



**HAL**  
open science

# Apport des NUIs pour les Applications de Réalité Virtuelle et Augmentée: État de l'Art

Fahima Djelil, Samir Otmane, Siju Wu

## ► To cite this version:

Fahima Djelil, Samir Otmane, Siju Wu. Apport des NUIs pour les Applications de Réalité Virtuelle et Augmentée: État de l'Art. 8èmes journées de l'Association Française de Réalité Virtuelle, Augmentée, Mixte et d'Interaction 3D, Oct 2013, Laval, France. hal-02455134

**HAL Id: hal-02455134**

**<https://hal.science/hal-02455134>**

Submitted on 25 Jan 2020

**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.

# Apport des NUIs pour les Applications de Réalité Virtuelle et Augmentée: État de l'Art

Fahima Djelil \*

Laboratoire IBISC - Université d'Évry Val d'Essonne

Samir Otmane †

Laboratoire IBISC - Université d'Évry Val d'Essonne

Siju Wu ‡

Laboratoire IBISC -  
Université d'Évry Val  
d'Essonne

## RÉSUMÉ

L'évolution des Interfaces Homme-Machine de la notion de Graphical User Interface (GUI) au paradigme naissant de Natural User Interface (NUI), offre à l'utilisateur une interaction directe, sans l'intermédiaire d'outils dédiés. Le contrôle est réalisé par le seul moyen du corps humain. L'intégration des NUIs (Natural User Interfaces) dans les applications de réalité virtuelle et augmentée permet de définir de nouvelles métaphores et techniques d'interaction 3D. Cependant, la conception et l'évaluation de ces techniques d'interaction naturelles est problématique, car il n'existe à notre connaissance, aucun élément de conception permettant de les formaliser. Ceci est principalement dû au manque de connaissances suffisamment complètes relatives à ce nouveau concept que sont les NUIs. Dans ce papier, nous présentons un état de l'art sur les NUIs visant à connaître d'une part, les éléments essentiels caractérisant les NUIs, et d'une autre part, leur apport pour les applications de réalité virtuelle et augmentée, notamment en termes d'interaction utilisateur.

**Index Terms:** H.5.2 [Information interfaces and presentation]: User Interfaces, Natural language—Interaction styles; H.5.1 [Information interfaces and presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

## 1 INTRODUCTION

À mesure que la technologie évolue, de nouveaux paradigmes d'interaction homme machine émergent. De nombreuses recherches dans le domaine des HCI (Human Computer Interaction) ont complètement révolutionné notre rapport avec la machine. De l'écran tactile multitouches à la reconnaissance automatique du geste, on assiste aujourd'hui à une explosion de possibilités d'interaction.

Les environnements de réalité virtuelle et augmentée n'en sont pas moins épargnés. Vis-à-vis de l'utilisateur, l'interaction devient de plus en plus transparente et conviviale. L'utilisateur emploiera sa voix, sa main, son regard, le mouvement de son corps ou encore ses pensées comme simples moyens d'interaction et de communication avec son environnement instrumenté.

L'expérience d'interaction est alors centrée sur le comportement naturel de l'homme. La communication se fait de façon naturelle, directe et intuitive. L'interface d'interaction n'est plus une barrière pour l'utilisateur, mais elle est faite de ses propres sens. On parle alors d'interfaces utilisateur naturelles, connues sous l'acronyme NUI (Natural User Interface).

\*e-mail: fahima.djelil@gmail.com

†e-mail:samir.otmane@ibisc.univ-evry.fr

‡e-mail:siju.wu@ibisc.univ-evry.fr

Cependant, le terme NUI prête souvent à confusion, et il est souvent mal interprété. En effet, les interfaces naturelles ne sont pas suffisamment claires, car il n'existe à notre connaissance, aucune représentation précise de ce concept. Leur modélisation relève de la recherche et nécessite une connaissance suffisante [36].

L'avènement des NUIs et leur exploitation dans les applications de réalité virtuelle et augmentée, permet de définir de nouvelles métaphores et techniques d'interaction 3D, qui sont dites naturelles. Cependant, l'absence de formalismes liés au concept de NUIs, constitue un réel handicap pour la conception et l'évaluation de telles techniques. De plus, les NUIs sont tout aussi diverses que complexes, et leur déploiement dans les applications de réalité virtuelle et augmentée commence à peine à se généraliser.

Ce papier est un état de l'art sur les NUIs visant à montrer les notions relatives aux NUIs afin de connaître leur apport pour les applications de réalité virtuelle et augmentée. Il s'organise en deux parties essentielles. Il présente en premier lieu, une vue globale des NUIs faisant connaître leur forme, les moyens d'interaction offerts par ces interfaces, ainsi que leur principe de fonctionnement. Dans un second temps, il donne les résultats de cette première étude qui concernent, une définition claire et précise du concept de NUI, l'ensemble des caractéristiques communes à ce type d'interfaces, et enfin, une classification de ces interfaces suivant deux axes, permettant de connaître d'une part, la satisfaction de chacune des caractéristiques identifiées, et d'une autre part, les tâches d'interaction 3D qui sont possibles à réaliser par ces interfaces.

## 2 DÉFINITION D'UNE NUI

Il n'existe dans la littérature aucune définition claire et précise des NUIs. Néanmoins, certains auteurs ont donné quelques indications concernant certains aspects de ce concept. Selon [36] le mot "natural" dans NUI, dénote à la fois l'aspect intuitif et facile à utiliser ou à apprendre de ces interfaces. Ce terme ne se réfère pas à l'interface elle-même et encore moins à sa technologie, mais à la façon dont l'utilisateur interagit avec le système et la manière dont il le sent et l'utilise [49].

## 3 MODALITÉS D'INTERACTION NATURELLE

Plusieurs travaux se sont intéressés à l'exploitation des sens de l'homme dans la définition de techniques d'interaction utilisateurs naturelles, notamment dans le domaine de réalité virtuelle et augmentée. Ainsi, sont exploités la main [30, 44], la voix [31], la direction du regard [43], le mouvement du corps [18], le mouvement des lèvres [19], les pensées ou l'activité cérébrale [24, 13], ou encore l'expression faciale [5]. Dans ce qui suit, nous détaillerons chacune de ces modalités naturelles.

### 3.1 La main

L'interaction directe par la main est considérée comme le moyen d'interaction le plus naturel étant donné que dans le monde réel, nous interagissons avec les objets physiques par la main.

### 3.1.1 Interaction par la main avec contact

L'interaction par la main directe avec contact se fait par contact direct avec la surface de visualisation qui peut être un écran tactile, ou une surface physique augmentée. On parle alors d'interaction *tactile* ou d'interaction *tangible*.

- Interaction tactile : se fait par le contact direct de la main sur une surface tactile. Le contact peut être multi-doigts, et peut se faire par la main entière (Figure 1). Il peut être statique (postures des doigts ou de la main), ou dynamique (geste des doigts ou de la main) [16]. Le geste est défini comme une séquence de postures séparées par des intervalles de temps continus [14].



FIGURE 1: Interaction tactile par la main entière [52].

- Interaction tangible : se fait par le contact direct de la main avec des représentations visuelles d'entités numériques en réalité augmentée et mixte, ou par l'emploi d'objets réels attachés à la main [42, 25] (Figure 2).



FIGURE 2: Interaction tangible [48].

### 3.1.2 Interaction par la main sans contact

L'interaction par la main directe sans contact se fait sans contact direct avec la surface de visualisation. Une certaine distance est maintenue entre les deux entités (Figure 3)<sup>1</sup>. Afin de traduire l'action de l'utilisateur, certains systèmes se basent sur des techniques d'apprentissage automatique [41, 9, 7]. Dans certains d'autres systèmes [22, 45], seules les positions et orientations de la main sont requises pour agir dans l'environnement virtuel.



FIGURE 3: Interaction par la main directe sans contact dans un site chirurgical stérile.

1. <http://research.microsoft.com/apps/video/default.aspx?id=159629>

### 3.2 La voix

La voix a été exploitée comme moyen d'interaction dans plusieurs applications de réalité virtuelles. En effet, en plus d'être un moyen d'interaction naturel et familier, la modalité vocale offre à l'utilisateur la possibilité de sélectionner et de manipuler des objets qui ne sont pas forcément visualisés dans l'espace virtuel [32]. Nous citons à titre d'exemple, les systèmes QuickSet [8] et DIVERSE [31] qui permettent la sélection et la manipulation d'objets virtuels via des commandes vocales. La navigation est également facilitée par cette modalité [35].

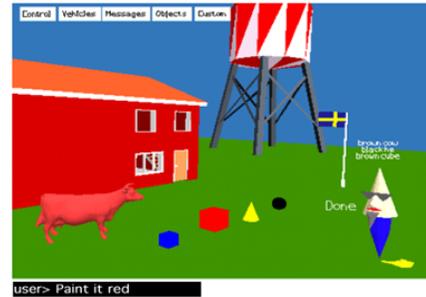


FIGURE 4: Interaction par la voix : le message vocal est retranscrit sur l'écran [31].

### 3.3 Le Regard

L'interaction par le regard est l'une des modalités naturelles les plus intéressantes, notamment pour les personnes atteintes de paralysie [27]. Cette modalité a été expérimentée afin de réaliser des tâches de sélection (Figure 5) et de manipulation [10, 53], notamment dans des applications de téléconférence virtuelles [47] et augmentées [43]. Le regard est très souvent combiné avec d'autres modalités plus explicites afin d'améliorer le déterminisme de l'interaction et de prédire les actions de l'utilisateur en vue de prévoir un meilleur rendu en temps réel [39, 54].



FIGURE 5: Sélection par le regard dans un environnement virtuel immersif [40].

### 3.4 Le mouvement du corps

L'interaction par le corps en entier est largement exploitée en réalité virtuelle, notamment dans les jeux vidéo virtuels où tous les membres de l'utilisateur, essentiellement la tête, les mains et les pieds, sont traqués afin de manipuler un objet dans la scène virtuelle, ou de reproduire le mouvement du corps à travers un avatar représentant l'utilisateur [50] (Figure 6). La navigation et le contrôle d'application sont également possibles à l'aide de cette interface, elles ont été respectivement expérimentées dans les systèmes SURVIVE (Simulated Urban Recreational Violence IVE) et ALIVE (Artificial Life IVE) [51].

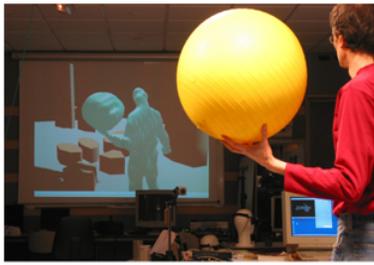


FIGURE 6: Interaction par le corps [18].

### 3.5 Le mouvement des lèvres

Le mouvement des lèvres ne constitue pas en lui seul une modalité d'interaction à part entière. En effet, il est le plus souvent exploité pour améliorer la reconnaissance du signal vocal, notamment dans des environnements bruyants [4]. Cependant, on estime qu'on peut l'utiliser afin de construire un moteur d'animation d'acteurs virtuels assez puissant [1, 2]. Il a également été démontré que cette modalité peut offrir un moyen d'interaction très puissant pour les personnes âgées et les personnes atteintes de paralysie. Dans ce contexte, un prototype d'interface basée sur le mouvement des lèvres pour la commande d'un robot a été expérimenté ( Figure 7).<sup>2</sup>



FIGURE 7: Interaction par le mouvement des lèvres.

### 3.6 L'activité cérébrale

L'envoi de commandes par l'activité cérébrale est un autre moyen d'interaction naturelle en réalité virtuelle. Plusieurs travaux ont démontré que grâce à une interface cerveau-ordinateur, il est possible de se déplacer, de sélectionner et de manipuler des objets virtuels [28]. Il est ainsi possible de déplacer des objets en faisant varier le niveau de concentration mentale [37, 24], ou de communiquer avec des avatars et de se déplacer dans l'environnement virtuel, par l'imagination du mouvement des pieds et des mains [26, 13] ( Figure 8).



FIGURE 8: Interaction par l'activité cérébrale [26].

### 3.7 L'expression faciale

L'expression faciale est également utilisée comme moyen d'interaction Homme-Machine, la simulation des expressions humaines par des avatars virtuels est l'une de ces applications [23]. L'expression faciale est également combinée au signal de parole, afin d'étudier son influence sur la modalité vocale [5]. Il a été ainsi démontré, que l'expression du visage possède une grande influence sur le contenu du message vocal (Figure 9).

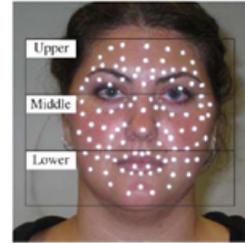


FIGURE 9: Interaction par l'expression du visage [5].

## 4 INTERFACES UTILISATEURS NATURELLES MULTIMODALES ET/OU COLLABORATIVES

### 4.1 Interfaces utilisateurs naturelles multimodales

#### 4.1.1 Intérêt de la multimodalité dans une NUI

D'après [39], une interface utilisateur naturelle multimodale combine les modalités d'interaction naturelles, comme la voix, le regard et le geste, afin d'assurer un moyen de communication assez puissant. Cependant, l'intérêt d'une interface naturelle multimodale est multiple [39, 32] :

- Accessibilité : les interfaces naturelles multimodales étendent l'accessibilité des systèmes informatiques à des utilisateurs multiples et novices, dans des contextes multiples.
- Stabilité et robustesse : elles offrent à l'utilisateur la possibilité de réaliser ses tâches de façon à minimiser les erreurs commises. Pour cela les modalités sont parallèles ou dupliquées. La Multimodalité permet également de gérer les ambiguïtés éventuelles, liées à des erreurs de reconnaissance des signaux d'entrée.
- Efficacité : elles offrent à l'utilisateur une puissante interface de manipulation avec des moyens de visualisation de plus en plus sophistiqués. Il a également été démontré, qu'elles constituent un moyen d'interaction plus rapide que les interfaces traditionnelles.
- Flexibilité : l'utilisateur reste libre de décider de la modalité à utiliser dans ses actions. A titre d'exemple, il peut choisir une interaction directe pour la navigation dans l'environnement virtuel, et une modalité vocale pour la manipulation.

#### 4.1.2 Modes de combinaison des modalités

Dans la combinaison de modalités, Oviatt et Cohen parlent de deux approches principales. Une première approche qui se base sur la *fusion de sources d'informations* et qui s'effectue de façon *parallèle*, et une deuxième approche qui fusionne la sémantique des entrées [39]. Une troisième approche a été définie par Zhai et al, [54], elle consiste à fusionner les modalités en *cascade* :

- Approche de combinaison parallèle : consiste à fusionner au moins deux sources d'informations. Cette approche est adéquate à l'interprétation de signaux ayant une échelle de temps similaire. Par exemple, dans la reconnaissance de la parole dans un environnement bruyant, on procède par fusionner les phonèmes extraits du signal de parole avec les visèmes qui sont des informations visuelles du mouvement des lèvres, produit lors de la prononciation de ces phonèmes [4].

- Approche de combinaison sémantique : consiste à fusionner la sémantique de deux signaux. Ce qui s’applique bien aux signaux non simultanés et qui peuvent être interprétés indépendamment comme le cas de la parole et de la gestuelle [8]. Cette approche entraîne peu de traitements et requiert moins de données d’apprentissage.
- Approche de combinaison en cascade : la combinaison se fait de sorte à ce qu’une modalité puisse précéder une autre, selon que la modalité soit *Active* (envoi de commandes explicites au système, comme la parole, le toucher ou toute autre commande manuelle) ou *passive* (l’utilisateur n’envoie pas des commandes explicites au système, c’est ce dernier qui traque les sens de l’utilisateur comme son regard ou la position de son corps). Cette approche est intéressante dans le cas où on veut améliorer ou prédire une modalité active, après avoir reçu des informations issues d’une modalité passive [54].

## 4.2 Interfaces utilisateurs naturelles collaboratives

Quand une interface utilisateur offre la possibilité de gérer des activités d’un groupe d’utilisateurs, elle est dite *multiutilisateur* ou *collaborative*. De plus en plus de dispositifs supportant l’interaction naturelle et collaborative sont entrain de voir le jour. Les grands écrans muraux et les tables multi-touches en sont un exemple. Ceci laisse le champ à la recherche et à l’évaluation de nouvelles techniques d’interaction toujours centrées sur le comportement naturel de l’utilisateur.

On peut citer les travaux menés par Morris et al, pour la conception d’un système offrant un espace de travail commun à une équipe constituée de quatre personnes, recevant chacune des données auditives sur des canaux individuels (Figure 10). Au cours de leurs travaux, ils avaient étudié l’impact d’un tel dispositif sur les stratégies de communication et de collaboration dans un groupe de travail [33].



FIGURE 10: Interaction naturelle et collaborative : quatre personnes partageant une table multitouche, recevant chacune des données auditives sur des canaux individuels [33].

Plusieurs travaux se sont, par ailleurs, intéressés à la conception d’espaces numérisés pour le travail collaboratif, comme les salles de réunion équipées de dispositifs d’affichage multiple et partagé offrant une interaction tactile et tangible, ou encore au moyen de stylos numériques pour le dessin et les annotations. Nous citons à titre d’exemple les plateformes NiCE Discussion Room [17], Media Room [21], RoomPlanner [52], et i-LAND [46].

## 4.3 Quelques interfaces naturelles multimodales et collaboratives existantes

Il existe plusieurs interfaces naturelles multimodales et collaboratives. Nous citons à titre d’exemple l’interface SHIVA (Several-Humans Interface with Vision and Audio) qui est une interface multimodale permettant à plusieurs utilisateurs d’interagir librement par le geste et la parole avec de grands écrans [6]. Elle a été testée sur un jeu d’échec, où les utilisateurs interagissent tour à tour

au geste et à la voix pour déplacer des pions sur un échiquier virtuel. SHIVA repose sur un système de suivi du corps et des mains et intègre un module de reconnaissance de la parole.

DAVE\_G (Dialogue-Assisted Visual Environment for Geoinformation) est un autre exemple de système reposant sur une interface naturelle multimodale et collaborative. Son interface supporte les gestes de la main libre ainsi que les commandes vocales [29]. L’interface offre aux utilisateurs la possibilité d’effectuer des requêtes d’interrogation de données, de naviguer et de tracer sur l’interface de visualisation de façon simultanée, tout en restant libres du choix de la modalité à utiliser.

Une autre interface naturelle multimodale et collaborative a été mise en place par König et al [21]. Elle est conçue pour offrir à un groupe d’utilisateurs une structure commune pour collaborer à la réalisation de croquis. Elle repose sur une infrastructure distribuée comprenant de grands écrans en haute résolution, des téléphones portables tactiles ainsi que des tables tactiles multi-touches. Cette interface intègre des modules de reconnaissance de la parole et de suivi du regard et supporte les gestes de la main libre. Les utilisateurs peuvent également se servir de stylets et d’outils physiques pour une interaction tangible ( Figure 11).



FIGURE 11: Interaction naturelle multimodale et collaborative [21].

## 5 PRINCIPE DE FONCTIONNEMENT DES NUIS

Dans une interface naturelle, l’interprétation des commandes naturelles comme la voix ou le geste de la main, nécessite plusieurs types de traitements. Dans de tels systèmes, une entrée est vue comme un événement qui est sujet à un processus de reconnaissance et qui peut être mal interprété, d’autant plus que les flux d’entrée sont continus et souvent simultanés [39]. De plus, avant d’aboutir à des données prêtes à être interprétées, celles-ci sont souvent issues d’une longue chaîne de traitement, appliquée depuis leur acquisition grâce à certains dispositifs de capture, jusqu’à la traduction des commandes naturelles proprement dites.

Ceci mène à dire que le fonctionnement des NUIS repose essentiellement sur trois éléments principaux : les technologies nécessaires à la capture et au traitement des données initiales, les méthodes essentielles à l’extraction des données pertinentes, et les algorithmes de reconnaissance et d’interprétation des commandes naturelles.

### 5.1 Les technologies NUIS

Il existe aujourd’hui, une multitude de technologies d’interfaces que l’on peut déployer dans une interface utilisateur naturelle. Murphy, S. parle essentiellement des six technologies suivantes [34] :

- Les écrans tactiles : les écrans tactiles les plus répandus aujourd’hui sont de type capacitif, que l’on retrouve essentiellement sur les Smartphones et les tablettes tactiles. Le contact des doigts sur la couche externe est traité par un contrôleur interne.
- La reconnaissance de la parole : les technologies de traitement de la parole existantes sont très évoluées, notamment grâce aux algorithmes de reconnaissance qui sont de plus en plus

sophistiqués, couvrant un vocabulaire très large, avec des accents de langage très variés, et pouvant être entraînés sur des voix multiples.

- Le graphisme : la qualité du graphisme affiché à l'écran dépend de la puissance du processeur graphique qui se charge de traiter les images en fonction de la résolution et de la profondeur des objets voulues, ainsi que de la capacité de stockage de la mémoire qui conserve les images traitées et les scènes 3D.
- Les technologies sans fil de courte portée : certaines technologies de communication sans fil de courte portée, sont entrain d'être intégrées dans les interfaces naturelles. Malgré leurs limites opérationnelles de leurs protocoles, celles-ci s'avèrent très avantageuses, puisqu'elles sont très peu consommatrices en énergie.
- La stéréoscopie 3D (S3D) : cette technologie permet de visualiser des contenus en 3 dimensions. Un dispositif de deux caméras 3D permet également d'offrir une interaction gestuelle en 3 dimensions.
- Les interfaces haptiques : il est possible aujourd'hui d'avoir un retour haptique sur tous les dispositifs impliquant une interaction par la main directe avec contact.

## 5.2 Techniques d'acquisition de données dynamiques

Dans une interface naturelle, l'acquisition de données dynamiques comme le geste de la main libre, ou encore le mouvement du corps fait appel à la vision artificielle, comprenant des techniques de traitement d'images, de suivi d'objets dynamiques, et de reconstruction 3D. Depuis déjà quelques années, les techniques basées vision sont les plus utilisées, étant donné qu'elles constituent pour l'utilisateur le moyen d'acquisition le moins intrusif [15].

- Techniques de traitement d'images : permettent d'extraire à partir des images capturées les formes voulues, comme la main [22], la silhouette du corps [18], la forme des lèvres, etc. Le traitement des images passe souvent par plusieurs étapes comme le *filtrage* qui permet d'enlever le bruit lié à l'environnement de capture, la *segmentation* qui permet de regrouper les points constituant la forme recherchée, ou encore le *lissage* qui permet d'avoir des formes plus régulières.
- Techniques de suivi d'objets dynamiques : permettent d'estimer les déplacements de la forme en mouvement. Les techniques les plus souvent utilisées sont le *filtre de Kalman* [22, 42], le *filtre particulaire* [11], et les *méthodes bayésiennes* [50].
- Reconstruction 3D : permet de reconstituer la forme de l'objet suivi en 3 dimensions. La reconstruction 3D des formes dynamique est très présente dans les applications à interfaces naturelles [11, 18, 50, 15].

## 5.3 Techniques de reconnaissance et d'interprétation de commandes naturelles

Dans une interface naturelle, l'interprétation des signaux d'entrée telles que la voix, la gestuelle, et l'activité cérébrale est sujette à l'incertitude. La reconnaissance de tels signaux est probabiliste et fait appel à des techniques d'apprentissage automatique.

Les méthodes d'apprentissage automatique reposent sur des algorithmes de classification, qui visent à identifier de façon unique le type ou la classe du signal d'entrée. La traduction d'une commande naturelle n'est possible qu'une fois la classe du signal est connue.

On retrouve parmi les algorithmes les plus souvent utilisés pour la reconnaissance des signaux vocaux, cérébraux, et gesticulatoires, les réseaux de neurones [41], les modèles de Markov cachés [9, 8], l'analyse linéaire discriminante [7], et les systèmes d'inférence flous [28].

Les algorithmes de classification, appelés également *classifieurs*, passent par une phase obligatoire d'apprentissage qui vise à entraî-

ner le système sur un ensemble de données initiales, qui sont des exemples de mouvements ou de signaux auxquels seront associés les différentes actions de l'utilisateur. La problématique de la collecte de ces données possède une part importante dans la conception de telles interfaces [12]. Cela vient s'ajouter au travail du choix des différents signaux élémentaires constituant les techniques d'interaction, qui sont par exemple, dans le cas de la gestuelle, les différentes postures des mains [52], des doigts [16] ou encore du corps entier [12].

## 6 BILAN ET ANALYSE

A l'issue de notre étude portée sur les NUIs, nous constatons que les travaux existants autour de l'interaction utilisateur naturelle sont nombreux, mais en même temps, il n'existe aucun travail de synthèse autour de ce concept. De la même façon, l'évolution rapide des technologies matérielles et logicielles a conduit à la prolifération de ce type d'interfaces, sans laisser place au travail de conception. Le peu d'écrits qui existent, traitant de quelques efforts de conception, se limitent uniquement à la modalité d'interaction par le geste de la main avec ou sans contact [49, 20].

Or, l'une des principales problématiques à laquelle se trouvent confrontés les concepteurs des techniques d'interaction naturelle, aujourd'hui, consiste en le manque de formalismes et d'outils de décision, qui permettent de les guider dans leurs travail de conception et d'apporter des solutions pour certains problèmes spécifiques [20].

Afin de mener à bien notre étude visant à connaître l'apport des NUIs pour les applications de réalité virtuelle et augmentée, et parvenir à long termes, à réunir les informations nécessaires à la définition d'une base de conception et d'évaluation de l'interaction naturelle, nous identifions en premier lieu, l'ensemble des caractéristiques communes aux NUIs, puis nous les classifions selon le respect de ces caractéristique et selon la réalisation des différentes tâches d'interactions 3D universelles.

Par ailleurs, nous retenons dans ce qui suit, la définition suivante d'une NUI : « *Une interface utilisateur naturelle est le couplage des sens de l'utilisateur avec des artefacts matériels non intrusifs, constituant un moyen direct, simple et intuitif pour la transmission, l'acquisition et/ou la restitution de l'information.* ».

### 6.1 Caractéristiques communes aux NUIs

Dans un environnement virtuel ou augmenté, une interface naturelle doit permettre à l'utilisateur de retrouver les mêmes sensations d'interaction expérimentées dans le monde réel. Pour cela, une interface naturelle doit répondre à un certain ensemble d'exigences et satisfaire un certain ensemble de conditions :

1. *Sensibilité aux commandes naturelles* : c'est la capacité du système à interpréter les actions qui paraissent les plus naturelles à l'utilisateur.
2. *Facilité d'utilisation* : l'interface doit assurer une interaction intuitive permettant à l'utilisateur de prévoir de façon claire et rapide les actions à effectuer.
3. *Facilité d'apprentissage* : l'interface doit permettre à un utilisateur novice de maîtriser rapidement les interactions.
4. *Interaction directe* : possibilité de l'utilisateur à effectuer des actions au moyen d'objets et d'outils physiques et réels. Il n'est pas contraint à porter des appareils intrusifs. Il lui suffira de se servir de ses propres sens ou d'objets physiques de son propre choix.
5. *Restitution naturelle* : c'est la capacité du système à répondre à l'utilisateur d'une façon qui lui paraît la plus naturelle. Les objets visualisés dans l'espace virtuel, doivent avoir les mêmes propriétés que dans le monde réel (masse, accélération, gravité . . . , etc.), et le problème d'occlusion des objets réels en réalité augmentée doit être géré et traité.

6. *Restitution temps réel* : la réponse aux actions de l'utilisateur doit être immédiate et instantanée.
7. *Transparence* : L'environnement virtuel doit apparaître comme une extension du monde réel et la frontière séparant les deux environnements ne doit pas être visible.
8. *Souplesse et flexibilité* : dans le cas où l'interface permet une interaction simultanée d'un groupe d'utilisateurs, le fonctionnement du système ne doit pas être altéré lorsqu'un utilisateur se joint ou quitte le groupe. De même, une telle interface doit être à la fois multi-utilisateurs et mono-utilisateur. Dans le cas où l'interface est multimodale, l'utilisateur doit être libre de décider de la modalité à utiliser.

## 6.2 Classification des NUIs en deux axes

La classification des NUIs en deux axes permet de montrer le respect des différentes caractéristiques identifiées, ainsi que les tâches d'interaction 3D universelles qui sont possible à réaliser par ces interfaces.

### 6.2.1 Classification des NUIs suivant l'axe des caractéristiques

Cette première classification montre la satisfaction et le degré de satisfaction de chacune des caractéristiques identifiées par les différentes modalités d'interaction naturelle. Cette classification est non exhaustive, pour son élaboration, nous avons utilisé environ vingt publications traitant de l'exploitation des différentes modalités naturelles dans les applications de réalité virtuelle et augmentée. Nous ne prenons pas en compte dans cette classification les caractéristiques "*Restitution naturelle et Restitution temps réel*", car elles sont plus liées à l'environnement virtuel et à la technologie de l'interface qu'à l'interaction elle-même (Table 1).

TABLE 1: Classification des NUIs suivant l'axe des caractéristiques - Cn désigne la n-ième caractéristique telle que énumérée précédemment - Les signes « + » et « - » désignent respectivement la satisfaction et la non satisfaction d'une caractéristique. La répétition du signe « + » indique un degré de satisfaction élevé, et le signe « + - » indique relativement la satisfaction d'une caractéristique à un degré plus ou moins élevé.

Modalité	(C1)	(C2)	(C3)	(C4)	(C7)	(C8)
Main (tactile)	++	++	++	++	+	+
Main (tangible)	+ -	++	++	+	+	-
Main sans contact	++	+	+	++	++	+
Voix	++	+	+	++	++	+
Regard	++	++	++	++	++	+ -
Mouvement du corps	++	+ -	+	++	++	-
Activité cérébrale	++	+ -	+	+ -	+ -	-
Mouvement des lèvres	-	-	-	-	-	-
Expression du visage	-	-	-	-	-	-

### 6.2.2 Classification des NUIs suivant l'axe des tâches d'interaction 3D universelles

Cette deuxième classification vise à identifier les tâches d'interaction 3D universelles qui sont possibles à réaliser par ces interfaces naturelles, à savoir, la navigation, la sélection, la manipulation et le contrôle d'application [3, 38] (Table 2). Pour ce faire, nous avons là aussi, utilisé des articles de recherche traitant de l'interaction naturelle dans le domaine de réalité virtuelle et augmentée, et nous avons cherché pour chaque modalité, les différents travaux l'ayant chacun exploitée, pour la réalisation d'une ou plusieurs tâches différentes.

TABLE 2: Classification des NUIs suivant l'axe des tâches d'interaction 3D universelles - Les tâches sont numérotées de 1 à 4, à savoir (la navigation, la sélection, la manipulation et le contrôle d'application) - Les signes « + » et « - » indiquent respectivement la réalisation et la non réalisation d'une tâche d'interaction 3D.

Modalité	(T1)	(T2)	(T3)	(T4)
Main (tactile)	+	+	+	+
Main (tangible)	+	+	+	+
Main sans contact	+	+	+	+
Voix	+	+	+	+
Regard	-	+	+	-
Mouvement du corps	+	+	+	+
Activité cérébrale	+	+	+	-
Mouvement des lèvres	-	-	-	-
Expression du visage	-	-	-	-

## 7 CONCLUSION

L'étude présentée dans ce papier, vise à connaître l'apport des NUIs pour les applications de réalité virtuelle et augmentée. Cette étude aboutit au constat que les applications exploitant les NUIs sont multiples et en perpétuelle expansion, et que très peu de travaux s'intéressent à la conception et à la formalisation de cette nouvelle notion, ce qui rend la compréhension des NUIs difficile.

Cette étude est construite autour des travaux menés sur les NUIs, essentiellement dans le domaine de la réalité virtuelle et augmentée. Elle a servi à cerner la notion de ces interfaces qui sont souvent méconnues, d'apporter des réponses quant à leur fonctionnement, et de connaître les différents moyens d'interaction naturelle offerts à l'utilisateur. Cette étude nous a également permis d'identifier l'ensemble des caractéristiques qui sont communes aux NUIs, puis de connaître d'une part, la satisfaction par ces interfaces de chacune de ces caractéristiques bien définies, et d'autre part, les tâches d'interaction 3D universelles qui sont possibles à réaliser par ces interfaces.

Cet état de l'art nous a, par ailleurs, aidés à définir un modèle conceptuel de NUIs pour les applications de réalité virtuelle et augmentées, que nous ne décrivons pas ici. Ce modèle considère les contextes sociale et environnementale de l'interaction, afin de planifier de façon stratégique le travail de conception d'une technique d'interaction 3D naturelle, tout en permettant la vérification de l'aspect « naturel » de l'interaction.

## RÉFÉRENCES

- [1] A. Blake and M. Isard. 3D position, attitude and shape input using video tracking of hands and lips. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 185–192. ACM, 1994.
- [2] M. D. Bondy, N. D. Georganas, E. M. Petriu, D. C. Petriu, M. D. Cordea, and T. E. Whalen. Model-based face and lip animation for interactive virtual reality applications. In *Proceedings of the ninth ACM international conference on Multimedia*, pages 559–563. ACM, 2001.
- [3] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. *3D user interfaces : theory and practice*. Addison-Wesley Professional, Boston, 2005.
- [4] C. Bregler, M. Covell, and M. Slaney. Video rewrite : Driving visual speech with audio. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 353–360. ACM Press/Addison-Wesley Publishing Co., 1997.
- [5] C. Busso and S. S. Narayanan. Interrelation between speech and facial gestures in emotional utterances : a single subject study. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(8) :2331–2347, 2007.

- [6] S. Carbini, S. L. D. Picard, A. Bouguet, and J. E. Viallet. Interaction collaborative sur des objets virtuels 3D à l'aide d'une interface multimodale oro-gestuelle. 2006.
- [7] Q. Chen, N. D. Georganas, and E. M. Petriu. Hand gesture recognition using haar-like features and a stochastic context-free grammar. *Instrumentation and Measurement, IEEE Transactions on*, 57(8) :1562–1571, 2008.
- [8] P. R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow. Quickset : Multimodal interaction for distributed applications. In *Proceedings of the fifth ACM international conference on Multimedia*, pages 31–40. ACM, 1997.
- [9] A. Corradini and H.-M. Gross. A hybrid stochastic-connectionist architecture for gesture recognition. In *Information Intelligence and Systems, 1999. Proceedings. 1999 International Conference on*, pages 336–341. IEEE, 1999.
- [10] W. Ding, P. Chen, H. Al-mubaid, and M. Pomplun. A gaze-controlled interface to virtual reality applications for motor- and speech-impaired users, 2008.
- [11] Z. Feng, B. Yang, Y. Li, Y. Zheng, X. Zhao, and Q. Meng. Real-time oriented behavior-driven 3D freehand tracking for direct interaction. *Pattern Recognition*, 2012.
- [12] S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin. Instructing people for training gestural interactive systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 1737–1746. ACM, 2012.
- [13] D. Friedman, R. Leeb, A. Antley, M. Garau, C. Guger, C. Keirnath, A. Steed, G. Pfurtscheller, and M. Slater. Navigating virtual reality by thought : First steps. In *Proceedings of the 7th Annual International Workshop on Presence*, volume 160, page 167, 2004.
- [14] P. Garg, N. Aggarwal, and S. Sofat. Vision based hand gesture recognition. *World Academy of Science, Engineering and Technology*, 49(1) :972–977, 2009.
- [15] D. M. Gavrilu. The visual analysis of human movement : A survey. *Computer vision and image understanding*, 73(1) :82–98, 1999.
- [16] E. Ghomi, O. Bau, W. Mackay, and S. Huot. Conception et apprentissage des interactions tactiles : le cas des postures multi-doigts. In *FITG '10 : French workshop on tactile and gestural interaction*, 2010.
- [17] M. Haller, J. Leitner, T. Seifried, J. R. Wallace, S. D. Scott, C. Richter, P. Brandl, A. Gokcezade, and S. Hunter. The nice discussion room : Integrating paper and digital media to support co-located group meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 609–618. ACM, 2010.
- [18] J.-M. Hasenfratz, M. Lapierre, and F. Sillion. A real-time system for full body interaction with virtual worlds. In *Proceedings of the Tenth Eurographics conference on Virtual Environments*, pages 147–156. Eurographics Association, 2004.
- [19] Z. Jian, A. Cheok, K. C. Chung, et al. Real-time lip tracking for virtual lip implementation in virtual environments and computer games. In *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, volume 3, pages 1359–1362. IEEE, 2001.
- [20] F. Klompaker, V. Paelke, and H. Fischer. A taxonomy-based approach towards NUI interaction design. In *Distributed, Ambient, and Pervasive Interactions*, pages 32–41. Springer, 2013.
- [21] W. A. König, R. Rädle, and H. Reiterer. Interactive design of multimodal user interfaces. *Journal on Multimodal User Interfaces*, 3(3) :197–213, 2010.
- [22] H. Lahamy and D. Lichti. Real-time hand gesture recognition using range cameras. In *Proceedings of the Canadian Geomatics Conference, Calgary, Canada*, 2010.
- [23] W. Lai and H. Huosheng. Towards multimodal human-machine interface for hands-free control : A survey. 2011.
- [24] A. Lécuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater. Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41(10) :66–72, 2008.
- [25] J. Y. Lee, M. S. Kim, J. S. Kim, and S. M. Lee. Tangible user interface of digital products in multi-displays. *The International Journal of Advanced Manufacturing Technology*, 59(9-12) :1245–1259, 2012.
- [26] R. Leeb, D. Friedman, M. Slater, and G. Pfurtscheller. A tetraplegic patient controls a wheelchair in virtual reality. In *BRAINPLAY 07 Brain-Computer Interfaces and Games Workshop at ACE (Advances in Computer Entertainment) 2007*, page 37, 2007.
- [27] C. Lim and D. Kim. Development of gaze tracking interface for controlling 3D contents. *Sensors and Actuators A : Physical*, 2012.
- [28] F. Lotte, A. Lécuyer, Y. Renard, F. Lamarche, B. Arnaldi, et al. Classification de données cérébrales par système d'inférence flou pour l'utilisation d'interfaces cerveau-ordinateur en réalité virtuelle. In *Ières journées de l'AFRV*, 2006.
- [29] A. M. MacEachren, G. Cai, R. Sharma, I. Rauschert, I. Brewer, L. Bolelli, B. Shaparenko, S. Fuhrmann, and H. Wang. Enabling collaborative geoinformation access and decision-making through a natural, multimodal interface. *International Journal of Geographical Information Science*, 19(3) :293–317, 2005.
- [30] F. Marton, M. Agus, E. Gobbetti, G. Pintore, and M. Balsa Rodriguez. Natural exploration of 3D massive models on large-scale light field displays using the FOX proximal navigation technique. *Computers & Graphics*, 2012.
- [31] S. McGlashan. Speech interfaces to virtual reality. In *Proceedings of 2nd International Workshop on Military Applications of Synthetic Environments and Virtual Reality*. Citeseer, 1995.
- [32] S. McGlashan and T. Axling. A speech interface to virtual environments. *Swedish Institute of Computer Science*, 1996.
- [33] M. R. Morris, D. Morris, and T. Winograd. Individual audio channels with single display groupware : effects on communication and task strategy. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 242–251. ACM, 2004.
- [34] S. Murphy. Design considerations for a natural user interface(NUI). *White paper*, 2012.
- [35] P. Nugues, C. Godéreaux, P.-O. El Guedj, and F. Revolta. A conversational agent to navigate in virtual worlds. *Luperfoy et al.* 1996.
- [36] K. O'hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor. On the naturalness of touchless : putting the interaction back into nui. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1) :5, 2012.
- [37] K. O'Hara, A. Sellen, and R. Harper. Embodiment in brain-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 353–362. ACM, 2011.
- [38] S. Otmane. *Modèles et techniques logicielles pour l'assistance à l'interaction et à la collaboration en réalité mixte*. Habilitation à diriger les recherches, Université d'Evry-Val d'Essonne, 2010.
- [39] S. Oviatt and P. Cohen. Perceptual user interfaces : multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3) :45–53, 2000.
- [40] T. Pfeiffer. Towards gaze interaction in immersive virtual reality : Evaluation of a monocular eye tracking set-up. In *Virtuelle und Erweiterte Realität-Fünfter Workshop der GI-Fachgruppe VR/AR*, pages 81–92, 2008.
- [41] P. Premaratne and Q. Nguyen. Consumer electronics control system based on hand gesture moment invariants. *Computer Vision, IET*, 1(1) :35–41, 2007.
- [42] D. W. Seo and J. Y. Lee. Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. *Expert Systems with Applications*, 40(9) :3784 – 3793, 2013.
- [43] M. Siggelkow. *Importance of gaze awareness in augmented reality teleconferencing*. PhD thesis, Masters thesis, Technische Universität München, 2005.
- [44] P. Song, H. Yu, and S. Winkler. Vision-based 3d finger interactions for mixed reality games with physics simulation. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, page 7. ACM, 2008.
- [45] P. Song, H. Yu, and S. Winkler. Vision-based 3d finger interactions for mixed reality games with physics simulation. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, page 7. ACM, 2008.
- [46] N. A. Streit, J. Gei, T. Holmer, S. Konomi, C. Müller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz. i-land : an interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 120–127, New York, NY, USA, 1999. ACM.
- [47] R. Vertegaal. The gaze groupware system : mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages

294–301. ACM, 1999.

- [48] M. Weiss, J. D. Hollan, and J. Borchers. Augmenting interactive tabletops with translucent tangible controls. In *Tabletops-Horizontal Interactive Displays*, pages 149–170. Springer, 2010.
- [49] D. Wigdor and D. Wixon. *Brave NUI world : designing natural user interfaces for touch and gesture*. Elsevier, 2011.
- [50] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfindex : Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7) :780–785, 1997.
- [51] C. R. Wren, F. Sparacino, A. J. Azarbayejani, T. J. Darrell, T. E. Starner, A. Kotani, C. M. Chao, M. Hlavac, K. B. Russell, and A. P. Pentland. Perceptive spaces for performance and entertainment untethered interaction using computer vision and audition. *Applied artificial intelligence*, 11(4) :267–284, 1997.
- [52] M. Wu and R. Balakrishnan. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 193–202. ACM, 2003.
- [53] R. C. Zeleznik, A. S. Forsberg, and J. P. Schulze. Look-that-there : Exploiting gaze in virtual reality interactions. Technical report, Technical Report CS-05-04, Brown University, Department of Computer Science, 2005.
- [54] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 246–253. ACM, 1999.