

## Comparatif des principales caméras 3D du marché

Grâce aux caméras TOF, le système peut apporter les avantages supplémentaires suivants par rapport aux classiques caméras couleurs : le fond n'a plus besoin d'être fixe et connu, le port de gants ou de marqueurs est inutile (puisque tous les algorithmes doivent se baser sur les distances et non plus sur les couleurs), la lumière ambiante peut varier (dans une certaine mesure) puisque la caméra fonctionne grâce à son propre projecteur infrarouge, les mains de (ou des) utilisateur(s) peuvent être suivies et ce quelque soient leur couleur, leur taille ou leur morphologie. Il n'y a plus besoin de faire attention à avoir des habits de couleur différente du fond et de la couleur de la peau. En revanche, l'inconvénient principal qui reste est l'occlusion possible d'une partie ou de la totalité des mains.

Avant 2011, quelques travaux ont porté sur l'utilisation de ces caméras 3D pour faire de la capture et de la reconnaissance de mouvements des mains (Breuer et al., 2007; Lahamy & Lichti, 2010; Li & Jarvis, 2009; Malassiotis & Srinivasan, 2008; Mo & Neumann, 2006; Suryanarayan et al., 2010) ou encore du corps entier (Holte & Moeslund, 2007; Kollorz et al., 2008). Ces travaux portent surtout sur la reconnaissance des postures de la main et ne permettent pas d'interagir avec l'environnement virtuel en temps réel. Hormis quelques laboratoires qui, au fil des années, se sont intéressés à ces périphériques, la majorité des techniques sans marqueurs portaient sur l'usage d'une ou de plusieurs caméras. Le prix très élevé des caméras 3D, du fait de leur faible commercialisation, ainsi que la faible résolution qu'elles offraient, constituaient très certainement les freins principaux. Une caméra 3D *SwissRanger* coûte environ 7000€ pour une résolution de 176x144 pixels. De même, puisque le nombre d'algorithmes permettant de traiter des nuages de points en temps réel est très faible comparé à ceux traitant les informations de couleur, peu de laboratoires se sont penchés sur la capture des mouvements de la main en temps réel à partir de caméras 3D. L'arrivée de la *Kinect* en novembre 2010 pour un prix de 150€ a permis de populariser largement ce type de périphérique et a, depuis, suscité un réel intérêt pour cette technologie.

## 5 Objectifs, Motivations et Hypothèses

L'objectif du travail présenté ici est la conception et l'évaluation d'un système permettant une interaction 3D transparente et temps réel en environnement virtuel, pour des tâches de sélection, de manipulation et de navigation. Le système est ici à comprendre comme l'association d'une technologie, de traitements algorithmiques et de modalités d'interactions. Nous souhaitons en évaluer de manière détaillée les performances d'une part, et la mesure de l'acceptabilité et des préférences des participants d'autre part. La finalité de notre travail est la mise à disposition d'un système efficace, efficient, confortable, performant et améliorant le sentiment d'immersion. Notre objectif général, et la finalité de notre travail, est motivée par la revue des matériels et techniques existantes, ainsi que par la quasi absence d'études démontrant l'intérêt de ce type de matériel pour l'interaction 3D temps réel. Nous allons dans un premier temps rappeler brièvement les raisons qui ont motivé le choix de la technologie de caméra 3D, puis nous présenterons le modèle d'interaction qui en découle. Nous présenterons ensuite nos hypothèses générales, définirons les critères d'évaluation et finirons cette partie par la structuration de notre travail.

### 5.1 Le choix d'une technologie

Une technologie ayant pour objectif de faire « disparaître » l'interface afin de permettre un mode d'interaction « naturel » aux utilisateurs est souvent présentée comme l'aboutissement final des interfaces de manipulation directe. (Fuchs & Moreau, 2003) parlent à ce propos d'interfaces

comportementales, ce qui implique que les recherches doivent se tourner vers la conception de systèmes « visant à exploiter un comportement humain, naturel et sans acquis préalable ». De même, comme le suggèrent Fuchs et Mathieu, dans le volume 2 du *Traité de la Réalité Virtuelle* (Fuchs & Mathieu, 2003), de manière idéale, les réponses motrices (dans le cadre de la capture de mouvements) doivent être transmises sans support matériel entre l'homme et la machine. Enfin, d'après (Winkler et al., 2007), le fait de n'avoir aucun matériel à porter et de rendre le système transparent pour l'utilisateur est de nature à améliorer le sentiment d'immersion et de présence ce qui est primordial dans le domaine de la réalité virtuelle.

Tout système possède naturellement des inconvénients mais les caméras 3D permettent de s'affranchir de la majorité des contraintes que l'on peut rencontrer avec les autres périphériques (tels que la couleur, la luminosité ambiante, les perturbations liées au son ou au métal, l'utilisation de capteurs, etc.). L'absence de périphérique en contact direct avec l'utilisateur est à même d'apporter une plus-value substantielle par rapport aux périphériques actuels. Ce sont ces raisons qui ont motivé notre choix pour cette technologie. Nous nous sommes orientés vers une caméra *SwissRanger* dans un premier temps, car elle était celle qui avait le meilleur rapport taux de rafraîchissement / résolution au moment du développement (la *Microsoft Kinect* n'était pas encore disponible à la vente et faisait alors l'objet de beaucoup de spéculations) et la précision était quasiment du même ordre entre les différentes caméras 3D. Notre système a ensuite été amélioré par l'utilisation de la *Kinect*, qui a été commercialisée durant la dernière année de notre travail de thèse. Nous avons décidé d'adopter ce nouveau périphérique car il offre une résolution supérieure tout en conservant un taux de rafraîchissement compatible avec la réalisation d'une interaction temps réel.

## 5.2 Notre vision du modèle d'interaction de Norman

Le choix de concevoir un système transparent, et donc de nous baser sur une caméra 3D nous conduit à proposer notre modèle de l'interaction homme-machine inspiré, en partie, du modèle de (Norman, 1988).

Le modèle de Norman (Figure 24) suggère que l'utilisateur envoie des stimuli à un périphérique quelconque (souris, clavier, space mouse, etc...) qui convertit ces informations et les envoie à l'ordinateur ; celui-ci retourne des données visuelles à l'utilisateur. Ce modèle a été enrichi (flèches grises) par (Nedel et al., 2003) par l'ajout des retours entre l'ordinateur et le périphérique, puis entre ce dernier et l'utilisateur lors de l'utilisation de périphériques haptiques.

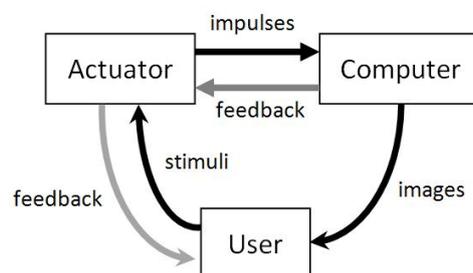


Figure 24 : Modèle de l'interaction homme-machine, tiré de (Norman, 1988)

Dans le système que nous proposons nous n'avons aucun feedback entre l'utilisateur et le périphérique, ni entre l'ordinateur et le périphérique, car le système est totalement transparent pour l'utilisateur. Il n'y a aucun capteurs ni matériel à porter et de ce fait aucun retour possible au niveau du périphérique.

La Figure 25 présente le fonctionnement de notre système, en lien avec le modèle de Norman présenté précédemment.

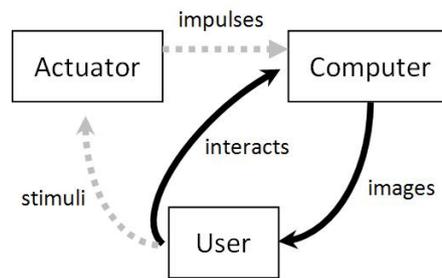


Figure 25 : Notre modèle de l'interaction homme-machine

Le cycle représenté en noir sur le schéma correspond à ce que l'utilisateur perçoit. Quand l'utilisateur bouge sa main, son geste lui est visuellement retransmis par l'ordinateur : ce retour visuel, associé à l'absence de port de matériel rend totalement transparent le périphérique. Il n'en reste pas moins qu'il est présent. La boucle représentée en gris sur notre schéma correspond à ce qui se passe réellement. L'utilisateur bouge son bras dans l'espace, le périphérique (une caméra 3D) se charge d'envoyer les informations à l'ordinateur qui peut ainsi déterminer la position et l'orientation de la main de l'utilisateur, ainsi que les actions qu'il est en train de réaliser. Enfin, un retour visuel est restitué à l'utilisateur. Il est donc très important, pour l'utilisateur, qu'il n'y ait pas de décalage temporel entre l'action réelle qu'il effectue et la restitution visuelle donnée par l'ordinateur. Il faut donc, d'une part que les traitements se fassent en temps réel et d'autre part avec une latence la plus faible possible.

Dans le cas d'une interaction en environnement virtuel, il existe plusieurs tâches (Coquillart et al., 2003) dont les principales sont :

- La sélection, c'est-à-dire positionner sa main (ou plus exactement l'avatar représentant sa main) sur un objet, pour pouvoir par la suite interagir avec lui. Le mouvement réciproque, naturellement, est la désélection ;
- La manipulation, c'est-à-dire la modification de l'état d'un objet supposé ici indéformable. Il s'agit de modifier sa position, son orientation et sa taille ;
- La navigation, c'est-à-dire la possibilité de se déplacer dans l'environnement virtuel et de l'explorer.

Nous avons donc choisi d'implémenter ces 3 tâches au sein de notre système.

### 5.3 Hypothèses générales

Les études expérimentales présentées dans les chapitres suivants visent à évaluer le système que nous proposons. Celui-ci a pour objectif d'obtenir des performances a minima équivalentes aux systèmes couramment utilisés pour les tâches de sélection et de manipulation d'objets 3D ainsi que pour les tâches de navigation. Nous ne visons pas de performances meilleures car nous savons qu'il est très difficile, avec un nouveau système, de dépasser les performances des systèmes avec lesquelles les utilisateurs ont une très grande pratique. L'influence de l'apprentissage étant particulièrement importante dans le cas de l'usage de la souris pour les tâches de sélection. Notre système vise par contre une meilleure acceptabilité, compte tenu des avantages que nous avons mis en avant précédemment. L'évaluation d'un gain se fait comparativement à des systèmes couramment utilisés : la

souris, pour la tâche de sélection d'un objet et les gants de données pour la manipulation et la navigation en environnement virtuel. Nos hypothèses de recherche sont donc les suivantes :

- Hypothèse A : les performances obtenues avec notre système, en termes de temps d'exécution, sont sensiblement équivalentes à celles obtenues avec la souris pour une tâche simple de sélection en environnement virtuel. Ces performances sont indépendantes de la taille de la main des participants, de leur sexe et de leur expertise en réalité virtuelle.
- Hypothèse B : l'acceptabilité est meilleure avec notre système, en termes d'utilisabilité, d'utilité et de sentiments d'immersion, pour une tâche de sélection, en comparaison de la souris.
- Hypothèse C : la précision et le temps de réalisation des tâches de manipulation et de navigation avec notre système sont sensiblement équivalents à ceux obtenus avec des gants de données associés à des capteurs électromagnétiques.
- Hypothèse D : notre système est à même de recueillir une meilleure acceptabilité et de meilleurs jugements subjectifs par les participants, pour des tâches de manipulation et de navigation en environnement virtuel, comparativement à des gants de données associés à des capteurs électromagnétiques. En particulier, du fait de l'absence de port de matériel et d'entrave due aux fils, nous pensons que notre système apporte un meilleur confort, une meilleure précision et une meilleure liberté de mouvements.
- Hypothèse E : les modalités d'interaction que nous proposons sont adaptées à notre système et permettent de réaliser de manière simple, confortable, efficace et efficiente les tâches de sélection, manipulation et navigation.

#### 5.4 Critères d'évaluation

Pour vérifier ces hypothèses, nous avons étudié, d'une part, les performances obtenues et, d'autre part, l'acceptabilité des utilisateurs envers le système que nous avons créé, lors de la réalisation de tâches de sélection, de manipulation et de navigation.

Les performances sont évaluées sous l'angle du temps nécessité par l'exécution d'une tâche, et par la précision dans le cas plus particulier des tâches de déplacement, d'orientation, de redimensionnement et de navigation. Cette précision est évaluée comme une fidélité entre la position, l'orientation et la taille souhaitées d'un objet et ces paramètres effectifs.

D'après (Tricot et al., 2003) « *l'acceptabilité [en anglais : acceptability] désigne la valeur positive ou négative de l'image que se fait un individu (ou un groupe) à propos de l'outil ou du système. En d'autres termes, l'acceptabilité se mesure au niveau des attitudes et opinions des utilisateurs potentiels vis-à-vis du produit lui même, de son utilité et de son utilisabilité.* » Nous avons donc défini l'acceptabilité selon trois sous-catégories : le sentiment d'immersion, l'utilité et l'utilisabilité en comparant à chaque fois notre système à des périphériques considérés comme les plus adaptés aux tâches à effectuer. D'autre part, nous avons décomposé l'utilisabilité en termes d'efficacité, d'efficience et de satisfaction. Ce découpage repose sur les notions d'ergonomie et les définitions tirées de Burkhardt. D'après (Burkhardt & Sperandio, 2004) « *les critères généraux de l'ergonomie concernent l'utilité, l'utilisabilité, l'accessibilité, la dangerosité, la satisfaction et l'acceptabilité. Ces critères délimitent de grandes familles d'objectifs ergonomiques pour la conception et l'évaluation* ». Parmi ces critères certains sont plus importants que d'autres et peuvent être hiérarchisés de façon variable en termes de priorité suivant le projet. Les deux critères principaux d'utilité et d'utilisabilité ressortent cependant plus que les autres. Nous adoptons ici, les définitions d'utilité et d'utilisabilité selon les termes employés par (Burkhardt, 2006).

« En ergonomie, le critère d'utilité [anglais : usefulness] correspond à un avantage ou un agrément significatif pour l'utilisateur dans une activité précise (en termes d'efficacité, de coût, de rapidité, de précision, d'agrément...). Cet avantage est toujours relatif : relatif aux objectifs de l'utilisateur, relatif aux outils existants ou habituellement utilisés, relatif à l'environnement d'utilisation, et relatif aux dépendances avec les autres activités. » (Burkhardt, 2006)

« L'utilisabilité [en anglais : usability] peut se définir comme la mesure (ou une estimation probable) de la performance des utilisateurs, dans le contexte et pour l'utilisation prévue. L'utilisabilité se mesure par divers indicateurs tels que la facilité d'apprentissage, la facilité de mémorisation, l'utilisation sans erreurs, etc. [...] Il en découle que l'utilisabilité n'est pas la propriété absolue d'un dispositif ou de telle technique d'interaction : elle est toujours relative à une tâche donnée, un contexte de réalisation et une population ciblée d'utilisateurs. » (Burkhardt, 2006)

Les critères d'accessibilité et de dangerosité ne sont pas évalués dans nos études. Toujours, d'après Burkhardt, « l'accessibilité concerne l'aménagement de l'accès aux systèmes d'information pour les personnes handicapées » et « la dangerosité regroupe l'ensemble des conséquences négatives possibles associées à l'utilisation du système notamment pour l'utilisateur, sa santé (santé, accident) et son environnement. [...] Au-delà des risques liés à un domaine d'application particulier, la dangerosité des environnements virtuels dépend notamment du type de configuration d'interfaçage exploitée : gants, visiocasque, exosquelette, etc. » (Burkhardt, 2006)

Enfin d'après (Green & Jordan, 2002), « le critère de satisfaction ou de plaisir [en anglais : satisfaction, enjoyment] [...] désigne l'expérience globalement positive qui marque de façon durable les utilisateurs du produit ou de l'outil conçu. Cette expérience s'analyse à travers plusieurs dimensions : satisfaction sur le plan de la pensée, des sensations et des émotions, etc. L'évaluation de ce critère s'appuie sur des échelles d'attitudes ou des questionnaires. »

## 5.5 Structuration du travail

Les études présentées dans les parties suivantes sont au nombre de trois, deux études principales et une étude complémentaire. La première étude est basée sur l'évaluation comparative de notre système et d'une souris pour une tâche simple de sélection 3D en environnement virtuel. Elle s'appuie sur la caméra 3D *SwissRanger SR 4000*. L'objectif premier de cette étude est d'attester l'intérêt de notre système et d'avoir un premier retour des participants sur le système proprement dit ainsi que sur la modalité d'interaction. Cette étude permettra de vérifier les hypothèses A et B.

La seconde étude reprend de nombreux éléments du protocole expérimental de l'étude précédente, et s'appuie sur une version améliorée du système initial, grâce à l'emploi de la *Kinect* de *Microsoft* et l'intégration de nouvelles bibliothèques permettant un temps de calcul réduit tout en offrant une plus grande robustesse. Cette étude permettra de valider les hypothèses C et D et l'hypothèse E dans une moindre mesure.

La troisième étude est complémentaire aux deux premières. Elle vise à prendre en compte les commentaires des participants concernant l'implémentation des modalités d'interaction. Elle opposera donc les modalités d'interaction utilisées pour la manipulation dans l'étude 2 (déplacement, rotation et redimensionnement) à celles implémentées d'après les suggestions et commentaires des participants. Elle apportera ainsi une seconde réponse à l'hypothèse E.