



Interaction corporelle évolutive entre un humain et un personnage virtuel

Pierre de Loor, Romain Richard, Elisabetta Bevacqua

► To cite this version:

Pierre de Loor, Romain Richard, Elisabetta Bevacqua. Interaction corporelle évolutive entre un humain et un personnage virtuel: Proposition d'un modèle théorique et de son évaluation dans le cadre d'un exergame de fitness. *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle*, 2017, 10.3166/RIA.31.1-23 . hal-01590022

HAL Id: hal-01590022

<https://hal.science/hal-01590022>

Submitted on 23 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Résumé

Cet article présente un modèle permettant de mettre en œuvre une interaction adaptative et évolutive entre un humain et un personnage autonome. Les bases théoriques de ce modèle sont inspirées d'une analogie entre communication et système vivant. Elles sont présentées de façon synthétique. La partie suivante présente une expérimentation qui a permis d'évaluer ce modèle dans le cadre d'un exergame de fitness. L'évaluation montre entre autres que le modèle génère bien un comportement adaptatif en temps réel, comparable à celui ayant lieu entre les humains. Plus précisément, le comportement est associé à un fort sentiment de co-présence dans un contexte d'interaction nécessitant une certaine attention de la part de l'humain sur le comportement du personnage virtuel. Les résultats sont ensuite discutés car le sentiment de co-présence et l'engagement de l'humain semblent bien améliorés, mais ce n'est pas le cas de la crédibilité du comportement.

Interaction corporelle évolutive entre un humain et un personnage virtuel

Pierre De Loor, Romain Richard et Elisabetta Bevacqua

23 novembre 2017

1 Introduction

Cet article aborde certains aspects complexes de la mise en place d'interactions corporelles entre un humain et un personnage virtuel. L'hypothèse sous-jacente est que certains aspects temporels, ont une influence importante sur des propriétés que l'on cherche généralement à améliorer chez les personnages virtuels : leur crédibilité et leur présence. Par voie de conséquence, cette dynamique doit avoir un impact sur l'expérience de jeu et l'engagement de l'humain qui interagit avec le personnage. Souvent, co-présence et crédibilité sont évaluées selon un point de vue peu dynamique. Par exemple, on mesure l'influence de la taille du personnage, de ses expressions faciales ou encore de son statut social par rapport à l'utilisateur. Cependant, lors d'une interaction, il existe des phénomènes subtils, plus difficiles à qualifier mais qui sont ancrés corporellement en nous et donnent lieu à une dynamique réactive et évolutive de l'interaction. Ainsi, deux sessions d'interaction ne vont jamais produire les mêmes séquences de gestes ou de langage corporel. Mieux, un même geste utilisé par une même personne pour communiquer avec une autre, ne va pas toujours déclencher la même réaction selon le moment où il sera émis durant l'interaction. L'histoire de l'interaction entre donc en jeu durant celle-ci et provoque son évolution. En particulier, la réaction d'un des participants à l'interaction à un instant donné va être dépendante du ressenti de l'interaction généré lui-même par ce qui s'est passé avant cet instant. La difficulté est que ce ressenti est dépendant des deux interactants et qu'il n'est pas facile d'établir un modèle de décision intégrant cette inter-dépendance.

Dans [8], nous avons proposé un tel modèle ayant pour objectif de reproduire une partie de ce phénomène. Il est basé sur une évaluation du couplage sensorimoteur existant entre le personnage virtuel et l'humain et propose d'améliorer la régulation de ce couplage, régulation qui améliorera la qualité du couplage. Ce niveau de couplage étant l'entrée principale de notre modèle, là où les modèles habituels prennent en

*Lab-STICC, ENIB, UBL, Centre Européen de Réalité Virtuelle Brest, FRANCE.
[deloor,bevacqua]@enib.fr

entrée le comportement de l'humain seul. Cet article est dédié à l'évaluation de ce modèle. Il commence par un état de l'art sur l'évaluation de la crédibilité et de la co-présence d'agents virtuels ainsi que sur les modèles de décisions existants et pouvant être rapprochés du notre. La section 3 résume les principes de ce modèle ainsi que son application dans un exergame de fitness. La section 4 présente l'expérimentation que nous avons menée pour mener à bien l'évaluation. En particulier, nous y faisons des comparaisons du ressenti par les sujets selon que l'agent avec lequel ils interagissent est contrôlé par notre modèle, par un modèle n'utilisant pas la notion de couplage et également par un magicien d'Oz.

2 État de l'art

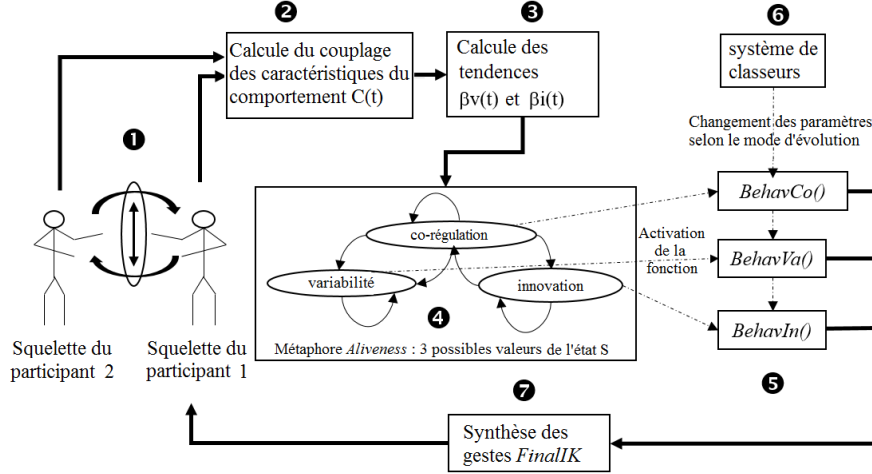
Un des objectifs liés à la réalisation de personnages virtuels est de faire en sorte qu'un observateur ait l'illusion qu'il se comporte comme une personne réelle. L'hypothèse sous-jacente étant que plus le personnage virtuel sera similaire à une personne réelle, plus l'observateur aura tendance à adopter un point de vue et un comportement similaire à celui qu'il adopterait devant une personne réelle. Il existe de nombreuses propriétés caractérisant cette similarité. Celles qui nous intéressent ici particulièrement sont des propriétés perceptives de *crédibilité* et de *co-présence*. La crédibilité est relative à l'adéquation du personnage virtuel à ce qu'en attendrait l'observateur s'il s'agissait d'une personne réelle [10], tandis que la co-présence fait référence au sentiment d'être *avec* une personne [14, 2]. Même si la nuance est subtile, elle permet d'imaginer qu'un personnage crédible ne suscite pas nécessairement une co-présence ou qu'un personnage co-présent ne soit pas nécessairement crédible. Le premier problème lié à ces propriétés est qu'elles font référence à un sentiment subjectif. L'autre problème est de déterminer les propriétés nécessaires et suffisantes que doit posséder le personnage pour que cela améliore la crédibilité et/ou le sentiment de co-présence. Comme mentionné en introduction, des facteurs tels que le rendu graphique sont évalués et ne donnent pas nécessairement un résultat probant [22]. En effet, une représentation quasi parfaite d'un personnage virtuel va rapidement être sujette au problème de la vallée de l'étrange si son comportement n'est pas lui aussi quasi-parfait [17]. Par ailleurs, [13] a montré que le comportement de l'agent et surtout sa congruence avec son apparence vont améliorer la perception qu'en aura un observateur. Wong et collègues montrent aussi que l'expressivité et la capacité à rendre compte d'émotions vont avoir également un impact [23]. [6] ont montré que le niveau d'expertise des sujets est un facteur impactant beaucoup les résultats.

L'évaluation de la co-présence reste un débat ouvert concernant l'utilisation d'évaluations subjectives, par le biais de questionnaires par exemple, ou objectives, par le biais de mesures physiologiques ou comportementales. Il s'agit en fait d'un problème qui va au delà de la co-présence et qui est récurrent en psychologie ou en sciences cognitives [19, 20]. La posture raisonnable consiste à accumuler et à comparer les

deux familles d'évaluation afin d'identifier les potentiels liens entre elles [1]. La difficulté sous-jacente est qu'il existe également de nombreuses façons de mettre en œuvre chacune de ces familles d'évaluation et que par exemple, le choix des variables objectives à observer n'est pas anodin et entraîne une certaine subjectivité. Le rôle de l'interaction est rarement étudié en terme de co-présence (généralement les chercheurs évaluent la co-présence des personnages virtuels sans que les sujets puissent interagir avec eux). Or, nous postulons que l'interaction peut jouer un rôle important dans le ressenti de la présence de celui avec qui on interagit. Dans des travaux précédents [5], nous avons montré que lorsque les sujets faisaient un effort pour maintenir le couplage sensorimoteur avec un agent virtuel, cela améliorait leur perception de sa présence. Cependant, le personnage était contrôlé par un magicien d'Oz. Depuis, nous avons implémenté le modèle de décision temps réel et le présent travail consiste à l'évaluer et en particulier à le comparer à un magicien d'Oz. Ce modèle a pour objectif de reproduire la dynamique des interactions, selon des considérations psychologiques. Il existe des propositions ayant des objectifs similaires dans la littérature. En particulier, les phénomènes tels que la 'mimicry', la synchronie ou encore les 'backchannels', tous considérés comme caractéristiques d'une interaction entre deux personnes, sont l'objet de certains modèles. Par exemple, dans [3], un agent virtuel 'attentif' utilise des signaux audiovisuels émis par le sujet humain pour déclencher des 'backchannels' (des signaux d'écoute tels que hochements de tête, onomatopées de type 'hum hum'). [16] ont proposé un modèle basé sur la notion de *resonance*. Il interprète le comportement de l'humain en termes des capacités dont lui même est doté. L'objectif étant que cette projection de l'humain relativement à l'agent artificielle va permettre d'améliorer la connexion entre eux. Les travaux de [12] s'inspirent également de données assez précises en provenance de la psychologie pour définir des règles comportementales d'interaction d'un personnage virtuel.

Notre proposition se différencie de ces précédents travaux par le fait qu'elle se focalise sur la notion de régulation. Cette dernière devant produire une interaction adaptative et évolutive basée sur des signaux assez bas niveau. Ce mécanisme de régulation est générique et peut alors s'appliquer à différents contextes d'interaction. Il est également basé sur des considérations psychologiques ancrées dans le courant interactionniste plutôt que cognitiviste. Enfin, il n'aborde que les interactions gestuelles sur le corps entier. Nous sommes conscients des limitations que cela implique. Cependant, le principe est générique et pourrait être appliqué à toute modalité ainsi qu'à une approche multimodale. Nous considérons néanmoins que, puisque le corps est notre premier moyen d'interagir avec les autres, des éléments fondamentaux de la perception de leur présence reposent sur cet ancrage physique et temporel qui est également propre au vivant. *Repartir du début*, i.e, du corps, est notre proposition pour identifier et reproduire les phénomènes principaux liés à la co-présence.

Principes du modèle de décision. Généralement, le participant 1 à l'interaction est l'agent virtuel et le participant 2 est un humain capté à l'aide d'un système de capture gestuelle. Cependant, nous pouvons utiliser notre architecture de multiple manière. Par exemple, les deux partenaires pourraient être deux personnages virtuels. Les principaux modules sont expliqués dans le corps du texte



3 Modèle de décision basé sur la métaphore du vivant

La formalisation complète du modèle de décision associée à des exemples est proposée dans [8, 4]. Dans cette section, nous nous contentons d'en rappeler les grandes lignes qui sont illustrées en figure 3. Le modèle repose sur les deux propositions suivantes :

1. Plutôt que de considérer que l'entrée du modèle est une représentation du comportement de l'humain qui est face à l'agent, nous utilisons une mesure de la qualité de l'interaction en cours entre l'humain et l'agent. Pour cela, les positions des différents angles du squelette des participants sont envoyés à un module d'analyse (**étiquette 2**, figure 3). Celui-ci calcule et compare des caractéristiques bas-niveau telles que les positions, les vitesses, la fluidité ou encore l'énergie des gestes déployés par les deux squelettes. 233 caractéristiques sont calculées [15] et inspirées de différents travaux portant sur l'analyse de gestes tels que ceux de [24, 21]. A partir de ces caractéristiques bas-niveau, nous définissons un *couplage* entre les deux partenaires. Ce couplage est représenté par une variable $C(t)$ calculée en temps réel sous la forme d'une somme pondérée de la différence entre les caractéristiques bas-niveau. Formellement, si l'on note $f_i(p_j, t)$ la valeur normalisée de la caractéristique i du participant j au temps t , $w_i \in]0, 1[$ le poids de la caractéristique i et n le nombre de caractéristiques, on peut définir $C(t)$ par l'équation 1.

$$C(t) = 1 - \left\{ \frac{\sum_{i=0}^n w_i * |(f_i(p_1, t) - f_i(p_2, t))|}{\sum_{i=0}^n w_i} \right\} \quad (1)$$

La qualité de l'interaction est une caractéristique difficile à identifier objectivement mais elle a un lien fort avec de l'évolution temporelle du couplage. Nous proposons un modèle qui a pour ambition d'améliorer cette qualité par le biais d'une régulation du couplage. Il s'agit bien là d'une ambition et d'une hypothèse de travail qui cependant a un sens eu égard aux théories sensorimotrices et interactionnistes de la cognition. Habituellement, les modèles de décision ont pour objectif d'améliorer l'interaction sans en mesurer explicitement la qualité, ni même considérer l'interaction comme une entrée. La différence peut paraître subtile mais elle est importante. Par exemple, les modèles utilisant des règles vont décrire ce que l'agent doit faire pour chaque comportement prévu de l'humain. Le problème est que si un comportement non prévus a lieu, la réaction de l'agent va être mal adaptée et casser la crédibilité et la présence qu'il suscite. A l'inverse, prendre en compte la qualité de l'interaction en entrée va limiter ce problème car les règles comportementales ne vont avoir pour objectif que d'améliorer cette qualité, quelque soit le geste de l'humain, et ceux à partir de données bas-niveau. Il est plus facile, à partir de la notion générique de couplage, de définir des comportements génériques pouvant améliorer la qualité de l'interaction. Évidemment, nous ne prétendons pas que le modèle est générique au point de trouver le comportement idéal pour n'importe quel comportement de l'humain. Nous verrons plus loin que certains éléments de sont comportements sont bien spécifiques au contexte de l'interaction. Un contexte étant pour nous une famille d'activité (par exemple, danse, théâtre, sport...). Néanmoins, nous considérons qu'il permet de définir des règles générales qui produisent une adaptation du comportement de l'agent quelque soit le comportement de l'humain. Les particularités spécifiques au contexte ne demandant ensuite qu'un faible effort de modélisation.

2. Les variables, les règles et les équations du modèle sont déterminées à partir d'une métaphore du vivant, issues de la psychologie, décrite dans [11]. Cette métaphore est basée sur l'observation que durant une interaction entre deux personnes, on peut observer trois phases :

i) La *co-regulation* durant laquelle chaque participant adapte son comportement à l'autre selon des phénomènes classiques socialement admis (mimicry, backchannel...).

ii) La *variabilité* apparaît lorsque les participants modulent la co-régulation. Les variations introduites sont également socialement reconnues et ont pour objectif de produire une sorte de *régulation de la co-régulation*. Par exemple, supposons que durant une discussion, une personne signifie à son partenaire, par un hochement de tête, qu'elle est d'accord avec lui et donc qu'il n'a plus besoin d'argumenter d'avantage sur tel ou tel sujet. Supposons que malgré cela, le partenaire en question continue d'argumenter, la première personne pourra lui indi-

quer à nouveau que ce n'est pas la peine de continuer, en amplifiant son hochement de tête. De cette façon, elle régule la co-régulation et si le partenaire arrête enfin d'argumenter, elle va également arrêter son hochement de tête. Ce mécanisme est subtil et très variable. Il n'est jamais complètement identique d'une interaction à l'autre car il est lié à l'influence réciproque entre les participants et donc à leurs variabilités respectives. C'est une des raisons qui rend sa reproduction à l'aide de programmes si difficile.

iii) L'*innovation* apparaît lorsque l'un des deux partenaires souhaite proposer quelque chose de nouveau, généralement par lassitude de l'échange en cours, qui va être interprété comme une cassure ou une évolution nette de cet échange. En effet, dans ce cas, le schéma de co-régulation socialement accepté est rompu. L'autre participant doit s'adapter à la nouvelle proposition et Foget et Garvey considèrent que si cette innovation est acceptée, elle va être progressivement intégrée à un nouveau schéma de co-régulation et faire évoluer les motifs de l'interaction. D'un point de vue psychologique, cet aspect évolutif de l'interaction donne lieu à une propriété plus globale qui est son aspect développemental.

Notre proposition n'intègre pas cet aspect développemental car même s'il est thématiqué en psychologie, il n'est pas formalisé (voir pas formalisable) dans un cadre mathématique fonctionnel. Cependant, elle intègre de nombreux aspects de la proposition de Fogel et Garvey. En effet, nous utilisons une variable aléatoire discrète S modélisant les trois phases possibles de l'interaction (*co-régulation*, *variabilité* et *innovation* ; **étiquette 4**, figure 3). Cette variable évolue dans le temps selon l'histoire du couplage sensorimoteur $C(t)$ des deux partenaires (l'humain et le personnage virtuel).

Pour modéliser l'évolution de S nous utilisons deux variables continues $\beta_v(t)$ et $\beta_i(t)$ qui vont représenter respectivement la tendance à introduire des variations et la tendance à innover (**étiquette 3**, figure 3).

La façon dont ces variables augmentent ou diminuent dépend de l'évolution du couplage $C(t)$, selon l'algorithme suivant :

```

if variation of  $C(t) \leq th_{\beta_v}$  then
     $\beta_v(t) \leftarrow \beta_v(t-1) + \alpha_v * C(t) * (1 - \beta_v(t-1))$ 
else
     $\beta_i(t) \leftarrow \beta_i(t-1) + \alpha_i * (1 - \beta_i(t-1))$ 
end if

```

th_{β_v} est un seuil de tolérance définissant une valeur du couplage minimum au dessous de laquelle il est nécessaire d'augmenter la tendance à entrer en variabilité. α_v et α_i sont respectivement de taux d'augmentation de la tendance à varier et à innover.

Les principes de cet algorithme sont les suivants : si le couplage est plutôt stationnaire, la tendance à produire des variations va s'accroître et ce proportionnellement à la force du couplage. Sinon, c'est la tendance à innover ($\beta_i(t)$) qui va s'accroître mais dans de proportions plus faibles que l'augmentation de la tendance à varier

car $\alpha_i < \alpha_v$. En d'autres termes, si le couplage est *trop* parfait, des variations vont être introduites. De même, si le couplage n'est pas stable, c'est plutôt une innovation qui va être favorisée au bout d'un certain temps. Ces principes reproduisent ceux de la métaphore de la communication vivante. Après que ces deux tendances aient été mises à jour, la phase de l'interaction représentée par la variable S est à nouveau estimée parmi les trois valeurs *co-régulation*, *variabilité* et *innovation*, par le biais d'un tirage aléatoire. Ce tirage est effectué en considérant les valeurs des tendances $\beta_v(t)$ et $\beta_i(t)$ normalisées par le biais d'une fonction *softmax* (voir [8] pour les détails techniques). Si la phase de l'interaction S change, les tendances $\beta_v(t)$ et $\beta_i(t)$ sont mises à jour de la façon suivante :

when S passes to *variability*

$$\beta_v(t) \leftarrow 0$$

$$\beta_i(t) \leftarrow \beta_i(t-1) + \alpha_i * (1 - \beta_i(t-1))$$

when S passes to *innovation* :

$$\beta_i(t) \leftarrow 0$$

En d'autres termes, à chaque fois que l'interaction passe en phase de variabilité, la tendance à entrer en variabilité est remise à zéro et la probabilité d'innover est légèrement augmentée. Un principe similaire est utilisé concernant le passage en phase d'innovation. En effet, lorsqu'une innovation (ou une variation) a lieu, la nécessité d'atteindre la phase correspondante est satisfaite. Remettre à zéro cette tendance permet de laisser l'interaction se stabiliser sur cette nouvelle phase et donc de relancer un processus de régulation réciproque. Les tendances à innover et à varier vont progressivement augmenter à nouveau, en fonction de l'évolution du couplage. Un autre effet de ce mécanisme est que s'il y a des allers et retours entre co-régulation et variabilités, la tendance à innover va augmenter. En effet, un tel cas se produit lorsque le couplage varie beaucoup et souvent. Cela peut signifier que les deux participants ont du mal à s'accorder et ceci peut se résoudre en 'cassant' cette interaction par le biais d'une innovation. À chaque fois que l'interaction entre dans une des trois phases, le comportement que doit adopter l'agent sera dépendant du contexte de l'interaction. C'est pourquoi, trois fonctions génériques (*BehavCo()*, *BehavVa()* et *BehavIn()*; **étiquette 5**, figure 3) sont prévues et doivent être instanciées en fonction du contexte. Chacune d'elle va produire les entrées d'un module de synthèse de mouvement (**étiquette 7**, figure 3). Ce module s'appuie sur la librairie 'Final IK' de unity3D qui permet de mixer captures de mouvements et cinématique inverse. Pour éviter qu'une fois définies ces fonctions génèrent des comportements trop réguliers et systématiques, elles peuvent être dotées de paramètres tels que par exemple la vitesse ou l'expressivité des mouvements générés. Ces paramètres vont alors évoluer grâce à un système de classeurs [7] (**étiquette 6**, figure 3) qui prendra en compte les différentes variables internes du modèle (phase S , tendances à innover et varier mais aussi

des caractéristiques gestuelles relevées par le module d'analyse). Les règles du système de classeurs sont exprimées dans un fichier XML, qui sera spécifique à chaque contexte d'interaction.

Actuellement, elles sont très simples. Par exemple, nous pouvons utiliser l'expressivité mesurée des mouvements de l'humain pour la recopier dans l'expressivité des mouvements du personnage virtuel. Nous provoquons alors une sorte d'imitation et ce quel que soit le geste des deux participants. Ainsi la variabilité intrinsèque de l'humain est utilisée afin de rendre le comportement du personnage également variable. Ceci entraîne également le fait que le personnage donnera toujours l'impression de proposer quelque chose en lien avec ce que fait l'humain, en dehors de toute expression exhaustive de tous les gestes et toutes les situations possibles. Le code sera mis à disposition sur le site du projet ¹.

3.1 Exergame de fitness

Nous avons appliqué notre modèle de décision à un exergame de fitness. Ce type d'exergame se diffuse largement avec l'arrivée des capteurs corporels à bas coût tels que la Kinect ou la wiifit. Cependant, les jeux grand public ne gèrent pas une régulation du couplage sensorimoteur. Ils utilisent parfois un personnage virtuel mais celui-ci est utilisé pour montrer des exemples prédéfinis non variables. Le choix de ce type d'exergame nous permet de nous limiter à une interaction corporelle tout en utilisant un contexte pour lequel l'influence mutuelle entre un coach et un élève joue un rôle important.

Il s'agit donc d'un exergame où un personnage virtuel réalise des mouvements de fitness en même temps qu'un utilisateur humain. L'humain ou le personnage virtuel peuvent jouer le rôle de coach ou d'élève. Le coach va simplement décider des mouvements et de leur variation durant une session, tandis que l'élève va tenter de l'imiter et de le suivre. Ils peuvent tous les deux faire varier leur vitesse. Ceci peut servir pour encourager l'autre à accélérer, par exemple, mais permet également de s'adapter à sa vitesse. Des exemples d'interaction réalisés sont visualisables sur les vidéos du site web du projet ².

L'influence mutuelle lorsque deux humains interagissent est très subtile voir inconsciente et donc difficile à évaluer par les participants ; lorsque l'un des humains est remplacé par un personnage virtuel, c'est également le cas. Cependant, dans le cadre de notre exergame, l'humain doit prêter attention à son partenaire. Nous avons opté pour deux interactions asymétriques – dans un cas l'humain est actif car il est le coach et dans l'autre cas il est passif car il est l'élève – car dans une étude précédente, nous avons montré que la perception subjective du couplage est d'autant plus forte que les utilisateurs produisaient un effort pour maintenir l'interaction [5]. Dans cette étude, il n'y avait pas d'asymétrie de rôle et il était délicat de discriminer si la nature

1. <http://www.ingredible.fr>

2. http://www.ingredible.fr/?page_id=201

adaptative de l’agent avait une réelle influence sur la perception de l’interaction.

Moyennant un paramétrage adéquat du modèle, celui-ci permet de générer les deux rôles distincts du personnage virtuel : Nous avons d’abord choisi que le calcul du couplage se ferait à partir de la vitesse et d’un étiquetage des gestes réalisés par les deux participants (agent et humain). Ceci se traduit par le fait que dans l’équation 1, des poids significatifs ont été attribués aux caractéristiques correspondant à la vitesse et à l’identification du geste en court. Les poids des autres caractéristiques ont été mis à 0. Ainsi, si l’humain réalise le même geste que l’agent et le réalise à la même vitesse, le couplage sera maximum. S’il fait le même geste mais à une vitesse très différente, le couplage sera proche de 0.5 (comme s’il fait un geste différent mais avec une vitesse similaire).

Dans un deuxième temps, nous avons défini les trois fonctions comportementales (*BehavCo()*, *BehavVa()* et *BehavIn()*) correspondant aux trois phases d’interaction possibles. Lorsque l’agent est le coach, la fonction de co-régulation consiste à adapter la vitesse de l’agent à celle de l’humain. La fonction de variabilité produit un changement de vitesse ou un changement de mouvement à choisir parmi une liste de gestes *connus*. Un geste est dit *connu* s’il a été produit au moins déjà une fois par les partenaires. Au début, cette liste est vide et elle se remplit au cours du temps, grâce à des passages par des phases d’innovation. En effet, la fonction liée à l’innovation consiste à introduire des gestes *inconnus*. En fait, il y a bien sûr un nombre maximum de gestes possibles (9 enregistrés dans une base de mocap) au delà duquel, l’innovation proposera un geste *connu*. Néanmoins, à l’usage, il fut rare que l’ensemble des gestes possibles ait été proposé lors de nos sessions de tests qui ne duraient que quelques minutes.

Lorsque l’agent est l’élève, la fonction associée à la co-régulation a pour but de reproduire le geste proposé par le coach humain et d’adapter sa vitesse à celui-ci. Durant la phase de variabilité, la fonction de comportement introduit des variations en vitesse (permettant de simuler de la fatigue ou de l’ennui par exemple). La fonction d’innovation consiste à signifier une incompréhension (signe non de la tête, bras croisé par exemple). En effet, lorsque l’agent est élève, l’innovation n’arrivant qu’en cas de mauvais couplages répétés, nous considérons qu’un tel cas n’arrive que si le coach fait un peu ‘n’importe quoi’ (gestes non reconnus par le système par exemple, donc ne correspondant pas aux exercices de fitness).

Pour finir, nous avons introduit deux règles au sein du système de classeurs. La première signifiant que si le couplage est fort pendant un certain temps, la marge de variation de la vitesse augmente. L’autre règle signifiant l’inverse (si le couplage ne peut être assez maintenu, la marge de variation de la vitesse va diminuer).

4 Évaluation

Nous avons utilisé l'exergame qui vient d'être présenté pour évaluer le modèle de décision présenté en section 3. Pour cela, nous avons défini deux scénarios se déclinant chacun selon trois conditions :

1. **Scénario Coach** : l'agent est un coach, l'humain est un élève qui doit suivre le coach en imitant ses mouvements de fitness.

- *Coach insensible* : l'agent ne prend pas en compte l'humain. Il réalise une leçon de façon prédéfinie, les mouvements se succédant de façon systématique quelle que soit la capacité de l'humain à le suivre et à l'imiter.
- *Coach sensible* : c'est dans cette condition que nous utilisons notre modèle de décision. L'agent est attentif au couplage existant entre lui-même et l'humain. Il peut introduire des nouveautés (innovation) à des moments liés à la qualité du couplage. Il peut également réguler sa vitesse pour éviter qu'une trop grande différence ne subsiste entre lui et l'humain et respecter ainsi les limites physiques de l'humain.
- *Coach MOz* : le comportement de l'agent est contrôlé par une technique de Magicien d'Oz.

2. **Scénario Élève** : cette fois, l'agent est l'élève et l'humain joue le rôle de coach. Il doit faire une leçon à l'agent. Pour cela, les trois conditions sont les suivantes :

- *Élève distrait* : la capacité de l'agent à prendre en compte le comportement de son coach est limitée. À chaque fois que l'humain propose un geste, l'agent a 50 % de chance de produire ce même mouvement, sinon il fait un mouvement choisi aléatoirement. De plus, l'agent n'adapte pas sa vitesse à celle de l'humain.
- *Élève attentif* : le comportement de l'agent est contrôlé par notre modèle de décision qui permet de réguler le couplage sensorimoteur existant entre eux. L'agent élève va donc reproduire les gestes proposés par son coach (sauf si le module d'analyse ne reconnaît pas le geste, auquel cas, l'agent signale qu'il ne comprend pas le mouvement par un geste caractéristique d'incompréhension). Il est également capable d'adapter la vitesse de son geste à celle de l'humain tout en étant capable d'introduire une certaine variabilité (accélérer et ralentir) si le couplage est trop 'parfait'. Ceci simule la fatigue ou au contraire une forte motivation signifiée au coach.
- *Élève MOz* : le comportement de l'agent est contrôlé par une technique de Magicien d'Oz.

Le magicien d'Oz est géré par une complice qui est habituée à pratiquer le fitness. Elle est dans une pièce à part et capturée par un capteur de type kinect. Elle peut observer l'humain qui est filmé par une caméra discrète. Chaque mouvement que fait la complice est analysé par le module d'analyse décrit dans [15], qui reconnaît ainsi le geste afin de le déclencher chez l'agent qui est face à l'humain, à partir d'une base de données de mouvements capturés préalablement. La vitesse de l'agent est également copiée sur celle de la complice. Il peut paraître surprenant de passer par une analyse suivie d'une synthèse du geste

sur l'agent alors qu'il serait possible de recopier fidèlement la position du squelette de la complice sur l'agent en temps réel ou de permettre à la complice de déclencher des comportements à l'aide d'une IHM. En réalité, si l'on fait une projection directe du squelette de la kinect sur le personnage virtuel, il est assez fréquent que le bruit associé à ce type de capteur perturbe de façon flagrante l'animation générée, trahissant ainsi le fait qu'il s'agit d'un magicien d'Oz. De plus, il est important que les gestes proposés par la complice ne s'altèrent pas au fil des expérimentations, ce qui n'aurait pas été le cas au bout des deux semaines d'expérimentations (une cinquantaine d'exercices pratiqués par la complice) sans passer par des captures de mouvement. Enfin, l'usage d'une IHM 'manuelle' aurait donné des temps de réactions très variable, voir des erreurs, liées à une interface devant permettre de sélectionner manuellement un geste parmi 9 autres possibles. En réalité, le délai du module d'analyse automatique de gestes est de 0,41 secondes avec un taux de reconnaissance de 96,18 %. Lorsque l'agent MOz est élève, il faut ajouter un délai d'environ 0,48 secondes pour que le complice reconnaisse le geste de l'humain [15]. Ce délai reste en dessous d'une seconde, ce qui est généralement accepté [1] et qui l'est d'autant plus dans ce contexte asymétrique où il est normal qu'il y ait un décalage entre les participants lors des changements de geste.

Le but de l'évaluation est double : d'une part, il s'agit de montrer que le comportement généré par notre modèle est similaire à celui qui serait réalisé par une personne réelle (dans une situation similaire). D'autre part, nous espérons éclairer les liens entre notre proposition de modèle basée sur la métaphore du vivant et les sentiments de crédibilité, de co-présence et d'expérience de jeu. Nous estimons que ces dimensions devraient être améliorées lorsque l'humain interagit avec un agent qui semble attentif à son comportement et capable de s'y adapter. Nous avons formulé 3 hypothèses :

- **Hypothèse 1.** L'évolution de la dynamique du couplage générée par l'interaction entre un humain et notre modèle est plus similaire à celle générée par le magicien d'OZ que par l'agent qui ne prend pas en compte le comportement de l'humain.
- **Hypothèse 2.** Les sujets qui passent l'expérimentation perçoivent bien que l'agent prend en compte leur comportement et s'y adapte. En particulier, nous pensons que le scénario a un impact sur cette perception. Dans le cas du *Scénario Élève*, les sujets sont plus attentifs car ils doivent s'assurer que l'agent arrive à les suivre. Ils devraient donc être plus sensibles à sa capacité d'adaptation.
- **Hypothesis 3.** La crédibilité du comportement de l'agent, le sentiment de présence qu'il suscite ainsi que l'expérience de jeu sont meilleurs dans les conditions *Coach sensible* et *Coach MOz* que dans la condition *Coach insensible*. De même, pour les conditions *Élève attentif* et *Élève MOz* par rapport à la condition *élève distrait*. En d'autres termes, le ressenti du comportement résultant de l'usage de notre modèle se rapproche de celui résultant du magicien d'Oz (qui est contrôlé par un humain réel). Nous espérons donc qu'il n'y aura pas de différences significatives entre les conditions

Configuration expérimentale



Coach sensible et Coach MOz et entre les conditions Élève attentif et Élève MOz.

4.1 Méthode

Les expérimentations ont été réalisées dans une école d'ingénieurs durant deux semaines. Cinquante deux sujets (14 % de femmes, 86 % d'hommes) ont pris part à celles-ci. Chacun d'eux devait passer les deux scénarios (une fois coach, une fois élève) mais dans une seule condition. Ils étaient en fait regroupés par deux et une session était organisée de la façon suivante : nous leur expliquions qu'ils allaient réaliser une séance de sport avec un agent virtuel et que le but de l'interaction était de réaliser une leçon de fitness ensemble, le mieux possible. Nous leur disions également qu'ils devraient être une fois coach, une fois élève. Ils ne savaient pas dans quelle condition ils se trouvaient. Nous leur disions seulement que lorsque l'un était coach, le second était élève et vice-versa. Nous ajoutons que l'agent qui était face à eux pouvait être contrôlé par un programme, par l'autre participant, ou par une personne tierce. La plupart des sujets étaient inquiets à l'idée d'être coach et de trouver les bons gestes que l'agent devait suivre. Pour éviter tout stress, nous les rassurions sur le fait que les gestes se décomposent en des petits sous-ensembles de gestes : trois pour le haut du corps, trois pour le bas et trois qui sont des combinaisons des précédents. Pour les aider à apprendre ces gestes, une vidéo leur était montrée. Cette vidéo commençait par la réalisation des gestes par une personne qui invitait ensuite les sujets à les réaliser avec elle. Pour finir, il était précisé aux participants que s'ils étaient coach, ils pouvaient faire n'importe lesquels des neuf gestes, les répéter autant de fois qu'ils le souhaitent et pouvaient également faire varier le rythme à leur gré. Un exemple de séquence possible de coaching était ensuite montré par vidéo.

Dès que les sujets étaient prêts, ils étaient emmenés dans deux

pièces différentes (comme cela est illustré par la photo de la figure 4.1). Les deux pièces étant équipées de manière similaire par un grand écran vertical, une kinect ainsi qu'un tapis définissant la zone au sein de laquelle les mouvements devaient être réalisés. Un rôle était assigné aux participants et il était laissé deux minutes au coach pour qu'il imagine le début de la séquence qu'il allait enseigner à son élève. Afin d'aider les sujet à se remémorer les gestes, des photos illustrant chacun d'eux, associées à leurs noms, étaient accrochées au mur.

Pendant ce temps, les programmes étaient lancés et un agent virtuel apparaissait sur chaque écran de TV. Puisque nous nous focalisons sur l'interaction corporelle, l'agent est représenté de façon minimaliste (stick figure).

Pour démarrer l'interaction, l'agent virtuel faisait un geste de la main pour signaler qu'il était prêt. Les participants ou l'agent pouvaient décider d'arrêter la session à n'importe quel moment en faisant une salutation ; en moyenne, une interaction durait 3 minutes environ, avec très peu de variation. Ensuite, les rôles étaient inversés pour une deuxième session d'interaction. À la fin de chaque session, les participants devaient renseigner un questionnaire pour évaluer leur expérience et le comportement de l'agent. Ce questionnaire contenait trente affirmations, affichées dans un ordre aléatoire, associées à une échelle de likert à 6 points que les sujets devaient renseigner (1 = en total désaccord ; 6 = tout à fait d'accord).

Sept questions étaient relatives à la perception de la crédibilité du comportement de l'agent, en comparant se comportement à ce qu'un humain devrait faire. Huit questions portaient sur le sentiment de co-présence et étaient inspirées d'un questionnaire proposé par [1]. Enfin, quatorze questions étaient relative à l'expérience de jeu et issues du GEQ In-Game proposé par [18], à laquelle s'ajoutaient une question sur le sentiment d'engagement qu'ils avaient ressenti. L'ensemble de ces questions est reporté dans le tableau 4.1. Enfin, durant les interactions, nous traçons le couplage ($C(t)$ présenté en section 3) et calculé par notre système toutes les 0,5 secondes (quelque soit la condition, ce couplage est mesurable). Les participants étaient ensuite invités à prendre une collation avant de repartir.

Nous avons enregistré 104 interactions parmi lesquelles cinq ont du être retirées : trois à cause d'un problème technique et les deux autres car correspondant à des participants qui ont eu de gros problèmes pour réaliser les mouvements de fitness. Nous avons donc obtenu des données correspondant à :

- 18 sujets, âgés de 17 à 36 ans (*Mediane* = 19,5), pour la condition *Coach insensible* ;
- 15 sujets, âgés de 18 à 23 ans (*Mediane* = 18), pour la condition *Coach sensible* ;
- 17 sujet, âgés de 15 à 45 ans (*Mediane* = 21), pour la condition *Coach MOz* ;
- 15 sujets, âgés de 17 à 42 ans (*Mediane* = 21), pour la condition *Élève distrait* ;
- 19 sujets, âgés de 15 à 45 ans (*Mediane* = 19), pour la condition

Les trente affirmations de notre questionnaire

Dimension	Question
<i>Crédibilité</i>	q1. Je pense que l'agent n'était pas contrôlé par un autre personne. q2. J'ai eu l'impression que le comportement du coach/élève était réaliste. q3. Le comportement de l'agent m'a fait penser à celui d'un humain. q4. Je ne pense pas que l'agent se comportait comme une personne réelle. q5. Je pense que l'agent était contrôlé par quelqu'un d'autre. q6. L'idée que l'agent virtuel ne se comportait pas comme le ferait un coach/élève réel m'a traversé l'esprit. q7. J'ai eu l'impression que l'agent n'était qu'un système artificiel et non une personne réelle.
<i>Co-présence</i>	q8. J'ai eu l'impression que l'agent prenait en compte ma présence pour agir. q9. J'ai eu l'impression que l'agent considérait mon comportement pour agir. q10. J'ai eu l'impression que l'agent essayait d'adapter son comportement au mien. q11. J'ai eu l'impression que l'agent n'avait rien à faire du fait que je sois là pour agir. q12. J'ai eu l'impression d'être en présence d'un autre (personne, être). q13. J'ai eu l'impression que l'agent virtuel était attentif ou réceptif à moi. q14. J'ai eu le sentiment d'être avec l'agent, dans la même pièce. q15. J'ai eu le sentiment de jouer avec l'agent.
<i>Expérience du jeu / Engagement</i>	q16. J'ai eu l'impression d'être impliqué(e) dans le jeu. q17. J'ai eu l'impression de bien m'en sortir, de réussir à jouer mon rôle. q18. J'ai eu le sentiment d'être compétent(e). q19. J'étais intéressé(e) par le principe du jeu. q20. J'ai trouvé le jeu impressionnant. q21. J'ai oublié ce qui m'entourait. q22. J'étais complètement absorbé(e). q23. Je me suis senti(e) frustré(e). q24. Je me suis senti(e) énervé(e), agacé(e). q25. J'ai eu l'impression d'avoir été défié(e) par l'agent. q26. J'ai dû faire un effort pour arriver à jouer. q27. Je me suis ennuyé(e). q28. J'ai trouvé ça très fatiguant. q29. Je me suis senti(e) content(e) durant l'interaction. q30. J'étais à l'aise durant l'interaction.

Résultats du test Wilcoxon unilatéral pour le *Scénario Coach*.
 W_- indique l'hypothèse alternative inférieur et
 W_+ indique l'hypothèse alternative supérieur

Questions	<i>Coach insensible– Coach sensible</i>	<i>Coach insensible– Coach MOz</i>	<i>Coach sensible– Coach MOz</i>
q8	$W_- = 65, p < ,05$		$W_+ = 191, p < ,05$
q9	$W_- = 57, p < ,05$		$W_+ = 196, p < ,05$
q13	$W_- = 86, p < ,05$		$W_+ = 177,5, p < ,05$
q17			$W_- = 88, p < ,05$
q18	$W_+ = 197,5, p < ,05$		$W_- = 82, p < ,05$
q21	$W_+ = 196,5, p < ,05$	$W_+ = 230, p < ,05$	
q22	$W_+ = 181,5, p < ,05$	$W_+ = 229,5, p < ,05$	
q23	$W_- = 75,5, p < ,05$		
q25	$W_- = 81, p < ,05$		
q28	$W_- = 91,5, p < ,05$		
q29	$W_+ = 212,5, p < ,05$		
q30	$W_+ = 201, p < ,05$		

Élève attentif ;

- 15 sujets, âgés de 18 à 36 ans (*Mediane = 20*), pour la condition *Élève MOz* ;

Il n'a pas été possible d'avoir exactement le même nombre de participants pour chaque condition en raison des aléas techniques mais aussi du fait que ces participations étaient liées à la bonne volonté des étudiants qui n'étaient pas toujours facile à convaincre. Pour finir, nous avons collecté autant d'interactions qu'il a été possible en deux semaines.

4.2 Résultats

Nous évaluons les données des deux scénarios de façon séparée. Dans un premier temps, nous représentons les histogrammes des couplages pour les 3 conditions du *Scénario Coach* et du *Scénario Élève*.

Ils représentent la répartition des intervalles de valeurs du couplage durant toutes les interactions. Nous avons découpé l'amplitude des valeurs de couplage en 10 intervalles de 0,1 chacun. Ce choix est pragmatique et n'a que pour justification que la lisibilité graphique de ces histogramme. Le résultat est reporté figure 4.3. Pour le *Scénario Coach*, l'histogramme de la condition *Coach sensible* (notre modèle) est plus similaire à la condition *Coach MOz* que de la condition *Coach insensible*. Pour le *Scénario Élève*, le temps passé pour des valeurs élevées de couplage est plus similaire entre les conditions *Élève attentif* et *Élève MOz* qu'entre les conditions *Élève MOz* et *Élève distrait*. Ce n'est par contre pas le cas pour le temps passé pour des valeurs faibles

Résultats du test Wilcoxon unilatéral pour le *Scénario Élève*.
 W_- indique l'hypothèse alternative inférieur et
 W_+ indique l'hypothèse alternative supérieur

Questions	<i>Élève distrait– Élève attentif</i>	<i>Élève distrait– Élève MOz</i>	<i>Élève attentif– Élève MOz</i>
q2		$W_- = 55,5$, $p < ,05$	
q3		$W_- = 57,5$, $p < ,05$	$W_- = 91,5$, $p < ,05$
q6			$W_+ = 204$, $p < ,05$
q8	$W_- = 46$, $p < ,05$	$W_- = 24,5$, $p < ,05$	
q9	$W_- = 75$, $p < ,05$	$W_- = 33$, $p < ,05$	
q10	$W_- = 60$, $p < ,05$	$W_- = 50,5$, $p < ,05$	
q11	$W_+ = 228,5$, $p < ,05$	$W_+ = 204$, $p < ,05$	
q13	$W_- = 81$, $p < ,05$	$W_- = 27,5$, $p < ,05$	$W_- = 89,5$, $p < ,05$
q14	$W_- = 74,5$, $p < ,05$	$W_- = 51$, $p < ,05$	
q15	$W_- = 73,5$, $p < ,05$	$W_- = 37$, $p < ,05$	
q16	$W_- = 65$, $p < ,05$	$W_- = 49$, $p < ,05$	
q17	$W_- = 76$, $p < ,05$	$W_- = 45,5$, $p < ,05$	
q18	$W_- = 86$, $p < ,05$	$W_- = 55,5$, $p < ,05$	
q21		$W_- = 40,5$, $p < ,05$	$W_- = 72,5$, $p < ,05$
q22			$W_- = 90,5$, $p < ,05$
q23	$W_+ = 215,5$, $p < ,05$	$W_+ = 184,5$, $p < ,05$	
q24	$W_+ = 196$, $p < ,05$	$W_+ = 163$, $p < ,05$	
q25		$W_+ = 154,5$, $p < ,05$	
q26			$W_+ = 190,5$, $p < ,05$
q29	$W_- = 81,5$, $p < ,05$	$W_- = 72,5$, $p < ,05$	
q30	$W_- = 70,5$, $p < ,05$	$W_- = 37,5$, $p < ,05$	

de couplage qui est plus faible pour la condition *Élève attentif* que pour les conditions *Élève inattentif* et *Élève MOz*. Une analyse statistique de ces résultats n'a pas de pertinence forte car comme dans toute tentative d'objectivation, le choix de la variable observée perturbe la mesure réalisée. Notre ambition est simplement ici de montrer des similitudes ou différences comportementales de bon sens, que permettent notre modèle.

Concernant les réponses aux questionnaires, nous procédons également en séparant les deux scénarios *Élève* et *Coach* en comparant les trois conditions deux à deux. Etant donné que chaque participant prend part à une seule condition par scénario et que nous nous sommes prémunis d'un effet d'ordre (pour chaque sujet commençant par un des scénarios, il y a un autre sujet qui commence par l'autre scénario), nous considérons que nous sommes dans un design de mesure indépendante. Nous réalisons donc un test de Wilcoxon (qui est un test non paramétrique équivalent au t-test), dans sa version unilatérale pour prendre en compte la direction des changements. Le test a été réalisé avec l'outil **R** et tous les résultats significatifs (rejetant l'hypothèse nulle avec $p < ,05$ correspondant au fait que la distributions des réponses à deux questions sont similaires) sont reportés dans les tableaux 4.1 et 4.1, associés aux hypothèses alternatives *inférieur* (W_-) ou *supérieur* (W_+) donnant l'information sur la direction du changement.

Pour le *Scénario Élève* (pour lequel l'agent est l'élève), nous obtenons des résultats significatifs pour plusieurs questions. Le comportement de l'agent dans la condition *Élève distrait* est perçu différent de celui des conditions *Élève attentif* et *Élève MOz* que ce soit relatif au sentiment de co-présence, d'engagement ou d'expérience de jeu. Ils montrent que les personnes ressentent une moins forte présence de l'agent lorsqu'il est en condition *Élève distrait* que lorsqu'il est en condition *Élève attentif* ou *Élève MOz*. Dans ces deux dernières conditions par exemple, l'agent est jugé plus concerné par le comportement de l'utilisateur ($q8$, $q9$, $q13$), davantage capable d'adapter son comportement à celui de l'humain ($q10$), plus présent ($q12$, $q14$), moins indifférent à l'humain ($q11$).

Les participants se sentent également plus engagés dans le jeu ($q16$) dans les conditions *Élève attentif* et *Élève MOz* que dans la condition *Élève distrait*. Ils sont également moins frustrés ($q23$) et irritables ($q24$). Comme nous l'espérions, il y a très peu de différences significatives entre les conditions *Élève attentif* et *Élève MOz*, ce qui est très positif pour notre modèle. La condition *Élève MOz* a toutefois tendance à être légèrement mieux évaluée en général. L'agent étant considéré plus crédible ($q3$, $q6$), plus attentif au comportement de l'humain ($q13$). De même, les participants se sont sentis plus absorbés dans l'interaction avec cette condition ($q21$, $q22$).

Dans le cas du *Scénario Coach* (pour lequel l'agent est coach), les résultats sont plus décevants. La condition *Coach MOz* n'est pratiquement jamais évaluée comme meilleure que les deux autres. Les participants se sentent juste plus compétents ($q17$) et habiles ($q18$) dans cette condition *Coach MOz* que dans la condition *Coach sensible*.

Trois questions ($q8$, $q9$, $q13$) parmi 8 dédiées à l'évaluation de la co-présence donnent un meilleur score à la condition *Coach sensible* qu'à la condition *Coach insensible* et *Coach MOz*. L'expérience de jeu est jugée meilleure en condition *Coach insensible* qu'en condition *Coach sensible* (8 questions sur 15) : Les participants se sentent moins défiés ($q25$) et plus fatigués ($q28$), compétents ($q18$), absorbés dans le jeu ($q21$, $q22$) et satisfaits ($q29$, $q30$) dans la condition *Coach insensible* que dans la condition *Coach sensible*.

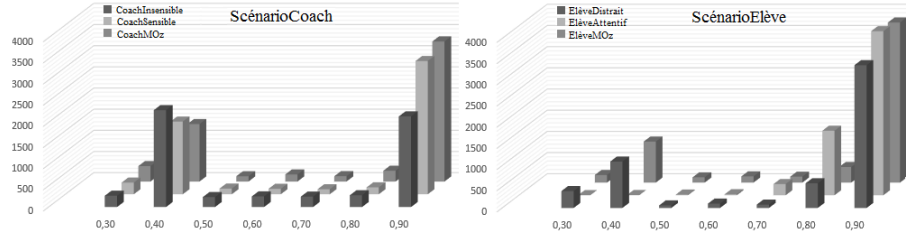
4.3 Discussion

Dans [8], nous avons décrit un modèle reproduisant les principes de la métaphore du vivant de Fogel et Garvey [11] et qui permet de gérer le passage entre les trois phases d'interaction que sont la co-régulation, la variabilité et l'innovation. Avec l'expérimentation que nous avons menée, nous pouvons comparer l'évolution temporelle du couplage obtenue grâce à ce modèle avec celle issue d'une interaction avec un humain (en magicien d'Oz) ou avec un programme ne régulant pas le couplage. Les histogrammes de la figure 4.3 sont des données objectives qui montrent que notre modèle est capable de reproduire des interactions dont certaines propriétés sont proches de celles obtenues avec le magicien d'Oz.

En effet, pour le *Scénario Coach*, les conditions *Coach sensible* et *Coach MOz* partagent des motifs de répartition du couplage dans le temps très similaires. La dynamique exprimée par ces patterns est très fortement liée à celle de l'interaction entre l'agent et l'humain. Notre modèle génère donc des comportements très proches de ceux d'un humain alors que ce n'est pas le cas pour l'agent non adaptatif. Pour le *Scénario Élève* les valeurs élevées du couplage sont plus fréquentes dans la condition *Élève attentif* que dans les conditions *Élève distrait* et *Élève MOz*, alors qu'elles sont moins fréquentes pour des valeurs faibles de couplage. Ceci est dû au fait que l'*Élève attentif* est programmé pour suivre les mouvements du coach de façon très similaire. Il l'imité un peu trop parfaitement, générant un taux de couplage très élevé. Cette stratégie a été mal paramétrée et comme nous le verrons un peu plus loin, elle nuit à la crédibilité de l'agent. Les données accumulées durant l'expérimentation pourront nous permettre de mieux paramétrer ce scénario, en réduisant par exemple le temps nécessaire de fort couplage avant d'entrer en variabilité. En conclusion, les histogrammes de la figure 4.3 montrent que nous obtenons tout de même un résultat globalement satisfaisant qui maintient notre première hypothèse.

Concernant l'analyse du questionnaire. Il est clair que nous n'obtenons pas de résultats probants sur la crédibilité du comportement de l'agent. Ce résultat décevant peut être expliqué par différents facteurs. Tout d'abord, les gestes produits par l'agent, même s'ils sont variés et modulés en vitesse, sont tous issus de captures de mouvements, quelque soit le modèle d'agent utilisé. L'adaptation de l'expressivité de l'agent sur simplement un critère de vitesse n'est manifestement pas suffisante

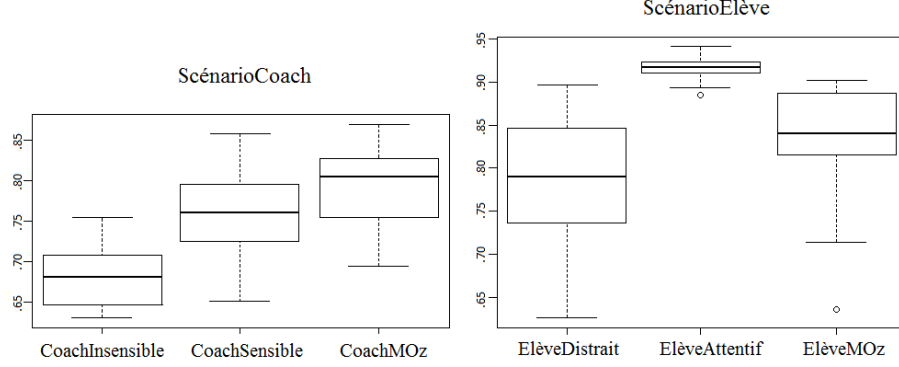
Histogrammes des couplages pour les 3 conditions du Scénario Coach
et du Scénario Élève



pour influencer la crédibilité perçue. Ensuite, même si nous avons insisté auprès des participants sur le fait de ne prêter attention qu’au comportement et non à l’aspect de l’agent, il n’est pas garanti que cet aspect (de type stick figure) n’ait tout de même pas eu une influence négative sur la réponse aux questions liées à la crédibilité. Néanmoins, les résultats liés à l’évaluation de la co-présence et à l’expérience de jeu sont plus riches d’enseignement.

Concernant le premier scénario, le test de Wilcoxon montre que les participants ressentent une différence entre le coach non adaptatif et le coach contrôlé par notre modèle. Notre agent est jugé plus présent, même si la différence n’est pas radicale puisqu’elle n’est statistiquement significative que pour un faible nombre de questions (3 sur 8). De plus, le coach sensible est perçu différemment du coach insensible relativement à l’expérience de jeu. Cette différence n’est par contre pas toujours en faveur du coach sensible. Si les participants sont moins frustrés d’interagir avec le coach sensible, ils sont plus absorbés dans le jeu et plus satisfait du coach insensible ! Ce qui est encore plus surprenant, c’est que le coach MOz n’a jamais de meilleurs scores que le coach insensible. Par conséquent, même si comme nous l’espérons, nous ne trouvons pas de différences significatives entre le coach sensible et le coach contrôlé en MOz, ce résultat n’est pas très intéressant et ne permet pas d’affirmer que le comportement de notre agent sensible est proche du comportement d’un humain dans un contexte similaire. Par ailleurs, les boîtes à moustaches de la figure 4.3 (graphes de gauche), montrent bien que le couplage entre l’agent et les participants est significativement plus élevé avec le coach sensible et le coach MOz qu’avec le coach insensible et qu’il est un peu meilleur en condition coach MOz qu’en condition coach sensible. Ceci confirme bien que notre modèle (coach sensible) permet de se rapprocher d’un comportement en MOz en ce qui concerne le niveau de couplage. Mais mis en regard avec les réponses au questionnaire, cela montre que ce n’est pas suffisant pour améliorer le ressenti des interactions par les participants. Nous supposons qu’une des raisons expliquant ce faible résultat est liée à l’objectif du scénario. En effet, pour ce dernier, les participants doivent simplement suivre le mieux possible les gestes du coach. L’effort qu’ils produisent leur permet de réussir relativement facilement cet objectif et peut leur donner l’impression que l’interaction *se passe bien*, quelque soit la condition (les différents types de coaches). La régulation du cou-

Boîtes à moustaches du couplage moyen pour les scénarios *Coach* et *Élève*



plage avec les participants, réalisée par le coach sensible ou le coach WOz, n'est en fait pas facilement perceptible en regard de l'objectif à atteindre par le participant. Ceci confirme les résultats que nous avons obtenus dans une évaluation précédente [5].

Pour le deuxième scénario, les résultats sont nettement plus concluants. Ils montrent que les participants ressentent une différence entre l'agent peu adaptatif (élève distract) et les agents pilotés par notre modèle (élève attentif) ou par un magicien d'Oz (élève MOz). De plus, cela va dans le sens espéré pour la présence qui est bien plus ressentie avec notre modèle et le MOz qu'avec l'élève distract. C'est également le cas pour l'expérience de jeu qui est meilleure avec l'élève MOz et l'élève attentif qu'avec l'élève distract. Rappelons que dans ce scénario, même l'élève distract n'est pas totalement indifférent à ce que fait l'humain, il n'est simplement pas basé sur la régulation de la valeur du couplage. Pourtant, notre modèle obtient de bien meilleures notes, ce qui montre l'importance de cette régulation. Comme le montre la partie droite de la figure 4.3, l'élève attentif (notre modèle de régulation du couplage) est plus fortement couplé avec les participants que l'élève distract ou que l'élève piloté en magicien d'Oz. Ceci s'explique en partie par le fait qu'en magicien d'Oz, le délai de réaction de l'agent est légèrement plus élevé que pour l'agent attentif (temps d'analyse et de reconnaissance du geste du magicien avant reproduction). Par ailleurs, il y a un facteur humain incontrôlable qui est la perte d'attention ou de motivation de l'humain qui pilote l'agent et qui adaptera donc son comportement de façon moins rigoureuse au fil du temps que ne le fait l'*Élève attentif*. Néanmoins, cette différence entre magicien d'Oz et comportement de l'*Élève attentif* n'a généralement pas d'impact sur les réponses au questionnaire. Les rares différences sont tout de même en faveur du magicien d'Oz qui est jugé par exemple plus réaliste. Par conséquent, la stratégie d'adaptation qui est utilisée par l'*Élève attentif* est trop stricte et il serait aisé de la relâcher quelque peu.

Ces résultats confortent en partie notre troisième hypothèse : pour le *Scénario Élève*, le comportement adaptatif de l'agent sensible a une influence sur le sentiment de co-présence et améliore l'expérience de jeu. Associés aux résultats faibles que nous avons obtenu pour le *Scé-*

nario Coach, ils confirment notre deuxième hypothèse, à savoir que le scénario a également un impact important sur la perception de l’agent. L’aspect adaptatif du comportement est plus important lorsque les participants ont un rôle actif, i.e. lorsqu’ils cherchent à influencer le comportement de l’agent et qu’ils sont donc attentifs à ses réactions, ce qui est le cas ici lorsqu’ils sont coachs. À l’inverse, lorsqu’il sont passifs, dans le sens où ils n’ont qu’à suivre l’agent, ce sont les efforts qu’ils doivent fournir qui exercent une plus grande influence sur le ressenti de l’interaction relativement au comportement de l’agent.

5 Conclusion et perspectives

Dans cet article nous avons argumenté pour un modèle comportemental d’agent virtuel interactif basé sur l’analyse du couplage qu’il exerce avec son partenaire humain et sur une métaphore dite de la communication vivante permettant de réguler ce couplage. L’idée étant que ce type de modèle pouvait améliorer la crédibilité, la co-présence et l’engagement suscités par le personnage virtuel. À travers notre expérimentation, nous avons montré que notre modèle peut générer une adaptation comportementale proche de celle qui est réalisée par un humain. Cependant cette proximité comportementale est déduite de mesures objectives mais n’est pas nécessairement validée par l’évaluation subjective des participants. Le dernier scénario est tout de même plus satisfaisant car l’agent est bien évalué en termes de co-présence et d’engagement et c’est vraisemblablement parce que les participants doivent être attentifs à son comportement pour mener à bien leur rôle de coach. Dans l’autre cas, les participants sont plus passifs, même s’ils doivent faire un effort physique, ils se contentent de suivre et cela n’est pas propice à l’observation du comportement adaptatif de l’agent. Dans l’ensemble, ces résultats sont tout de même encourageants et confirment que la régulation du couplage peut être prometteuse dans certains cas. Il faut juste préciser ces cas et sans nul doute, améliorer la richesse de cette régulation. Par exemple, si nous augmentions le nombre de caractéristiques du mouvement (comme l’amplitude, l’énergie ou l’accélération), l’agent pourrait moduler et évaluer pour estimer le couplage, la pertinence du modèle serait accrue.

Nous avons également pu constater qu’un couplage trop ‘parfait’ n’est pas jugé plus crédible et ne suscite pas une plus grande co-présence ou un meilleur engagement. Notre modèle est paramétrable et basé sur des probabilités. Ces éléments peuvent permettre de doser la part d’imprévisibilité du modèle et elle n’était pas optimisée ici. Nous allons néanmoins pouvoir utiliser les données obtenues grâce à ces expérimentations, et en particulier la condition ‘magicien d’Oz’ pour effectuer une telle optimisation. Nous allons également appliquer le modèle à d’autres types de scénarios qui montreront sa généricité. Enfin, les règles du système de classeur pourront être complexifiées et évoluer durant l’interaction afin d’aller vers un aspect développemental de l’interaction évoquée lors de la présentation de la métaphore de la communication vivante. Un tel modèle participera à l’avancée de nos

Acknowledgments

Ce travail a été financé par le projet ANR INGREDIBLE : ANR-12-CORD-001 (<http://www.ingredible.fr>).

Références

- [1] J.N. Bailenson, R.E. Guadagno, E. Aharoni, A. Dimov, A.C. Beall, and Ji. Blascovich. Comparing behavioral and self-report measures of embodied agents : Social presence in immersive virtual environments. In *7th Annual International Workshop on Presence*, pages 216–223, 2004.
- [2] J.N. Bailenson, K. Swinth, and C. Hoyt. Independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual. *Presence : Teleoperators and Virtual Environments*, 14(4) :379–393, 2005.
- [3] E. Bevacqua, E. de Sevin, S. Hyniewska, and C. Pelachanud. A listener model : introducing personality traits. *Journal of Multimodal User Interfaces, special issue Interacting ECAs*, 6(1) :27–38, 2012.
- [4] E. Bevacqua, R. Richard, J. Soler, and P. De Loor. INGREDIBLE : A platform for full body interaction between human and virtual agent that improves co-presence. In *Proceedings of the 3rd International Symposium on Movement and Computing*, July 2016.
- [5] E. Bevacqua, I. Stanković, A. Maatallaoui, A. Nédélec, and P. De Loor. Effects of coupling in human-virtual agent body interaction. In *14th International Conference on Intelligent Virtual Agents*, pages 54–63, 2014.
- [6] Cyril Bossard, Romain Bénard, Pierre De Loor, Gilles Kermarrec, and Jacques Tisseau. An exploratory evaluation of virtual football player’s believability. In *Proceedings of 11th Virtual Reality International Conference (VRIC’09)*, 2009.
- [7] C Buche, C Septseault, and P De Loor. Proposition d’un modèle générique pour l’implémentation d’une famille de systèmes de classeurs. *Revue des Sciences et Technologies de l’Information, série Intelligence Artificielle (RSTI-RIA)*, 20(1) :63–88, 2006.
- [8] 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 International Conference of Computer Animation and Social Agents*, May 2016.

- [9] Pierre De Loor, Kristen Manac'h, and Jacques Tisseau. Enaction-Based Artificial Intelligence : Toward Co-evolution with Humans in the Loop. *Minds and Machines*, 19(3) :319–343, 2009.
- [10] V. Demeure, R. Niewiadomski, and C. Pelachaud. How is believability of a virtual agent related to warmth, competence, personification, and embodiment ? *Presence*, 20(5) :431–448, 2012.
- [11] A. Fogel and A. Garvey. Alive communication. *Infant Behav Dev*, 30(2) :251–7, 2007.
- [12] J. Gratch, A. Hartholt, M. Dehghani, and S.C. Marsella. Virtual Humans : A New Toolkit for Cognitive Science Research. In *Cognitive Science*, 2013.
- [13] V. Groom, C. Nass, T. Chen, A. Nielsen, J.K. Scarborough, and E. Robles. Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, 67(10) :842–849, 2009.
- [14] C Heeter. Being there : the subjective experience of presence. *Teleoperators and Virtual Environments*, 1(2) :262–271, 1992.
- [15] C. Jost, P. De Loor, A. Nédélec, E. Bevacqua, and I. Stanković. Real-time gesture recognition based on motion quality analysis. In *7th International Conference on Intelligent Technologies for Interactive Entertainment*, pages 47–56, 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] R.S. Nagayama. The Uncanny Valley : Effect of Realism on the Impression of Artificial Human Faces. *Presence Teleoperators & Virtual Environments*, 16(4) :337–351, 2007.
- [18] R. Pugliese, K. Förger, and T. Takala. Game Experience When Controlling a Weak Avatar in Full-Body Enaction. In *15th International Conference on Intelligent Virtual Agents*, pages 418–431, 2015.
- [19] M.J. Schuemie, P. van der Straaten, M. Krijn, and C.A. van der Mast. Research on presence in virtual reality : a survey. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 4(2) :183–201, 2001.
- [20] J. Stewart. An enquiry concerning the nature of conceptual categories : a case-study on the social dimension of human cognition. *Front Psychol.*, 654(5) :1–13, 2008.
- [21] A. Truong, H. Boujut, and T. Zaharia. Laban descriptors for gesture recognition and emotional analysis. *The Visual Computer*, 2015.
- [22] H.C. van Vugt, J.F. Hoorn, and E.A. Konijn. Interactive engagement with embodied agents : an empirically validated framework. *Sp. Iss. Computer Animation and Virtual Worlds*, 20(2–3) :195–204, 2009.

- [23] Joshua Wei-Ern Wong and Kevin McGee. Frown more, talk more : Effects of facial expressions in establishing conversational rapport with virtual agents. In *12th International Conference on Intelligent Virtual Agents*, volume 7502 of *LNCS*, pages 419–425, 2012.
- [24] Y. Yin and R. Davis. Real-time continuous gesture recognition for natural Human-Computer Interaction. In *2014 IEEE Symposium on Visual Languages and Human-Centric Computing*, pages 113–120, 2014.