



HAL
open science

KIVA : environnement virtuel informé pour la formation au geste technique

Florian Jeanne, Indira Thouvenin

► **To cite this version:**

Florian Jeanne, Indira Thouvenin. KIVA : environnement virtuel informé pour la formation au geste technique. 10èmes journées de l'Association Française de Réalité Virtuelle, Augmentée, Mixte et d'Interaction 3D, Oct 2015, Latresne, France. hal-01228471

HAL Id: hal-01228471

<https://hal.science/hal-01228471>

Submitted on 3 Dec 2015

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.

KIVA : environnement virtuel informé pour la formation au geste technique

Florian Jeanne *

UMR CNRS 7253 Heudiasyc

Université de Technologie de Compiègne

Indira Thouvenin †

UMR CNRS 7253 Heudiasyc

Université de Technologie de Compiègne

RÉSUMÉ

Le projet KIVA (*Knowledge and Informed Virtual environment for gesture cApitalization*) vise à créer un nouvel environnement de formation au geste technique par le biais de la réalité virtuelle. Il se décompose en un module de reconnaissance de gestes, AN-GORA, et un module de décision, ALPAGA, proposant des retours sensoriels à l'apprenant. Grâce à une interaction informée en environnement virtuel, notre approche permet un guidage adaptatif personnalisé à l'apprenant, en fonction du contexte utilisateur et de la situation dans laquelle il se trouve.

Keywords: Réalité virtuelle, guidage, geste technique, retours sensoriels

1 LE GESTE

Mouvement du corps, principalement de la main, des bras, de la tête, porteur ou non de signification.

ou

Manière de mouvoir le corps, les membres et, en particulier, manière de mouvoir les mains dans un but de préhension, de manipulation.

Cette définition du dictionnaire Larousse montre que le geste n'est pas nécessairement réalisé par la ou les main(s), mais plutôt par une partie quelconque du corps, ou dans certain cas l'ensemble du corps.

Dans notre vie de tous les jours, les gestes font partie intégrante de notre communication, qu'ils soient intentionnels ou non. Les gestes peuvent notamment, remplacer la parole, enrichir une information en l'illustrant, ou tout simplement renseigner sur l'état émotionnel du locuteur.

1.1 Le geste technique

Dans [6], Bril définit le geste technique comme étant une habileté acquise par apprentissage permettant la réalisation d'une tâche orientée vers un but spécifique. Cet apprentissage peut créer des automatismes permettant le contrôle du geste "à un niveau infra conscient" [34].

La notion de geste professionnel ou de métier, selon Le Bellu [20], renvoie à l'idée d'une expertise acquise au cours du temps. Le geste se décompose alors selon trois composantes qui sont les opérations physiques (manuelles et kinesthésiques), attentionnelles et cognitives. Ainsi, le geste technique correspond à la maîtrise d'une action, d'une ressource matérielle ou immatérielle. Le geste professionnel quant à lui, est plus conséquent et plus riche puisqu'il comprend l'histoire d'un métier. Il s'est construit au cours du temps.

Aussi, Tomás différencie le geste de métier [34]. Il comprend une manière de faire qui est considérée comme acceptable et acceptée par les professionnels du milieu correspondant. Ces manières de

faire peuvent alors être apparentées à "une mémoire professionnelle collective" qui dirige et régule l'activité.

Les gestes sont au cœur de nos actions que ce soit d'un point de vue professionnel ou non. Différentes approches caractérisent et classent les gestes selon leur signification, leur description, leur lien à la parole, leur relation avec l'environnement ou tout simplement selon leur finalité.

1.2 Sémantique du geste

Il existe différentes classifications sémantiques du geste associé à la parole. Kendon [18], [19] analyse le geste selon sa symbolique. Cependant, il ne s'intéresse qu'aux gestes réalisés avec les bras et les mains, dans le cadre de la communication. Il élabore alors un *continuum* :

- Gesticulation (*gesticulation*) : geste spontané (main, bras, tête) qui accompagne la parole ;
- Gestes quasi-linguistiques (*speech-framed gestures*) : le geste complète la structure de la parole (e.g., "Rémy est allé par là [en pointant du doigt]) ;
- Pantomimes (*pantomimes*) : gestes ou séquence de gestes, sans paroles, qui ont une ligne narrative, une histoire à raconter (souvent utilisés au théâtre) ;
- Emblèmes (*emblems*) : gestes conventionnels, d'éloges ou d'insultes (e.g., pouce vers le haut) ;
- Langages des signes (*sign languages*) : les gestes permettant une communication linguistique complète sans paroles.

McNeill reprend ces travaux en 1992 [22], mais propose à la place une classification où le geste possède un sens plus large et n'est plus rattaché au discours. Il peut alors représenter une action que le locuteur simule, ou encore refléter ses sentiments ou ses désirs. Cette classification comprend alors :

- Les gestes iconiques (*iconic*) : ils décrivent concrètement une entité, un objet ou une action ;
- Les gestes métaphoriques (*metaphoric*) : représentation d'une idée abstraite ;
- Les gestes déictiques (*deictic*) : désignation d'un objet, d'un lieu ou même d'une personne, à l'aide d'une partie de son corps (doigt, main ou encore tête) ou d'un objet tenu ;
- Les pulsations ou battements (*beats*) : gestes qui rythment le discours, sans sens propre particulier.

En 2004, Kendon s'appuie sur la démarche de McNeill et complète son *continuum* en y incluant la parole (fig. 1). On y voit décroître la présence de la parole en le parcourant de la gesticulation vers la langue des signes. En effet, elle est absente dès la pantomime, mais n'est pas nécessaire pour la compréhension.

McNeill propose également une segmentation temporelle des gestes du bras sous la forme suivante :

Pre-stroke - Stroke - Post-stroke

Ici, le geste a besoin d'une préparation (pre-stroke). Il doit atteindre une position et avoir une orientation spécifique dans l'espace avant d'être effectué. De même, la phase de rétraction correspond au moment où le bras (ou plus simplement la main) retourne à sa position de repos. Ces phases peuvent être omises en fonction des actions précédentes ou suivantes. De plus, si nous comparons cette segmentation à l'écriture, les phases de pre-stroke et de post-stroke peuvent être assimilées aux espaces entre les mots.

*e-mail: florian.jeanne@hds.utc.fr

†e-mail: indira.thouvenin@utc.fr

Gesticulation

Gestes quasi-linguistiques

Pantomines

Emblèmes

Langages des signes

Présence de la parole

Absence de la parole

FIGURE 1: *Continuum* de Kendon

1.3 Description du geste

Dans certains domaines, comme le sport ou l'écriture par exemple, la forme du geste peut également être un sujet d'étude intéressant. On y distingue la topocinèse et la morphocinèse. [11] définissent une théorie nommée *embodied of integrative theory of gesture*, qui définit l'intégration des théories motrices et communicatives du geste. Elle distingue et identifie notamment les rôles joués par les processus moteurs et les processus communicatifs du geste :

- La topocinèse est le contrôle moteur, dépendant principalement de la proprioception ;
- La morphocinèse correspond au contrôle communicatif, lié à la sémantique du geste.

Cependant, Serre [30] avait quant à lui, défini différemment ces processus. La sémantique du geste n'est plus comprise dans la morphocinèse, mais est rattachée à une troisième catégorie, la sémiocinèse.

- La topocinèse est une activité motrice dont le but est d'atteindre un point de l'espace, et où le moyen est un mouvement pour atteindre un point, comme dans les sports collectifs par exemple ;
- La morphocinèse est une activité motrice dont le but est de produire une forme conforme à un code, et où l'espace, voire le temps, sont des moyens pour produire cette forme. Nous prendrons pour exemple la gymnastique et l'écriture ;
- La sémiocinèse est une forme corporelle pour produire du sens et des émotions chez les spectateurs, comme dans le cadre de la danse.

1.4 Fonction du geste

Cadoz [7] quant à lui, classe les gestes par rapport à leur fonction et à leur rapport avec l'environnement. Il distingue notamment différentes fonctions :

- La fonction ergodique qui concerne la manipulation, la régulation ou la modification de l'environnement physique ;
- La fonction épistémique qui consiste à percevoir l'information depuis l'environnement (e.g., le cas d'une exploration haptique) ;
- la fonction sémiotique qui permet la communication d'informations vers l'environnement.

L'utilisateur agit sur l'environnement et celui-ci lui effectue un retour sensoriel. La majorité des interactions seraient donc de nature sémiotique. Cependant, les retours sensoriels semblent jouer un rôle primordial dans le comportement de l'utilisateur, ce qui relèverait plutôt d'une fonction ergodique. Cette classification introduit la notion de relation entre l'utilisateur et l'environnement dans lequel il évolue (environnement virtuel dans notre cas).

1.5 Système éactif

Le biologiste Varela propose dans ses travaux [37] la notion de l'éaction qui est l'adaptation du comportement d'un individu, en réaction à l'évolution de l'environnement dans lequel il se situe. Ce paradigme est basé sur le principe biologique de l'autopoïèse qui vise à définir l'être vivant. Varela définit ce principe ainsi [35], [36] :

Un système autopoïétique est organisé comme un réseau de processus de production de composants qui régèrent continuellement par leurs transformations et leurs interactions le réseau qui les a produits, et qui constituent le système en tant qu'unité concrète dans l'espace où il existe, en spécifiant le domaine topologique où il se réalise comme réseau. Il s'ensuit qu'une machine autopoïétique engendre et spécifie continuellement sa propre organisation. Elle accomplit ce processus incessant de remplacement de ses composants, parce qu'elle est continuellement soumise à des perturbations externes, et constamment forcée de compenser ces perturbations. Ainsi, une machine autopoïétique est un système à relations stables dont l'invariant fondamental est sa propre organisation (le réseau de relations qui la définit).

Dans le paradigme de l'éaction, l'humain, qui interagit avec son environnement, perçoit des sensations en retour de ses actions, et va donc adapter son comportement et/ou ses gestes. Thouvenin rappelle néanmoins que ce n'est pas l'interface qui est éactive, mais bien l'humain qui rend le monde éactif par le biais d'une interface [33]. Dans [13], De Loor souligne que la réalité virtuelle est propice à cela puisqu'il y est possible de modéliser une boucle éactive comprenant le système et l'interaction entre l'humain et ce système.

2 ENVIRONNEMENTS DE FORMATION

2.1 Environnements virtuels informés

Selon Bowman [5], un environnement virtuel est un monde artificiel généralement en trois dimensions, observé depuis une vue à la première personne et contrôlée par l'utilisateur en temps réel.

Les environnements virtuels informés (ou EVI) sont un type particulier d'environnements virtuels. Différentes définitions ont été proposées.

Aylett et Luck [2] définissent en 2000 les environnements virtuels intelligents comme étant la combinaison de l'intelligence artificielle, de la vie artificielle et de la réalité virtuelle, intégrée dans des créatures ou agents autonomes qui interagissent entre eux.

Bowman quant à lui, propose en 2003 la notion d'*information-rich virtual environment* ou IRVE [5]. Les IRVE sont des environnements virtuels réalistes enrichis par l'ajout d'informations absentes relatives au contenu du monde. Ici, l'information est représentée sous forme de texte. Bowman propose ainsi une taxonomie de la représentation des textes dans les environnements virtuels (fig. 2).

Donikian introduit en 2004 la notion d'environnement virtuel informé (EVI). Un EVI est un environnement virtuel dont les modèles 3D contiennent non seulement la géométrie de la scène mais aussi toutes les informations pertinentes pour les entités comportementales à simuler comme les éléments symboliques ou sémantiques qui peuvent permettre à ces entités de percevoir, décider et agir [14].

Enfin, Thouvenin [33] considère un EVI comme étant un environnement virtuel doté de modèles à base de connaissances dans lequel il est possible à la fois d'interagir et de permettre des comportements par interprétation de représentations dynamiques ou sta-

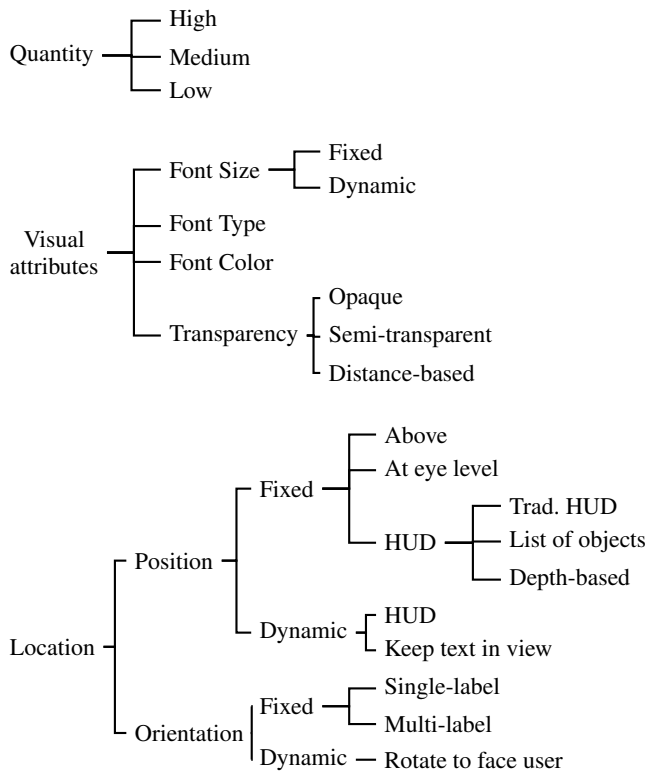


FIGURE 2: Taxonomie des textes en environnements virtuels - Bowman

tiques. Cette définition élargit le concept d'EVI en intégrant l'utilisation de modèles de représentation des connaissances comme les bases de connaissances, les ontologies, les arbres ou encore les modèles de raisonnement.

Ainsi, les EVI permettent de représenter les connaissances dans les environnements virtuels. Ces connaissances peuvent être communiquées à l'utilisateur comme dans le cadre de la formation. Aussi, certaines informations provenant de ce même utilisateur peuvent également être interprétées comme des connaissances et classifiées par le système.

2.2 Apprentissage en environnement virtuel

Pour Fuchs, "la finalité de la réalité virtuelle est de permettre à une personne (ou plusieurs) une activité sensori-motrice et cognitive dans un monde artificiel, créé numériquement, qui peut être imaginaire, symbolique ou une simulation de certains aspects du monde réel" [16]. L'idée principale ici est que l'interaction de l'utilisateur lui donne accès à un tout autre univers ; cette *activité sensori-motrice*, induit que la personne agit dans cet environnement et perçoit les différents éléments qui le composent. En nous intéressant à son application, nous nous apercevons que la réalité virtuelle est de plus en plus utilisée dans le cadre de l'apprentissage [1] [8] [15] [17] [27]. Elle permet de former les personnes dans des cadres très différents, comme le geste de soudure [31], les gestes obstétriques [24] ou encore la formation scénarisée [10].

2.2.1 Formation au transport fluvial

Dans le cadre de la formation au transport fluvial, Fricoteaux propose une interaction informée par un système décisionnel pour un retour multimodal en environnement virtuel [15] (fig. 3). Ici, le module décisionnel a pour objectif de renforcer le couplage entre l'humain et l'environnement. Pour cela, il adapte les retours mul-

timodaux envoyés à l'utilisateur en fonction du comportement de celui-ci. Ils peuvent prendre la forme d'aides visuelles, de messages sonores et de simplifications des conditions de navigation afin de l'aider à anticiper les manœuvres à effectuer.

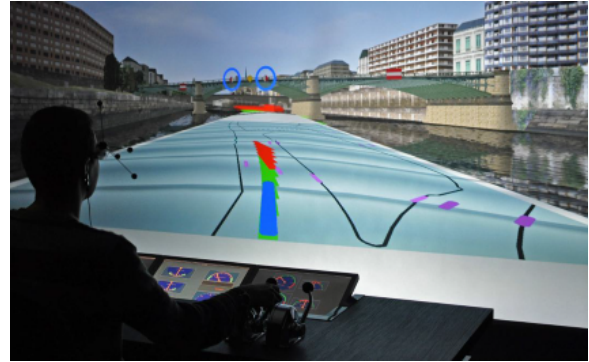


FIGURE 3: Formation au transport fluvial [15]

2.2.2 Scénarisation personnalisée

Carpentier propose un système de formation scénarisé adaptatif [8]. Au niveau le plus bas, l'adaptation doit moduler les conséquences des actions de l'apprenant. La suite d'évènements ou d'actions engendre des résultats différents selon le niveau d'expertise de l'utilisateur. Ensuite, l'adaptation a également lieu dans le processus de production de séquences complexes d'évènements menant à une situation d'apprentissage particulière. Enfin, l'adaptation est censée guider l'apprenant à travers les différentes étapes et situations d'apprentissage entre chaque session. Le cas d'étude de ces travaux était le projet NIKITA (*Natural Interaction, Knowledge and Immersion in Training for Aeronautics*) qui a pour but la formation d'opérateurs dans l'assemblage d'avions, en appliquant le modèle de génération de scénarios proposé (fig. 4).



FIGURE 4: Screenshot du projet NIKITA [9]

2.2.3 Formation aux gestes obstétriques

Moreau quant à lui, propose une méthode pour évaluer l'acquisition de gestes obstétriques lors d'une formation par la réalité virtuelle [24]. Il s'intéresse particulièrement à l'analyse des gestes de l'apprenant en les comparant à un geste expert. Pour cela, un système de métaphores visuelles, à savoir des sphères, était utilisé pour symboliser la trajectoire idéale des forceps (fig. 5). Ce chemin idéal correspondait à celui emprunté par un expert, l'apprenant devant le suivre pour réaliser le meilleur geste possible.

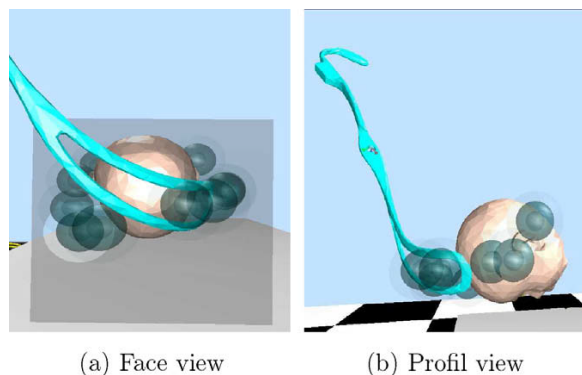


FIGURE 5: Sphères de guidage pour la formation aux gestes obstétricaux [24]

2.2.4 Le geste de soudage

L'apprentissage du soudage rencontre différentes difficultés au niveau de l'activité gestuelle et de l'environnement de travail. Les gestes sont fins, méticuleux et répétitifs. De plus au niveau de la charge cognitive, ils demandent une grande concentration afin de maîtriser les gestes de ses mains mais aussi du reste de son corps. L'environnement de travail peut être difficile à cause de la dangerosité du processus, de la chaleur, de l'éblouissement ou encore de la fatigue musculaire. De plus, "*le débutant doit apprendre à maîtriser les différents paramètres qui déterminent la qualité d'une soudure (vitesse d'avance, rectitude, distance buse-tôle, angle d'avance et angle bissecteur) sans repère visuel (éblouissement et étincelles)*" [23].

CS WAVE est une application de formation aux gestes de soudage [23] [31]. Elle comprend un environnement virtuel qui sert de support à l'interaction gestuelle. En effet, l'utilisateur apprend à souder sur un écran en réalisant les nécessaires à une bonne soudure et qui sont attendus d'un professionnel en situation réelle (fig. 6).



FIGURE 6: CS Wave - wave-ng.com [31]

2.2.5 Réapprendre les gestes du quotidien

Il est également possible de reformer des personnes aux gestes du quotidien, notamment celles qui ont des déficiences cognitives suite à des lésions cérébrales ou des maladies [17]. Avec la réalité virtuelle il est possible de former l'apprenant dans de bonnes conditions, dans un environnement sûr et contrôlé. Il est également possible de suivre sa formation à distance et d'intervenir si besoin dans le processus sans être à côté de la personne.

Ainsi, dans l'application de Gourlay [17], le patient doit déplacer des objets dans l'espace en utilisant notamment un gant de données, le VR glove. Chaque objet doit être positionné dans le bon emplacement (fig. 7). Une cuisine virtuelle est également implémentée et permet de réaliser les gestes du quotidien comme remplir une tasse avec du café par exemple. Les troubles de concentration et d'autres troubles de l'attention peuvent alors être surveillés et enregistrés.

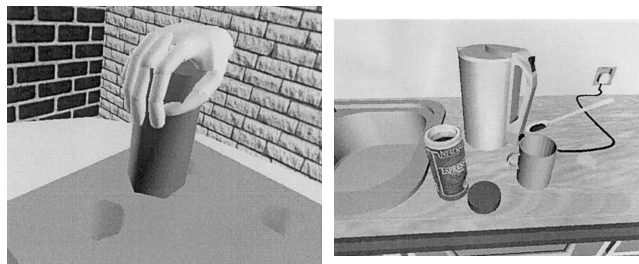


FIGURE 7: Exercice de déplacement dans l'espace et cuisine virtuelle [17]

3 KIVA

3.1 Le projet

KIVA est un acronyme pour *Knowledge and Informed Virtual environment for gesture cAPitalization*. Le projet ambitionne de créer un nouvel environnement de formation utilisant la réalité virtuelle afin d'accélérer le processus d'apprentissage de gestes techniques. Il s'inscrit dans la forte dynamique de la formation par la réalité virtuelle, qui est de plus en plus utilisée dans l'industrie. Le groupe Montupet, membre du projet et un des leaders mondiaux dans la conception et la fabrication de pièces automobiles en aluminium, cherche à améliorer son efficacité et à réduire ses taux de rebuts, qui sont des conséquences directes de la formation. En effet, la fabrication des culasses (partie supérieure d'un moteur à pistons), nécessite une haute technicité dans le geste. La formation en environnement virtuel permettra de restituer et d'explicitier un savoir-faire essentiellement tacite.

3.1.1 Cas d'étude

Le cas d'étude comprend trois postes clés du processus de fabrication de culasses chez Montupet, à savoir l'ébavurage et le contrôle d'un noyau au poste de noyautage, le soufflage du moule au poste de moulage, et enfin le poste de contrôle visuel final.

Poste de noyautage Le noyautage est l'atelier de la fonderie où sont fabriqués les noyaux, composés de sable et de liant, qui seront insérés par la suite dans le moule, afin de générer les conduits internes de la culasse (fig. 8).

Le mélange de sable et de résine est versé dans un moule. Une fois ce noyau moulé, il est récupéré par l'opérateur qui effectue alors un contrôle visuel pour vérifier qu'il n'y ait pas d'anomalies. Si la pièce est conforme, il procède à un ébavurage du noyau pour éliminer les éventuels surplus et dépôts de sable. La pièce est ensuite placée sur un charriot afin d'être conduite à l'étape suivante de la chaîne.

Poste de moulage Dans l'industrie de la fonderie, l'opérateur à ce poste se certifie de la conformité de l'équipement avant la prochaine coulée (fig. 9).

Le poste de moulage chez Montupet est en partie situé sur un carrousel qui comprend 6 moules chauffés à températures élevées. L'opérateur agit sur un des moules, pendant que les autres subissent des opérations robotisées. L'une de ses missions de nettoyer le moule en enlevant les morceaux de noyaux de la pièce précédente qui sont éventuellement restés, en soufflant le sable restant et en



FIGURE 8: Un noyau après ébavurage (Montupet)



FIGURE 9: Soufflage d'un moule (Montupet)

enlevant les morceaux de métal qui ont pu rester accrochés. Cette étape intervient juste avant la coulée de l'aluminium liquide, et à la suite du prélèvement d'une pièce suffisamment refroidie pour que sa solidification soit suffisamment avancée avant les traitements à venir.

Poste du contrôle visuel final Le contrôle final est le troisième et dernier poste étudié dans le cadre du projet KIVA (fig. 10). Il est également le dernier poste du cycle de production et le dernier repart qualité avant l'envoi des culasses chez le client.



FIGURE 10: Contrôle visuel de la qualité de la pièce (Montupet)

Ici, l'opérateur suit une procédure de contrôle visuel et dimensionnel. Il détermine la correspondance des diamètres et mesures

des pièces avec celles définies dans le cahier des charges client, et si la pièce ne présente aucun défaut. Ce contrôle visuel est un gage de qualité nécessaire et exigé par les clients de Montupet.

3.1.2 L'environnement de formation de KIVA

Afin de former les opérateurs de la meilleure des façons, les postes sont en cours de numérisation et de modélisation dans l'environnement virtuel par une des partenaires du projet, Reviatch. L'objectif est d'avoir un environnement visuellement proche de la réalité pour qu'une fois la formation terminée, les apprenants soient déjà familiarisés avec leur poste de travail. C'est un point non négligeable puisque l'environnement réel contient un certain nombre de contraintes comme la chaleur, le bruit ou la posture de travail. Nous visons donc à mettre l'opérateur dans les meilleures conditions possibles une fois sa formation achevée.

Lors de la formation, les gestes de l'utilisateur sont capturés par un système de capture de mouvements. Nous utilisons des caméras infra-rouges couplées à des marqueurs passifs réfléchissants (fig. 11 - Image provenant du site creagames.fr) positionnés sur l'apprenant (la tête, les membres et le tronc). Ainsi, nous pouvons obtenir la position et l'orientation des différentes parties de son corps à chaque instant. Cela nous permet alors de déterminer l'action qu'il est en train de faire et d'évaluer sa gestuelle.

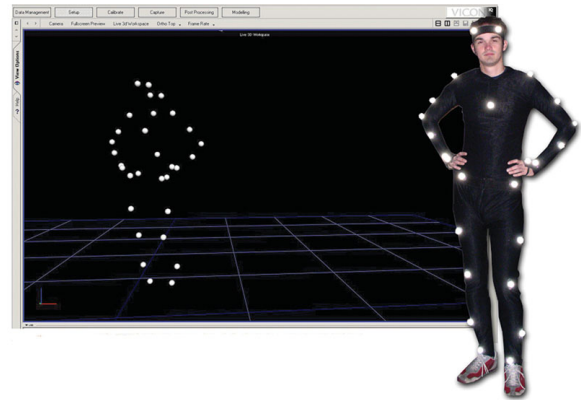


FIGURE 11: Capture de mouvements humain - creagames.fr

Les données récupérées vont nous permettre de travailler sur le geste en lui-même (e.g. nettoyage avec la soufflette à un endroit précis) ou sur une séquence de gestes (e.g. nettoyage complet du moule). Elles sont capturées et traitées en temps réel afin de fournir des retours, sur la formation, pendant l'exécution des gestes.

3.2 ANGORA

Partie intégrante du système intelligent développé par le laboratoire Heudiasyc dans le cadre du projet KIVA, le module ANGORA est un système de reconnaissance automatique de gestes permettant à la fois de qualifier le geste effectué et de quantifier sa performance. De nombreux travaux de recherche se sont intéressés à la reconnaissance de gestes, notamment par l'emploi de méthodes de correspondance (kNN, DTW,...) [38] [29] [32] ou de modèles d'apprentissage statistique (HMMs, CRFs, Neural Networks, SVMs, ...) [3] [25] [39] [26] [4]. En particulier, divers chercheurs se sont récemment penchés sur la reconnaissance de gestes techniques sur une chaîne d'assemblage [12] ou dans le cadre de la poterie [21].

Le module ANGORA doit faire face à plusieurs difficultés. Tout d'abord, le système doit pouvoir informer l'utilisateur en temps réel sur sa performance afin notamment qu'il puisse rectifier immédiatement son geste. Le module de reconnaissance doit donc permettre une évaluation du geste de l'utilisateur (ou de la séquence

de gestes) pendant son exécution. De plus, l'ensemble des méthodes décrites ci-dessus requièrent une base d'exemples composée de gestes types. L'acquisition d'une telle base nécessite la participation d'un ou plusieurs experts pour valider les données ce qui ne peut se faire généralement qu'à petite échelle. Enfin, les résultats en sortie du module ANGORA sont fournis en entrée d'un module de décision inclu dans ALPAGA. Ces sorties doivent donc être interprétables par ce module.

Pour répondre à ces problématiques, notre intérêt s'est porté sur les modèles de Markov à états cachés (HMMs, [28]). En effet, les HMMs sont des modèles de séquences, ils sont donc définis pour traiter des données temporelles de longueur variable et sont applicables en temps réel. De plus, la phase d'apprentissage des paramètres du modèle possède une extension intuitive à un cadre d'apprentissage semi-supervisé. Ceci permet de potentiellement renforcer les capacités du modèle en prenant en compte des données dites non-étiquetées lors de l'apprentissage, c'est à dire des données non validées par un expert. Enfin les sorties d'un HMM sont des scores probabilités et sont donc interprétables par le module de décision.

3.3 ALPAGA

Le module ALPAGA de KIVA est la partie qui traite de la pédagogie. Il comprend le modèle décisionnel qui s'occupe de l'annotation du geste et du choix des retours sensoriels à émettre à l'apprenant.

Tout d'abord, le module récupère les données pré-enregistrées de l'utilisateur, comme son niveau d'expertise par exemple. Lorsqu'un geste est effectué, il communique avec l'environnement virtuel afin de récupérer le contexte de celui-ci. Il obtient donc la liste des objets manipulés pendant le geste et leur localisation dans l'espace. Ces informations sont transmises au module ANGORA et serviront pour l'évaluation du geste.

Le module ANGORA transmet par la suite un score du geste réalisé au module ALPAGA. Le modèle décisionnel annote alors ce geste. Il le contextualise, lui donne un sens dans une situation donnée, puis détermine les retours sensoriels pertinents qu'il doit envoyer à l'utilisateur (fig. 12).

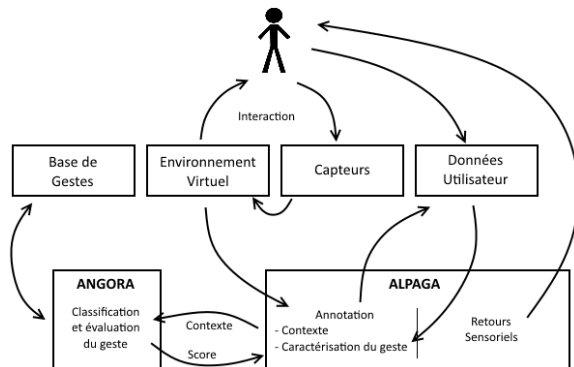


FIGURE 12: Architecture du système pour la caractérisation du geste dans le cadre du projet KIVA

Retours sensoriels Ces retours sensoriels peuvent servir par exemple pour le guidage de l'apprenant dans certaines situations ou tout simplement des commentaires textuels sur sa gestuelle.

Ces derniers peuvent être affichés en temps réel à l'utilisateur afin de lui donner des indications sur sa gestuelle, lui permettant de se corriger directement. Ces commentaires peuvent dans certains cas, par exemple pour un utilisateur un peu plus expérimenté, s'effectuer *a posteriori*. Il n'a alors aucune aide et n'a des retours sur ces gestes qu'à la fin de l'exercice.

Concernant le guidage, des métaphores visuelles auront pour but d'alerter l'apprenant lorsqu'il s'éloigne du chemin optimal ou de la tâche à effectuer. Cette métaphore serait basée sur le principe de l'excitation de particules. Si les particules sont agitées, cela signifie que l'utilisateur n'effectue pas son action correctement ou est éloigné de la trajectoire idéale, alors qu'au contraire, si les particules sont dans un état stable, il est sur la bonne voie.

Adaptation Notre approche vise à être adaptative. En effet, le module de reconnaissance de gestes, comparera celui de l'apprenant à un geste expert et lui attribue un score. Cependant, ce score ne sera pas nécessairement satisfaisant. En effet, le système décisionnel (fig. 13), en annotant le geste, visera à ajuster l'évaluation du geste en tenant compte des informations de l'utilisateur, comme son niveau d'expertise par exemple. Les retours sensoriels proposés à l'utilisateur seront également adaptatifs. Le système s'adaptera à l'évolution de l'apprenant et proposera différents retours selon la situation.

De plus, le système possèdera une partie adaptable pour corriger l'ignorance du système. Il est possible que le système de reconnaissance hésite par exemple entre deux gestes. L'utilisateur sera alors questionné et renseignera le système sur la nature de son geste.

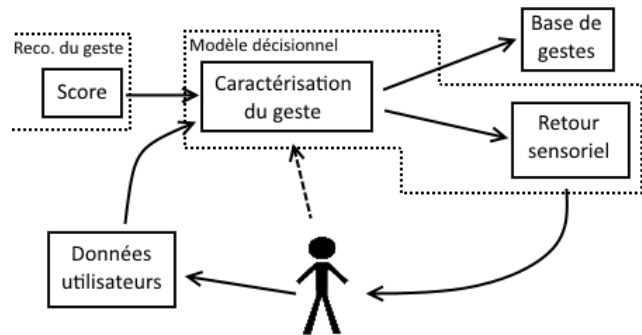


FIGURE 13: Modélisation du système décisionnel pour ALPAGA

4 EXPÉRIMENTATION

Nous avons développé un premier scénario d'usage dans le but de récolter une première série de données sur des gestes. Il comprenait différents exercices basés sur des gestes effectués sur le poste de moulage à savoir gratter des amas de métal à l'aide d'un grattoir, ainsi que de disperser et d'enlever du moule des résidus de métal et de sable à l'aide d'une soufflette à air comprimé.

Le dispositif comprenait un vidéoprojecteur qui affichait une scène en stéréoscopie active (fig. 14) sur une surface plane. L'utilisateur était équipé de différents marqueurs situés sur les lunettes, l'épaule droite, le coude et la main (fig. 15 et 16). Ces marqueurs nous permettaient de déterminer la position et l'orientation des parties du bras à chaque instant, et donc des mouvements effectués.

Les différents outils possédaient également un jeu de marqueurs afin de capturer leurs mouvements dans l'environnement et de gérer les conséquences de leur utilisation (à savoir retirer les tâches à l'aide du grattoir et disperser les résidus avec la soufflette). Sur la figure 14, le grattoir est représenté par une forme rectangulaire rouge. Cet outil permet de retirer la tache représentée par un cercle noir. Les trois autres taches présentes sur cette capture d'écran représentent du sable et des résidus de métal que l'utilisateur doit enlever à l'aide de la soufflette.

L'expérience comprenait une quarantaine d'exercices. Pour les résoudre l'utilisateur devait réaliser des gestes différents selon ce qui était demandé à l'aide de l'outil adéquat. Lors de certains exercices aucun geste n'était imposé alors que pour certains, il devait déplacer la soufflette de haut en bas par exemple. Aussi, sur un des

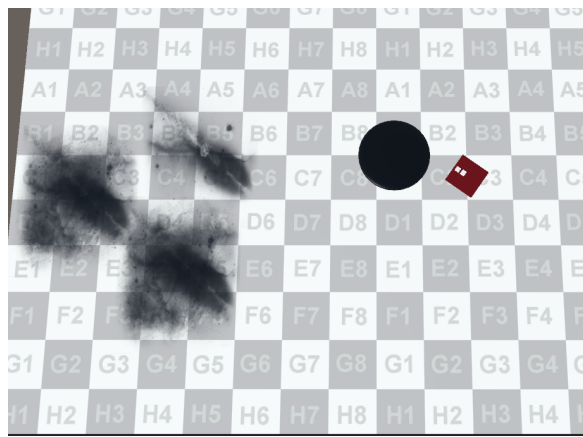


FIGURE 14: Capture d'écran de la scène projetée

types de gestes, nous avons demandé aux utilisateurs d'effectuer des erreurs dans leur gestuelle. Les gestes imposés nous ont permis d'avoir un geste commun à tous les utilisateurs.

Nous avons effectué cette expérimentation avec 15 personnes différentes. Les résultats nous ont permis d'évaluer les différents modèles de reconnaissance de gestes que nous avons sélectionnés, à la fois pour une tâche de classification et pour une évaluation de l'écart avec un geste que nous considérons comme expert. De plus, cela nous aidera pour la définition d'un protocole expérimental et de futures expériences avec des opérateurs experts au niveau des postes étudiés. Nous avons également pu établir une première typologie des erreurs effectuées. Nous distinguons les erreurs effectuées lors d'un geste de celles effectuées lors d'une séquence de geste. Les plus courantes sont des erreurs au niveau de l'angle d'incidence entre l'outil et le plan (sa valeur devant être comprise dans un intervalle spécifique pour que l'opération soit efficace), l'utilisation d'un mauvais outil pour une tâche donnée et enfin l'oubli d'une action à effectuer.



FIGURE 15: Utilisateur équipé du grattoir

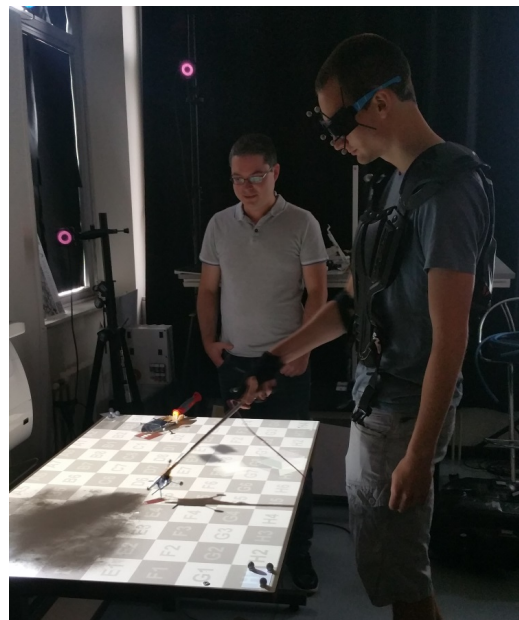


FIGURE 16: Utilisateur équipé de la soufflette à air comprimé

5 CONCLUSION

Le projet KIVA comporte ainsi deux modules. Le premier, AN-GORA, est un système de reconnaissance automatique de geste et est basé sur des méthodes d'apprentissage statistique. Le résultat de cette reconnaissance est fourni en entrée d'un second module, ALPAGA, qui vise à l'associer avec le contexte utilisateur afin de fournir des retours sensoriels personnalisés à l'apprenant dans l'environnement virtuel.

Nous avons pour l'instant des pistes de travail qui nous semblent prometteuses. Nous nous intéressons à la création et l'utilisation de divers retours sensoriels multimodaux adaptés à la formation aux gestes techniques. Le projet en est encore à ses débuts. Le développement d'une première application aura lieu durant les prochains mois et les premières expérimentations avec des experts auront lieu au cours de l'année prochaine. Nous espérons pouvoir en publier les résultats au cours de la même année.

ACKNOWLEDGEMENTS

Le projet KIVA est cofinancé par l'Union européenne. L'Europe s'engage en Picardie avec le Fond européen de développement régional (FEDER)

RÉFÉRENCES

- [1] S. Aubry. *Annotations et gestion des connaissances en environnement virtuel collaboratif*. PhD thesis, Université de Technologie de Compiègne, 2007.
- [2] R. Aylett and M. Luck. Applying artificial intelligence to virtual reality : Intelligent virtual environments. *Applied Artificial Intelligence*, 14(1) :3–32, 2000.
- [3] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana. Continuous realtime gesture following and recognition. In *Gesture in embodied communication and human-computer interaction*, pages 73–84. Springer, 2010.
- [4] K. K. Biswas and S. K. Basu. Gesture recognition using microsoft kinect®. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, pages 100–103. IEEE, 2011.
- [5] D. A. Bowman, C. North, J. Chen, N. F. Polys, P. S. Pyla, and U. Yilmaz. Information-rich virtual environments : theory, tools, and re-

- search agenda. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 81–90. ACM, 2003.
- [6] B. Bril and V. Roux. Le geste technique. réflexions méthodologiques et anthropologiques. 2002.
- [7] C. Cadoz. Le geste canal de communication homme/machine : la communication "instrumentale". *Technique et science informatiques*, 13(1) :31–61, 1994.
- [8] K. Carpentier. *Scénarisation personnalisée dynamique dans les environnements virtuels pour la formation*. PhD thesis, Compiègne, 2015.
- [9] K. Carpentier and D. Lourdeaux. Generation of learning situations according to the learner's profile within a virtual environment. In *Agents and Artificial Intelligence*, pages 245–260. Springer, 2014.
- [10] K. Carpentier, D. Lourdeaux, and I. M. Thouvenin. Dynamic selection of learning situations in virtual environment. In *ICAART (2)*, pages 101–110, 2013.
- [11] J. Cole, S. Gallagher, and D. McNeill. Gesture following deafferentation : A phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences*, 1(1) :49–67, 2002.
- [12] E. Coupeté, S. Manitsaris, and F. Moutarde. Real-time recognition of human gestures for collaborative robots on assembly-line. In *3rd International Digital Human Modeling Symposium (DHM2014)*, pages 7–p, 2014.
- [13] P. De Loor, K. Manac'h, A. Fronville, and J. Tisseau. Requirement for an enactive machine : Ontogenesis, interaction and human in the loop. In *International Conference on Enactive Interfaces*, 2008.
- [14] S. Donikian. Modélisation, contrôle et animation d'agents virtuels autonomes évoluant dans des environnements informés et structurés. *Hdr, IRISA*, 2004.
- [15] L. Fricoteaux. *Interaction informée par un système décisionnel pour un retour multimodal en environnement virtuel : application à la formation au transport fluvial*. PhD thesis, Compiègne, 2012.
- [16] P. Fuchs. *Les interfaces de la réalité virtuelle*. éditeur AJIIMD, 1996.
- [17] D. Gourlay, K. C. Lun, Y. Lee, and J. Tay. Virtual reality for relearning daily living skills. *International journal of medical informatics*, 60(3) :255–261, 2000.
- [18] A. Kendon. How gestures can become like words. *Cross-cultural perspectives in nonverbal communication*, 1 :131–141, 1988.
- [19] A. Kendon. *Gesture : Visible action as utterance*. Cambridge University Press, 2004.
- [20] S. Le Bellu, S. Lahlou, and V. Nosulenko. Capter et transférer le savoir incorporé dans un geste professionnel. *Social science information*, 49(3) :371–413, 2010.
- [21] S. Manitsaris, A. Glushkova, F. Bevilacqua, and F. Moutarde. Capture, modeling, and recognition of expert technical gestures in wheel-throwing art of pottery. *Journal on Computing and Cultural Heritage (JOCCH)*, 7(2) :10, 2014.
- [22] D. McNeill. *Hand and mind : What gestures reveal about thought*. University of Chicago Press, 1992.
- [23] D. Mellet-d'Huart and G. Michel. Réalité virtuelle et apprentissage. *Les environnements informatiques pour l'apprentissage humain. Paris : Hermes*, 2006.
- [24] R. Moreau, V. Ochoa, M. T. Pham, P. Boulanger, T. Redarce, and O. Dupuis. A method to evaluate skill transfer and acquisition of obstetric gestures based on the curvatures analysis of the position and the orientation. *Journal of biomedical informatics*, 41(6) :991–1000, 2008.
- [25] L.-P. Morency, A. Quattoni, and T. Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [26] J. Nagi, F. Ducatelle, G. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L. M. Gambardella, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on*, pages 342–347. IEEE, 2011.
- [27] J. Olive. *Annotations gestuelles pour capitaliser les interactions en environnement virtuel informé : Application à la maintenance industrielle*. PhD thesis, Université de Technologie de Compiègne, 2010.
- [28] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2) :257–286, 1989.
- [29] Z. Ren, J. Yuan, J. Meng, and Z. Zhang. Robust part-based hand gesture recognition using kinect sensor. *Multimedia, IEEE Transactions on*, 15(5) :1110–1120, 2013.
- [30] J.-C. Serre. La danse parmi les autres formes de la motricité. *La recherche en danse*, 3 :135–156, 1984.
- [31] D. Steib, L. Dalto, and D. Mellet-d'Huart. Apprendre le geste du soudage avec cs-wave : l'expérimentation de l'afpa. In *First International VR-Learning Seminar*, pages 20–21, 2005.
- [32] G. Ten Holt, M. Reinders, and E. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. 2007.
- [33] I. M. Thouvenin. *Interaction et connaissance : construction d'une expérience dans le monde virtuel*. PhD thesis, Université de Technologie de Compiègne, 2009.
- [34] J.-L. Tomás. La transmission des gestes de métier en chirurgie cardiaque. *Centre de Recherche sur le Travail et le Développement (EA 4132), CNAM, équipe clinique de l'activité. Site de l'Institut de recherche Robert-Sauvé en santé et sécurité du travail (en ligne)*. <http://www.irsst.qc.ca>, 2008.
- [35] F. G. Varela, H. R. Maturana, and R. Uribe. Autopoiesis : the organization of living systems, its characterization and a model. *Biosystems*, 5(4) :187–196, 1974.
- [36] F. J. Varela, P. Dumouchel, and P. Bourguine. *Autonomie et connaissance*. Seuil, 1989.
- [37] F. J. Varela, E. Thompson, and E. Rosch. *The embodied mind : Cognitive science and human experience*. MIT press, 1992.
- [38] J. Varona, A. Jaume-i Capó, J. González, and F. J. Perales. Toward natural interaction through visual recognition of body gestures in real-time. *Interacting with computers*, 21(1-2) :3–10, 2009.
- [39] S. B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1521–1527. IEEE, 2006.