InGredible : Une plateforme dédiée à l'interaction corporelle entre humain(s) et/ou agent(s) virtuel(s) basée sur la notion de couplage

Elisabetta Bevacqua, Romain Richard, Julien Soler, Pierre de Loor

▶ **To cite this version:**

Elisabetta Bevacqua, Romain Richard, Julien Soler, Pierre de Loor. InGredible : Une plateforme dédiée à l'interaction corporelle entre humain(s) et/ou agent(s) virtuel(s) basée sur la notion de couplage. Workshop Affect, Compagnon Artificiel Interaction (WACAI 2016), Jun 2016, Brest, France. hal-01342321

HAL Id: hal-01342321
https://hal.science/hal-01342321

Submitted on 8 Jul 2016

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.

InGredible : Une plateforme dédiée à l'interaction corporelle entre humain(s) et/ou agent(s) virtuel(s) basée sur la notion de couplage

Elisabetta Bevacqua
Lab-STICC, ENIB, UBL
Centre Européen de Réalité Virtuelle
bevacqua@enib.fr

Romain Richard
Lab-STICC, ENIB, UBL
Centre Européen de Réalité Virtuelle
richard@enib.fr

Julien Soler
Virtualys
julien.soler@virtualys.com

Pierre De Loor
Lab-STICC, ENIB, UBL
Centre Européen de Réalité Virtuelle
deloor@enib.fr

RÉSUMÉ

Cet article dresse une synthèse de travaux menés dans le cadre d'un projet sur l'interaction corporelle entre un humain et un agent virtuel ou entre deux agents virtuels. Nous présentons principalement la plateforme logicielle support qui a été développée. Elle est composée de modules interconnectés capables de réaliser la capture des mouvements, leur analyse, une prise de décision interactive basée sur la définition d'un geste ou d'une expressivité à transmettre, ou encore une synthèse et un rendu comportemental. Au delà de cette synthèse, nous insistons sur l'originalité du travail qui repose sur la notion de couplage sensorimoteur entre les interactants et qui donne lieu à des comportements adaptatifs et évolutifs, que nous relions au sentiment de co-présence. Elle peut être exploitée dans des domaines aussi différents que le sport ou les arts interactifs.

Mots clés

agent virtuel autonome, couplage humain-agent, temps interactif, co-presence, architecture cognitive, cognition incarnée, analyse gestuelle, plateforme

1. INTRODUCTION

Un problème délicat lié au développement d'agents virtuels interactifs, est de faire en sorte qu'ils permettent de susciter une relation proche de celle que l'on établie habituellement entre humains. Le réalisme, la crédibilité et le sentiment de co-présence contribuent, à leur manière, à cette illusion d'avoir affaire à un *semblable*. La co-présence est sans doute la plus complexe des propriétés à aborder car elle fait référence au sentiment subjectif d'être en présence d'un autre (*being with*) [1, 5].

Le projet `InGredible` [1], aborde ce problème sous l'angle des aspects sensorimoteurs et sur son rôle en termes de sentiment de co-présence. En effet, le courant de la cognition incarnée qui s'est beaucoup développé ces dernières années, montre que le corps joue un rôle important dans la construction d'un sens propre à chacun [19, 24]. Etant donné que notre corps est notre seul moyen de nous approprier le monde qui nous entoure, nous l'utilisons depuis le début de notre vie, pour découvrir des *régularités*. Ces régularités sont des liens entre nos perceptions et nos actions. Elles constituent la base de la création de nos représentations. Certains auteurs comme [11, 14] emettent l'idée qu'il est possible détendre ce principe aux relations sociales. Les neurosciences confirment également le lien existant entre capacités corporelles et cognition sociale [17]. Tous ces éléments sont aussi en lien avec les recherches en psychologie qui montrent que durant les interactions entre humains, chacun s'adapte continuellement tout en influençant l'autre de façon réciproque [25, 4, 8, 6]. En fait, nos premières interactions avec les autres (souvent nos parents) se font sur la base de phénomènes sensorimoteurs et d'influences mutuelles. Notre hypothèse est que ces phénomènes sont ancrés dans notre mémoire comme une marque de la présence des autres et qu'il est donc très important de les reproduire durant une interaction avec un agent artificiel pour favoriser le sentiment que cet agent est présent, avec nous. Pour développer cette idée, nous proposons une plateforme dédiée à l'interaction du corps entier entre un agent et un humain, s'appuyant sur certains principes issus du courant incarné de la cognition et de l'interactionisme.

Un des objectifs du projet `InGredible` est également de mettre la plateforme à disposition d'une compagnie de thêatre [2] avec laquelle nous collaborons, afin que les artistes puissent l'utiliser dans le cadre de performances de réalité mixte intégrant des interactions entre acteurs et agents virtuels. Le paragraphe 2 porte sur le contexte scientifique de ce travail et sur sa différentiation d'avec les approches habituelles sur les comportements interactifs d'agents virtuels. Le paragraphe 3 est une description de la plateforme et de ses différents modules, tandis qu'un exemple illustratif de son usage dans un contexte sportif est présenté par le paragraphe 4. En effet, son usage dans un contexte artistique fait partie de nos perspectives qui montrera la généricité de

1. Interaction gestuelle crédible, www.ingredible.fr

2. www.derezo.com

la proposition.

2. CONTEXTE

Dans la litterature, différentes approches peuvent être rapprochées de la notre. Par exemple, l'équipe de Stephan Kopp travaille sur la notion de *résonance* [16]. Cette notion pouvant s'appliquer au control moteur ou aux relations sociales. Dans ce dernier cas, plusieurs phénomènes contribuent à maintenir une connexion entre les personnes. Ils reposent sur une coordination corporelle et une adaptation mutuelle qui génèrent des motifs d'interaction que l'on retrouve au travers d'imitation [20], de mimétisme [23] ou encore de synchronisation dans les comportements [2]. Pour favoriser l'apparition de ces phénomènes, les auteurs utilisent des agents dotés d'un modèle de leurs actions et gestes dont ils se servent pour *interpréter* les actions et gestes observés d'un humain. [21]. Cette interprétation du comportement de l'humain à partir de ce que sait faire l'agent, favoriserait la résonnance entre eux. Une autre inspiration de notre travail se trouve chez Gratch et al. [13] qui proposent que les agents virtuels puissent être des outils pour les sciences cognitives. Par exemple, leur atelier contient un générateur de comportements non verbeaux basé sur des résultats de travaux d'observations provenant de la psychologie sociale [18]. Le système à base de règles résultant de ce travail permet la génération de comportements en lien avec les phrases qui doivent être produites par l'agent ainsi que son état émotionnel supposé.

Notre travail est moins général dans le sens où il n'aborde pour l'instant que les interactions gestuelles. Nous sommes bien conscient du fait qu'à longs termes, il faudra y intégrer des aspects de plus haut niveau. Cependant, nous préférons concentrer nos efforts sur le langage du corps avant d'aborder le reste du problème, car conformément au champ de la cognition incarnée, c'est lui qui est à la base du reste. Par ailleurs, notre approche présente l'originalité de faire des propositions intégrant d'avantage le rôle du corps entier dans le ressentie de la co-présence d'agents virtuels. L'aspect temps-réel y est très important et le modèle de décision est basé sur une notion appelée *couplage* qui met en avant cette relation temporelle entre le comportement des deux participants (l'agent et l'humain). Ce couplage fait partie intégrante de la dynamique sensorimotrice mise en œuvre, qui est elle inspirée des travaux en psychologie proposés par [8]. Ceux-ci comparent la communication entre personnes à un phénomène *vivant* qui constitue la métaphore que nous utilisons pour construire notre modèle comportemental. Un des aspects fondamentaux du vivant est sa capacité à se réguler á se maintenir et à résister aux perturbations de son environnement (système immunitaire, systèmes nerveux, systèmes sociaux). Métaphoriquement, le couplage est un phénomène qui possède ce type de propriétés. Nous le définissons comme une influence mutuelle et continue entre deux personnes, influence qui se maintient, se régule et evolue. La mesure de cette influence est possible en observant des corrélations entre les comportements mutuels des interactants qui sont considérés comme des signaux [3]. En effet, quelques soient les modalités utilisées pour interagir, nous les assimilons á des signaux qui vont faire l'objet d'une comparaison et d'une régulation du résultat de cette comparaison. L'essence du couplage est de n'exister que parce qu'il peut disparaître. Il est à l'interaction, ce que l'autpoièse est au vivant. Il y a une *lutte* pour son maintien et les personnes qui interagissent vont faire un effort dans ce sens, lorsqu'elles sentiront son affaiblissement. Ainsi, si un signal est interprété comme étant trop faible par l'interactant qui le reçoit, celui-ci va envoyer un signal en retour, suscitant une reaffirmation du signal prédédent (par exemple, froncement de sourcils, onomatopée de type 'hein ?'). A l'inverse, si le signal est trop fort (par exemple, un ton de voix élevé), l'interactant qui le reçoit va réagir différemment (par exemple, geste de la main signifiant qu'il faut baisser le ton ou onomatopée de type 'ok, ok'). Les deux interactants réagissant de façon analogue, avec leurs propres personnalités. Il apparait alors une sorte d'équilibre de l'interaction, propre è la dyade, toujours en mesure d'être rompu et donc sans cesse évalué et corrigé. La difficulté pour un programme est de reproduire la subtilité des variations réciproques entre les interactants ainsi que les mécanismes de régulation associés, surtout lorsque l'un des interactants est un humain. Or, nous faisons l'hypothèse que c'est cette attention mutuelle, cette capacité á reconnaitre ce phénomène de régulation et cet effort de recherche de l'équilibre, qui peut être la source du sentiment de co-présence et peut-être également de l'engagement et de la crédibilité des agents virtuels [3]. Un deuxième point important est que de ce phénomène résulte une évolution temporelle des réactions des interactants : si à un instant donné, l'emission d'un geste A provoque en retour un geste B et bien, à un autre instant, ce même geste A ne provoquera pas le même retour. En effet, la mesure du couplage et de sa variation temporelle, à une échelle de temps différente du *backchannel* ou de la synchronie, va venir moduler les réactions mutuelles des protagonistes. Notre modèle de décision, s'appuyant sur des variables continues, prend en compte ce principe de régulation et permet de faire apparaître une interaction que l'on pourrait qualifier d'évolutive. In fine, pour Fogel et al. [8], la communication est développementale dans le sens où il va y avoir une introduction progressive de nouveaux types de signaux qui vont progressivement être assimilés au repertoire des signaux de régulation. Notre approche ne va que trés modestement vers cet aspect développemental. Elle se focalise surtout sur l'aspect évolutif même si la frontière entre évolution et développement est ténue. L'évolution dans toute sa subtilité temporelle est déja un problème complexe qu'il serait très difficile de résoudre avec un système purement discret et que nous abordons et que l'architecture du projet `Ingredible` permet d'aborder.

Une conséquence de notre proposition basée sur le couplage est que l'agent virtuel va être capable de réagir à n'importe quel comportement, y compris ceux qui n'ont pas été envisagés a priori. En effet, même si le comportement de l'humain n'a aucun sens au regard d'un scénario ou d'un contexte d'interaction, il sera possible d'évaluer un couplage sensorimoteur entre ce comportement et celui de l'agent et donc de proposer une réaction qui aura pour objectif d'améliorer le couplage. Ainsi, l'interaction n'est jamais interrompue par un arrêt subit des mouvements de l'agent. La figure 1 illustre les différences de principe entre des approches basées règles et notre approche.

3. LA PLATEFORME INGREDIBLE

La plateforme est basée sur une architecture modulaire,

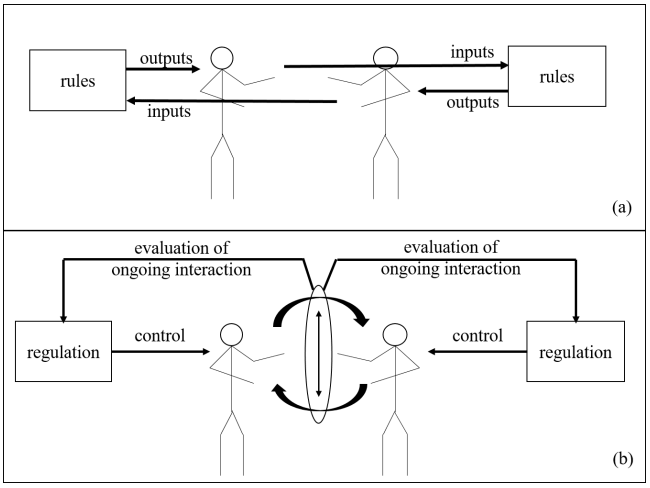

Figure 1: a) approches entrées/sorties classiques, basées sur des règles : le comportement de l'agent est décrit à partir des actions de l'humain qui sont considérées comme des entrées. b) Notre approches : le comportement de l'agent est une régulation du couplage qu'il opère avec l'humain, c'est en quelque sorte le couplage qui est l'entrée.

certains modules étant prédéfinis pour la *capture*, l'*analyse*, la *decision*, la *synthèse* et le *rendu*. Ces modules sont bien évidemment des propositions qui peuvent être enrichies et modifiées. Les modules communiquent grâce à un *framework* de communication.

3.1 Framework de communication

Ce *framwork* s'appuie sur le protocole de sérialisation de données structurées `protobuf` [3]. Il est indépendant de la plateforme d'acceuil et optimisé pour la communication temps-réel de messages. Nous avons besoin d'une telle optimisation car certains modules sont producteurs de nombreuses informations (la capture de mouvements et l'analyse de mouvements par exemple) et la réaction doit être produite en temps interactif.

Notre plateforme possède une API pour `C++`, `JAVA` and `PYTHON` [4]. Lors de la création d'un module, il faut simplement définir les modules avec lesquels il va échanger (un fichier de configuration contient toutes ces informations ainsi que les adresses IP des modules) et renseigner les fonctions de *callback* associées aux messages correspondants.

Plusieurs messages sont prédéfinis ainsi que les traitements qui leur sont généralement associés. Il s'agit de `Skeleton`, `Bone`, `Rotation` et `Position` qui sont relatifs au corps (de l'humain comme de l'agent) ; `SkeletonDetected` et `SkeletonLost` donnent des information sur l'ètat des systèmes de capture. `GestureRecognition` et `Features` contiennent des informations permettant le calcul du couplage. `Decision` et `GestureControl` sont destinés aux modules de synthèse de mouvements. Enfin, les messages de type `Generic` sont utilisés pour tout autre type de message qui semblerait utile aux utilisateurs. Ils doivent alors créer eux-même

3. https://developers.google.com/protocol-buffers/
4. pour le moment, l'API est disponible sur le site web du projet, le lecteur interressé peut nous contacter pour obtenir les sources qui seront rapidement mises en ligne

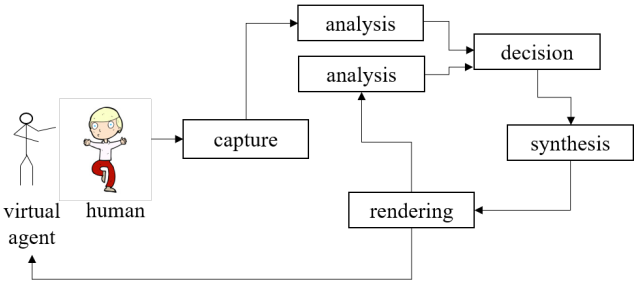

Figure 2: Configuration du framework pour une interaction entre un agent virtuel et un humain.

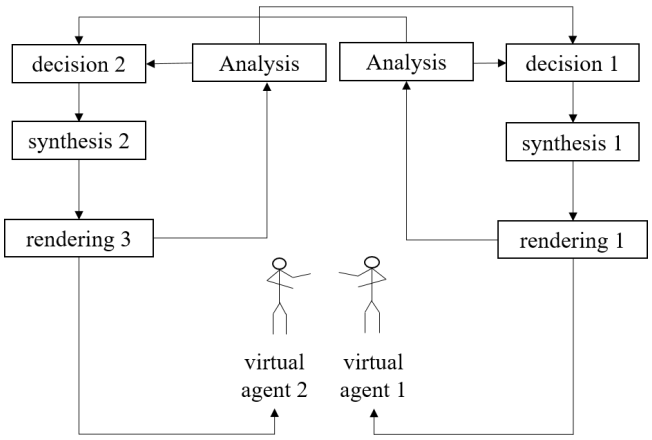

Figure 3: Configuration de la plateforme pour une interaction entre deux agents virtuels.

le code de leur utilisation. En fait, nous utilisons ce dernier type lorsque nous hésitons entre plusieurs propositions pour résoudre tel ou tel problème. Dès qu'une proposition est pérenne, un message qui lui est dédié est créé.

Cette approche modulaire permet de créer des configurations différentes très rapidement, s'appuyant sur des outils de capture ou de rendu différents mais surtout, permettant des expérimentations différentes.

Les figures 2 et 3 illustrent deux configurations. La première pour une interaction humain-agent, l'autre pour une interaction agent-agent. Seul le module de capture diffère pour ces deux configurations, tous les autres modules ne *savent pas* si les données transmisent viennent de la capture des mouvements d'un humain ou d'un agent virtuel. Il est alors très facile de développer un agent capable d'interagir avec un humain, puis de remplacer cet humain par un autre agent pour voir émerger une interaction entre deux agents !

Une des originalités du module de décision proposé par la plateforme est qu'il prend en entrée, non pas des informations provenant de l'analyse des mouvements d'un seul des interactants, mais des deux interactants, c'est à dire, celui qui interagit avec l'agent controlé par le module mais également l'analyse du comportement de l'agent controlé, lui même. C'est la mise en pratique de la prise en compte du couplage comme élément central de la prise de décision.

3.2 Module de capture

Le module de capture permet de connecter la plateforme à de nombreux dispositifs techniques de capture existants

ou à venir. L'important est qu'en sortie du module, on n'ait plus besoin de savoir comment est capté l'humain ou l'agent qui est en entrée ni de savoir si c'est un humain ou un agent qui est capté. Le module renvoie alors un message `skeleton` normalisé de 15 articulations. Ce squelette, assez minimaliste, présente l'avantage d'être compatible avec des systèmes de capture grand public, et même s'il peut être vu comme une dégradation de la qualité d'un squelette obtenu avec des dispositifs professionnels, il est clairement suffisant pour réaliser la tâche du module suivant qui consiste à analyser et reconnaitre les mouvements. Actuellement, cinq dispositifs de captures sont intégrés : `Microsoft Kinect` [5], `Moven suit` [6], `OptiTrack` [7], `ARTTrack` [8] et bien sûr notre module de rendu qui permet donc de capturer un agent virtuel comme si c'était un humain. Dans un tel cas, le module n'a pour rôle que de transmettre le squelette qu'il reçoit du rendu. Si le rendu travaille avec un squellette normalisé, on peut ommettre le module de capture (c'est le cas pour la figure 3).

3.3 Le module d'analyse

Ce module analyse en temps-réel le squelette envoyé par le module de capture (ou le module de rendu). Il emet deux types d'informations : 1) un ensemble de caractéristiques des mouvements, calculées à partir des successions temporelles des postures du squelette. 2) Le nom du geste qu'il reconnait, calculé à partir des valeurs des différentes caractéristiques. Les caractéristiques des mouvements portent à la fois sur des parties du corps (mains, pieds, bras, jambes ...) et sur le corps entier. Nous calculons la fluidié, la densité ou l'ernergie d'un geste, d'après les équations proposées par [12] et [22]. A cela s'ajoute des grandeurs de vitesses et d'accélélrations. Le tout, associé à tout ou partie du corps, donne lieu, au total, à 233 caractéristiques potentielles. L'algorithme de reconnaissance de geste s'appuie sur ces caractéristiques ainsi que sur une analyse en composantes principales qui a été préalablement faite à partir d'un échantillon des gestes que l'on aura besoin de reconnaitre. Les détails et performances de l'algorithme sont fournis dans [15].

En résumé, il opère une reconnaissance à partir du flot continu des données d'entrée. Ceci veut dire que le geste est généralement reconnu bien avant sa fin. C'est une proporiété très interessante dans le cadre de la régulation de comportements telle que nous la proposons. Nous avons montré que sur un ordinateur standart, l'algorithme peut reconnaître jusqu'à 32 gestes en temps-réel. Il possède un taux de reconnaissance de $96,18\%$ sur une base de gestes composée pour moitié de notre système de capture, et pour l'autre moitié de la base MSRC-1 Kinect de Microsoft [9]. L'algorithme est très robuste à la morphologie (en particulier grâce à la normalisation du squelette), mais également à l'expressivité (rapide, lent, large, dense ..), tout en nécessitant peu de données d'apprentissage (5 répétitions de chaque geste suffit à obtenir des résultats satisfaisants dans la majorité des cas). Le module d'analyse émet des messages `Features` en continu et des messages `GestureRecognition` dès qu'un geste est reconnu.

5. http://www.xbox.com/fr-FR/xbox-one/
6. https://www.xsens.com/products/xsens-mvn/
7. https://www.optitrack.com/products/motive/
8. http://www.ar-tracking.com/products/tracking-systems/arttrack-system/

3.4 Module de décision

Ce module est le représentant principal de notre approche, qui base le comportement d'un agent virtuel sur la mesure du couplage qu'il opère avec l'autre interactant. Son fonctionnement est formalisé et décrit en détail dans [7]. Il s'appuie sur une proposition théorique venant de la psychologie, faisant un rapprochement entre la communication entre deux personnes et un système *vivant* [8]. En résumé, la communication est un processus évolutif au sein duquel on peut distinguer trois modes d'interactions bien spécifiques : La *co-regulation*, la *variabilité ordinaire* et l'*innovation*. La *co-regulation* consiste en un ajustement mutuel des signaux envoyés par les deux interactants. Ces signaux sont socialement admis comme le fait que *tout se passe bien*, chacun signifie à l'autre qu'il *le suit*. Parmi ces signaux, on trouve le *backchannel* [26] ou le *mimétisme* par exemple. Ces signaux sont toutefois sujets à des variabilités intrinsèques propres à chacun, qu'il faudra reproduire dans un agent virtuel. La *variabilité ordinaire* dépasse cette variabilité intrinsèque de la co-régulation. Elle apparait après un temps de *co-régulation* qui permet aux interactants d'oser *de nouvelles choses*. Ils introduisent des changements subtils mais perceptibles, qui appellent une réaction de la part de l'autre interactant. Ces propositions vont progressivement modifier l'objet ou le but de l'interaction. L'*innovation* est une variabilité qui sort des limites socialement admises dans le contexte initial de l'interaction. C'est une proposition plus franche qui suggère de nouveaux comportements. Tout comme la *variabilité ordinaire* est utilisée pour faire évoluer la *co-régulation* aprés un certain temps, l'*innovation* permet d'aller encore plus loin et de faire évoluer la *variabilité*. En effet, une innovation va finir par être considérée comme une variabilité admise par les interactants et le processus pourra recommencer, donnant à la communication son caractère développemental. Les personnes ressentent ces modes d'interactions et leur évolution. Ils s'y adaptent et c'est cette adaptation perpétuelle et la perception de cette adaptation que nous supposons être partie prenante du sentiment de co-présence.

Le module de décision fait une proposition générique dans le sens où il implémente l'evolution temporelle entre les modes d'interactions (*co-régulation, variabilité et innovation*) à partir d'une mesure du couplage et de son évolution temporelle. Pour cela, des grandeurs réelles permettent de rendre compte d'une combinaison entre la qualité du couplage et sa dynamique. Ces variables évolutent grâce à des équations temporelles et permettent de calculer la probabilité qu'a l'interaction de passer d'un mode à un autre. Par exemple, si le couplage est assez élevé depuis un certain temps, la probabilité d'entrer en variabilité va progressivement augmenter. A l'inverse, si le couplage varie sans cesse, c'est que les interactants ne s'accordent pas. Une innovation pourrait alors être salvatrice. L'ensemble des combinaisons et conditions de passage a été modélisé. Si il y a un changement de mode d'interaction, le système estime également lequel des deux interactants est le plus vraissemblablement à l'origine de ce changement car cela a un impact sur le comportement à décider. Par exemple, si c'est l'humain qui a changé brutalement son expressivité, c'est lui l'*initiateur* du changement. Dans le cas inverse, c'est l'agent qui a provoqué un changement, vraissemblablement à cause de l'évolution des grandeurs réelles évoquées ci-dessus. La figure 4 illustre le fonctionnement de ce module. A chaque mode d'interaction, est associée une fonction de comportement

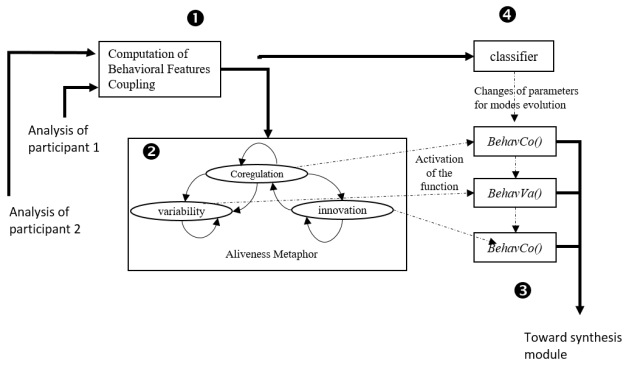

Figure 4: Partie générique du module de décision : (1) le couplage entre les deux interactants est calculé, il est utilisé pour estimer le mode d'interaction au travers de l'usage d'une variable discrète aléaltoire (2). En (3) se trouvent les 3 fonctions comportementales dédiées à chaque mode d'interaction. (4), modulation du comportement des fonctions comportemtales, par le biais de paramêtres dépendant du couplage.

(`BehavCo(), BehavVar(), BehavInov()`). Elle sera appelée en boucle durant tout le temps d'activation du mode. Ces fonctions, tout comme certaines autres informations, doivent être spécifiées par l'utilisateur car elles sont dépendantes du contexte dans lequel la plateforme est utilisée. Par exemple, la co-régulation dans un exercice de sport peut consister à adapter sa vitesse à celle de l'autre, alors que dans un contexte de théâtre, ce sera plutôt la posture qui sera adaptée. En plus des trois fonctions comportementales précédentes, le mécanisme générique central, repose sur la qualité du couplage dont la définition est elle aussi dépendante du domaine d'application. Par exemple, si l'application est du domaine de l'interaction artistique (la dance), la mesure du couplage fera vraisemblablement intervenir des caractéristiques de fluidité alors que dans le cadre d'une interaction sportive, ce seront les caractéristiques de vitesses qui seront plus pertinentes. Le couplage est en fait une mesure de corrélation des charactéristiques des gestes pondérée en fonction du contexte de l'interaction. Tout ou partie des 233 caractéristiques gestuelles calculées par les modules d'analyse peuvent être utilisées (en leur affectant un poids réel compris entre 0 et 1). Tout usage du module passe donc par le choix de ces poids. Notons que l'identification du geste par le module d'analyse peut être considérée comme une des caractéristiques importantes pour calculer le couplage entre deux personnes. Il n'est pas pertinent dans les contexte d'improvisation, pour lequel, par définition, il ne peut y avoir de gestes prédéfinis donc de gestes à reconnaître. Il l'est certainement beaucoup plus dans un contexte de travail collaboratif où à chaque geste d'un interactant peut correspondre un geste de l'autre interactant. Pour finir, un système de classeur est utilisé pour simuler une variabilité intrinsèque de l'agent. Il modifie certains paramètres des fonctions de comportement, en ligne, selon certaines valeurs des caractéristiques des gestes. Dans les faits, ce système est consitué de peu de règles relativement simples. Par exemple, une règle peut exprimer le fait que si le couplage est supérieur à 80%, la vitesse peut être augmentée aléatoirement de 0 à 5 %. Ce genre de règles suffit déjà pour entretenir une variabilité suffisante au sein de chaque mode d'interaction. A termes ce système de classeurs devrait être exploité davantage pour mieux intégrer les aspects développementaux de l'architecture.

Les messages envoyés par un module de décision est de type `Decision`. Ce type permet de spécifier un identifiant de geste, une expressivité et/ou des positions clefs que le squelette de l'agent doit atteindre (en respectant l'expressivité si elle est renseignée). L'hypothèse est faite que le module de synthèse pourra prendre en compte ces informations pour générer le geste adéquat.

3.5 Rendu et synthèse

Le cœur de nos recherche porte sur la décision (basée couplage), l'analyse de l'interaction et la reconnaissance de gestes en continu. Pour ce qui est de la synthèse et du rendu, nous nous appuyons sur le moteur de jeu Unity3D [9] associé à la librairie FinalIK [10]. Grâce à ces deux outils, synthèse et rendu forment un seul module qui peut tout aussi bien jouer un geste pré-enregistré à partir de captures de mouvements, moduler ce geste par le biais de contraintes d'expressivité ou de positions clefs à maintenir pour certaines parties du corps, ou encore mixer deux postures ou deux gestes. Tout autre outils de synthèse et de rendu capable de réaliser ce type de traitement peut bien sûr être utilisé sur la plateforme. Notons qu'actuellement, notre module de synthèse/rendu génère un squelette normalisé qui peut être directement géré par notre module d'analyse.

4. RÉSULTATS ET PERSPECTIVES

La figure 5 montre l'usage de la plateforme dans un contexte d'exergame. Il s'agit ici de simuler un coach, un élève ou un compagnon de fitness. En résumé, les caractéristiques des gestes permettant d'évaluer le couplage sont principalement liées à la vitesse (et au fait que l'humain fasse le même geste que l'agent). Selon qu'il est coach ou élève, les fonctions comportementales de l'agent diffèrent. Par exemple, en tant que coach, il pourra introduire une expressivité de compensation : si l'élève humain est trop lent, il va exagérer la vitesse du geste à produire. Ce qui nous semble important est que l'agent est toujours en mesure de proposer une réaction qui évolue et qui semble adaptée. Si le geste de l'humain n'est pas reconnu, l'agent va proposer d'autres gestes (innovation), si l'humain est très couplé avec l'agent, il va lui proposer des variations. Il est peu probable que les temps au bout desquels les changements ont lieu soient identiques car ils dépendent du couplage (qui est une variable réelle qui évolue continuellement). De même, il est peu probable qu'un geste soit joué exactement de la même façon à des moments différents. Nous sommes actuellement en train de mener des expérimentations pour valider l'hypothèse qu'une telle approche améliore le sentiment de co-présence.

Nous utilisons également cette plateforme dans le cadre d'une collaboration avec des artistes. Ici, le contexte est celui de l'improvisation. Il n'y a pas de gestes prédéfinis et la synthèse ne reçoit que des positions à atteindre par certaines parties du corps. Cette performance est inspirée par un exercice de théâtre. Deux interactants doivent s'imiter tout en

9. https://unity3d.com
10. http://www.root-motion.com/final-ik.html

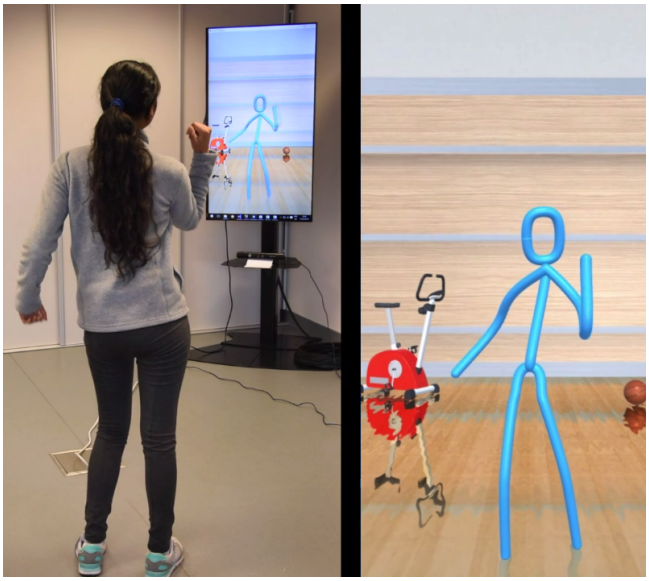

Figure 5: Image extraite d'une vidéo de l'exemple du jeu de fitness. A gauche, l'ensemble du système où ici, l'humain est capturé par une kinect. A droit, la vue de l'agent virtuel. Afin de nous concentrer sur les aspects sensorimoteur, nous avons choisi ce type de représentation, plutôt abstraite mais qui rend compte de l'expressivité tout en évitant les eccueils d'une *uncany valley*.

proposant à l'autre des variations. Il est alors impossible de prédire ce que va devenir l'interaction. Notre approche est bien adaptée à ce principe et nous espérons que les spectateurs tout comme les acteurs ressentirons cette ambiguité dans les comportements de l'agent et de l'humain, ambiguité qui empêche réellement de savoir qui est le guide et qui est le guidé. Toujours dans le cadre de ces collaborations artistiques, nous développons actuellement un module permettant d'interfacer la plateforme avec le protocole OSC [10] largement diffusé dans le monde du spectacle. L'idée est de pouvoir intégrer `InGredible` à des installations de plus grande envergures, destinées à des spectacles vivants. Enfin, puisque la plateforme le permet, nous utilisons la configuration de la figure 3 pour mettre deux agents virtuels face à face. L'idée est d'observer le comportement interactif qui émerge, sachant que les deux modules de décisions s'appuient chacun sur le couplage constaté entre les deux agents mais que pour chacun d'eux, le poids des caractéristiques consitutant la mesure du couplage est différent. Le comportement obtenu est alors évolutif et peut donner lieu à des performances artistiques tout comme à des études sur l'interaction sociale dans le cadre de la métaphore de la communication vivante.

5. REMERCIEMENTS

Nous tenons à remercier l'ANR qui a financé ce projet (ANR-12-CORD-001), Céline Jost, Aléxis Nédélec, ainsi que les partenaires qui sont associés au projet, en particulier l'équipe de Jean Claude Martin du LIMSI et l'équipe de Sylvie Gibet à l'IRISA.

6. REFERENCES

[1] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence : Teleoperators and Virtual Environments*, 14(4) :379–393, 2005.

[2] C. Bernieri and R. Rosenthal. Interpersonal coordination : Behavior matching and interactional synchrony. *Fundamentals of nonverbal behavior*, 1991.

[3] E. Bevacqua, I. Stanković, A. Maatallaoui, A. Nédélec, and P. De Loor. Effects of coupling in human-virtual agent body interaction. In B. et al., editor, *14th International Conference of Intelligent Virtual Agents*, 8637, pages 54–63, 2014.

[4] J. K. Burgoon, L. A. Stern, and L. Dillman. *Interpersonal Adaptation : Dyadic Interaction Patterns*. Cambridge University Press, 1995.

[5] J. Casanueva and E. H. Blake. Presence and co-presence in collaborative virtual environments. *Proceedings of the 2nd South African Telecommunications, Networks and Applications Conference*, pages 120–125, 1999.

[6] P. De Loor, E. Bevacqua, I. Stanković, A. Maatallaoui, A. Nédélec, and C. Buche. Utilisation de la notion de couplage pour la modélisation d'agents virtuels interactifs socialement présents. In *Conférence III*. Oxford University Press, 2014.

[7] P. De Loor, R. Richard, J. Soler, and E. Bevacqua. Aliveness metaphor for an evolutive gesture interaction based on coupling between a human and a virtual agent. In *Proceedings of the 29th conference on computer animation and social agents, CASA*, pages 147–155. ACM Press, 2016.

[8] A. Fogel and A. Garvey. Alive communication. *Infant behavior & development*, 30(2) :251–7, may 2007.

[9] S. Fothergill, H. M. Mentis, S. Nowozin, and P. Kohli. Instructing people for training gestural interactive systems. In *ACM Conference on Computer-Human Interaction*, 2012.

[10] A. Freed, A., Schmeder. Features and Future of Open Sound Control version 1.1 for NIME. In *NIME Conference*, 2009.

[11] T. Froese, H. Iizuka, and T. Ikegami. Embodied social interaction constitutes social cognition in pairs of humans : A minimalist virtual reality experiment. *Scientific Reports 4*, (3672) :1–10, 2014.

[12] D. Glowinski, M. Mortillaro, N. Dael, and A. Camurri. Towards a Minimal Representation of Affective Gestures (Extended Abstract). In *International Conference on Affective Computing and Intelligent Interaction*, pages 498–504, 2015.

[13] J. Gratch, A. Hartholt, M. Dehghani, and S. Marsella. Virtual Humans : A New Toolkit for Cognitive Science Research Presenters. *McNeill 2005*, (2010) :41–42, 2012.

[14] H. D. Jaegher and E. D. Paolo. Participatory Sense-Making An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Science*, 6 :485–507, 2007.

[15] C. Jost, P. De Loor, E. Nédélec, Alexis I. Stanković. Real-Time Gesture Recognition Based On Motion Quality Analysis. In *7th International Conference on Intelligent Technologies for Interactive Entertainment - INTETAIN*, 2015.

[16] S. Kopp. Social Resonance and Embodied Coordination in Face-to-Face Conversation with Artificial Interlocutors. *Speech Communication*, 52(6) :587–597, 2010.

[17] J. Krueger and J. Michael. Gestural coupling and social cognition : Möbius Syndrome as a case study. *Frontiers in human neuroscience*, 6(81) :1–14, 2012.

[18] J. Lee, S. Marsella, and M. D. Rey. Nonverbal Behavior Generator for Embodied Conversational Agents. In *6th International Conference of Intelligent Virtual Agents*, volume 4133 of *LNCS*, pages 243–255, 2006.

[19] J. R. Matyja. Travelling in Style from Standard Cognitive Science to Embodied Cognition. *ConstruCtivist*, pages 231–233, 2011.

[20] A. Meltzoff and M. Moore. Imitation et développement humain : les premiers temps de la vie. *Terrain*, 44 :71–90, 2005.

[21] A. Sadeghipour and S. Kopp. A probabilistic model of motor resonance for embodied gesture perception. In *9th International Conference of Intelligent Virtual Agents*, volume 5773 of *LNCS*, pages 90–103, 2009.

[22] A. A. Samadani, S. Burton, R. Gorbet, and D. Kulic. Laban effort and shape analysis of affective hand and arm movements. In *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 343–348, 2013.

[23] R. B. Van baaren. *Mimicry : a social perspective*. PhD thesis, University of Nijmegen, 2003.

[24] F. J. Varela, E. Thompson, and E. Rosch. *The Embodied Mind*. MIT Press, 1993.

[25] W. Warren. The dynamics of perception and action. *Psychological Review*, 113 :358–389, 2006.

[26] V. Yngve. On getting a word in edgewise. In *Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–577, 1970.