes écran vers la droite. De même lorsque ces mêmes groupes regardent à plus cinq colonnes écran vers la gauche.

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

Nous allons maintenant présenter les résultats de l'étude de la variabilité de la moyenne sur la rotation de lacet 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 ce cas n'est significativement pas différente d'un groupe à l'autre. Nous présentons donc ci-dessous les moyennes sur la rotation de lacet selon les 3 distances possibles à l'écran (LS1, LS2 et LS3) et en fonction des 3 hauteurs (LE1, LE2 et LE3).

**3.4.1 Cas 1 :** les groupes regardent la ligne supérieure de l'écran. Nous n'obtenons aucune différence significative pour la moyenne de rotation de lacet de la tête lorsque les groupes regardent la ligne supérieure de l'écran LE1 et ceci peu importe la distance à l'écran.

**3.4.2 Cas 2 :** les groupes regardent la ligne intermédiaire de l'écran. Nous n'obtenons aucune différence significative pour la moyenne de rotation de lacet de la tête lorsque les groupes regardent la ligne intermédiaire de l'écran LE2 et ceci quelle que soit la distance à l'écran.

**3.4.3 Cas 3 :** les groupes regardent la ligne inférieure de l'écran. Nous n'obtenons aucune différence significative pour la moyenne de rotation de lacet de la tête lorsque les groupes regardent la ligne inférieure de l'écran LE3 et ceci indépendamment de la distance à l'écran.

## 4 DISCUSSIONS

Les premiers résultats présentés en 3 montrent que plus les utilisateurs sont éloignés du capteur 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 on devrait 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 car plus les rotations sont importantes moins l'utilisateur fait face à la caméra car il tourne la tête de manière importante afin d'éviter des angles inconfortables pour les cervicales et donc complique la détection des points caractéristiques du visage nécessaire à un *head tracking* de qualité.

Les résultats présentés en 3.1 montrent de manière significative que, dans notre configuration physique, la Kinect V2 est assez précise pour discriminer les différences d'angle de rotation de tangage 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 lacet, les résultats significatifs en 3.2 montrent que la Kinect V2 est suffisamment précise pour discriminer les groupes regardant la ligne supérieure, intermédiaire ou inférieure de notre écran.

L'ensemble des résultats présentés en 3.3 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 de tangage vers la droite. Or ce sont les mêmes groupes qui réalisent les cas « symétriques » de rotation vers la gauche. De plus, en regardant en détail ces résultats, il apparaît que ce sont principalement les cas concernant la consultation de la colonne écran 6 qui présentent des différences significatives. Malgré le soin apporté à l'installation, il semble qu'une légère inclinaison parasite du capteur ait grevé la symétrie attendue du dispositif. Cette leçon est à retenir pour les expérimentations futures. Les cas montrant des différences significatives laissent également penser que nous sommes dans des conditions limitées de détection du capteur, 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. Ce problème pourrait être dépassé en couplant trois Kinect V2. En effet, en ajoutant une 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 une des 3 Kinect V2.

Enfin, l'analyse de la rotation de tangage 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. Pour les cas où des différences significatives sont mises en avant, seul les rotations de tangage vers la droite sont concernées. Par ailleurs, l'analyse des résultats pour la rotation en lacet ne met en évidence aucune différence significative selon les 3 distances et 3 hauteurs à l'écran.

L'ensemble de cette étude nous permet de conclure que la Kinect V2 semble être un capteur assez précis pour réaliser des interactions, à base de *head tracking*, avec des écrans de grand format. Enfin, les résultats concernant les différents cas de consultation d'information en fonction des deux rotations de la tête conduisent à penser que de manière générale nous orientons tous la tête de manière comparable pour consulter confortablement une information nécessitant une lecture d'au moins 3 secondes.

## 5 CONCLUSIONS

Cette expérience nous a permis de caractériser les mouvements moyens de rotation de tangage et de lacet de la tête des utilisateurs via une Kinect V2 même si les résultats ne confirment pas dans tous les cas nos hypothèses de départ. Dans le cas de notre configuration physique, la Kinect V2 est suffisamment précise pour discriminer des cas de consultation différents. De plus, il apparaît que nous orientons tous la tête de manière comparable pour une tâche de consultation d'information sur un grand écran nécessitant une lecture. Ceci nous a permis de concevoir un système permettant des interactions individualisées sur écran de grand format via un *head tracking* réalisé par une Kinect V2.

## REMERCIEMENT

Les auteurs remercient la Direction Générale de l'Armement qui finance ce programme de recherche ainsi que l'ensemble des volontaires qui ont pris part gracieusement à cette expérience.

## RÉFÉRENCES

- [1] Schiavo, G., et al. Sensing and reacting to users' interest: an adaptive public display. in CHI'13 Extended Abstracts on Human Factors in Computing Systems. 2013. ACM.
- [2] Marion, D., et al. Etude outillée de l'attention visuelle flottante dans la recherche collaborative d'informations sur grands écrans. in EPIQUE 2015. 2015. Aix-en-Provence, France: Arpege Science Publishing.
- [3] Shotton, J., et al., Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 2013. 56(1): p. 116-124.
- [4] Lachat, E., et al., First experiences with Kinect v2 sensor for close range 3D modelling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2015. 40(5): p. 93.
- [5] Khalidov, V. and J.-M. Odobez, Real-time multiple head tracking using texture and colour cues. 2017, Idiap.
- [6] Jeni, L.A., J.F. Cohn, and T. Kanade, Dense 3D face alignment from 2D video for real-time use. *Image and Vision Computing*, 2017. 58: p. 13-24.
- [7] Eichenlaub, J.B., Single Person and Multiple Person Full Resolution 2D/3D Head Tracking Display. 2017, Google Patents.
- [8] Ariz, M., et al., A novel 2D/3D database with automatic face annotation for head tracking and pose estimation. *Computer Vision and Image Understanding*, 2016. 148: p. 201-210.
- [9] Yan, Y., et al., A multi-task learning framework for head pose estimation under target motion. *IEEE transactions on pattern analysis and machine intelligence*, 2016. 38(6): p. 1070-1083.
- [10] Ranjan, R., V.M. Patel, and R. Chellappa, Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *arXiv preprint arXiv:1603.01249*, 2016.
- [11] Drouard, V., et al., Robust head-pose estimation based on partially-latent mixture of linear regressions. *IEEE Transactions on Image Processing*, 2017. 26(3): p. 1428-1440.
- [12] Fanelli, G., J. Gall, and L. Van Gool. Real time 3d head pose estimation: Recent achievements and future challenges. in *Communications Control and Signal Processing (ISCCSP)*, 2012 5th International Symposium on. 2012. IEEE.
- [13] Murphy-Chutorian, E. and M.M. Trivedi, Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2009. 31(4): p. 607-626.
- [14] Foucambert, D., *Syntaxe, vision parafovéale et processus de lecture: contribution du modèle structural à la pédagogie*. 2003, Université Pierre Mendès France (Grenoble).
- [15] Darby, J., et al., An evaluation of 3D head pose estimation using the Microsoft Kinect v2. *Gait & Posture*, 2016. 48: p. 83-88.
- [16] Berthoz, A., *Effets des vibrations sur l'homme*. Scherrer J: *Précis de physiologie du travail*, 2ème édition, Masson édit, 1992.
- [17] Bingham, G.P. and M.G. Stassen, Monocular egocentric distance information generated by head movement. *Ecological psychology*, 1994. 6(3): p. 219-238.