Believability and Co-presence in Human-Virtual Character Interaction

Elisabetta Bevacqua, Romain Richard, Pierre de Loor

Believability and Co-presence in Human-Virtual Character Interaction

Elisabetta Bevacqua, Romain Richard, and Pierre De Loor*

November 23, 2017

Abstract

This paper presents an evaluation of the believability, the feeling of co-presence and the game experience induced by a decision model that allows for adaptive and evolutive body interaction between a human and a virtual agent. The theoretical bases are presented and the model is briefly described. We illustrate and discuss an experiment we conducted to evaluate the model applied in the context of a fitness exergame. This evaluation shows that our model can generate an adaptive body behavior for virtual agents comparable to that a human would perform in context where the user has to pay attention to the agent. This evaluation shows also that this behavior improves the perception of the agent in terms of co-presence and game experience. It also shows that the role of the agent in the scenario has an important impact on the human perception of the agent itself.

1 Introduction

This paper addresses some issues about the real-time bodily interaction between a human and a virtual agent. The assumption behind our study is that, beyond graphical criteria like photo realism or physical realism, some dynamical properties of an interaction have an important influence on the believability and the feeling of co-presence of the virtual character. We also hope to show the link between these properties and the game experience of the user. Generally, co-presence and believability are evaluated regarding rather static parameters like the size of the agent, its facial expression or its role. However, some subtle phenomena that are rooted in our body concern the reactive and evolutive part of the interaction. Two sessions of interaction never produce exactly the same sequences of gestures. Moreover, a gesture proposed by an interactant to communicate with another, will not always prompt the same reaction from the other. This reaction depends on the history of the interaction and particularly

*E. Bevacqua, R. Richard and P. De Loor are with Lab-STICC, ENIB, UBL, Centre Européen de Réalité Virtuelle, Brest, FRANCE. E-mail: {bevacqua,deloor}@enib.fr

on the partners' feeling about the interaction. In [3], we proposed a decision model that aims to reproduce part of these phenomena by introducing an evaluation of the quality of the coupling between the virtual agent and the human as an input of the decision. This paper is mostly dedicated to the evaluation of this model. A minimalist representation of the agent, a stickman, is used in order to focus solely on bodily interaction. Our key findings are promising: 1) An objective comparison shows that the model is able to reproduce some dynamical properties observed in a human-human interaction, even if some aspects of the animation are not perfect. 2) Even in a rigorous independent-measure design, the evaluation of the user's subjective perception indicates that the character controlled by our model can improve the feelings of co-presence and engagement and the game experience like a character piloted by a "Wizard of Oz" (WOz) (that is an unseen human being who operates the agent), and significantly better than a character that does not try to reproduce a sensorimotor coupling with the user. 3) These results also confirm that the role of the user in the scenario has an important impact on the perception of co-presence. In particular, results are positive when the user has to be *attentive* to the agent. The paper is structured as follows: Firstly, we present a state of the art on the evaluation of the believability and co-presence of virtual agents and on decision models close to ours. Section 3 summarizes the main principles of our model and its application to a fitness exergame. Section 4 presents the conducted experiment, the obtained results and their discussion. Our future works are presented in section 5.

2 State of the art

One of the goals when implementing virtual characters is to give users the illusion that the agents behave like real people. A general hypothesis is that the more a virtual agent is similar to a real person, the more the user adopts the same behavior they would adopt with another human being. Two important properties are generally studied: *believability* and *co-presence*. Believability is relative to how the agent fits the human's expectations in terms of behavioral similarity with another human being [4], while co-presence is the feeling to be with another person [8].

The first problem one encounters when addressing these properties, and people's *feeling* about them, is that they are very subjective. The other problem consists in determining the capabilities a virtual agent must be endowed with to improve its believability and to favor the user's feeling of co-presence. Improving only the graphical representation does not seem to be the most effective solution [12] and a perfect representation of a virtual agent could even fall in the uncanny valley if its behavior is not perfect as well. For example, [7] shows that the role of the agent behavior and the congruence of such a behavior with its appearance improve how the agent is perceived by the user. The evaluation of co-presence is an open debate concerning the use of subjective evaluation techniques like questionnaires or objective measures like physiological responses. A fair proposal is to accumulate and compare objective measures and subjective feelings to stress possible causal links between them [1]. The problem is that there are a lot of possible objective measures, too.

The role of the interaction itself is rarely studied in terms of co-presence (experiments often consist in evaluating an agent standing in its virtual environment without interacting with it) and we consider that it is important and maybe crucial to favor the feeling of this property. For instance, we conducted an evaluation in [2] and showed that the effort produced by the human to maintain a sensorimotor coupling with an interactive virtual agent improves effectively the feeling of co-presence. Nonetheless, in this previous study, the agent was not controlled by an autonomous decision model but by a WOz. Since then, we implemented a decision model for real-time autonomous interactive agents and this work is about such a model and its evaluation.

This model focuses on the simulation of the interaction dynamics, following psychological considerations. In the domain of interactive virtual agents, there are different propositions close to ours. For example, some phenomena like mimicry, synchrony or backchannel that are identified as relevant during an interaction between two people, are reproduced in some models. Kopp and colleagues implemented a model based on the notion of *resonance* that interprets the behavior of a human in terms of agent capabilities [10]. This proposition aims at improving the connection between the interactants. In [6], a psychological observation is used to define interactive rules for the virtual agents behavior.

Our proposition is a bit different from the previous ones. It focuses on regulation mechanisms that aim to produce adaptive and evolutive interactions based on low level data. These mechanisms are generic, can be applied to different contexts, and are based on psychology researches. Moreover, differently from the previous cited approaches, it only addresses the full body gestural interaction. We are aware of the limitations due to the utilization of just one communication modality, nonetheless, the underlying mechanisms are generic and could be applied to a full modal case. Moreover, as the body is our first mean of interact with others, we consider that fundamental principles relative to the feeling of co-presence are rooted in the physical and temporal aspect of our life.

3 Aliveness metaphor based decision model

The formalization of the decision model and some examples of its temporal evolution are described in details in [3]. Here, we resume the main principles which are illustrated in Figure 1. This model is based on two main propositions:

1) The input of the model is not just the behavior of the human in front of the agent, but rather a measurement of the quality of the interaction that occurs between the human and the agent. For that, the joint positions of the two participant's skeletons (**tag 1**, Fig. 1) are sent to an analysis module (**tag 2**, Fig. 1). It extracts up to 10 main features, such as limbs position, speed,

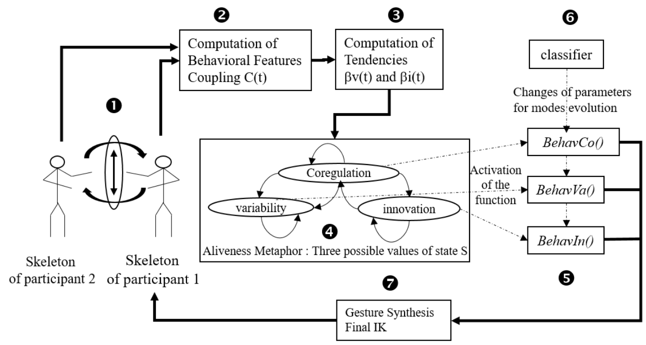

Figure 1: Overview of the decision model. Participant 1 is the virtual agent and participant 2 is a human being tracked with a mocap system.

fluidity, etc.. Since they are computed for each skeleton joint (or a subset of joints) and in all direction (x, y and z), we obtain 233 low-level features. The analysis module can also recognize the current movement performed by the user [9]. We use these data to compute the level of *coupling* between the behavior of the two participants. This coupling is formalized by a variable $C(t)$ computed in real-time as the weighted sum of the difference between the low-level features of the two participants. Let $f_i(p_j, t)$ be the normalized value of the feature i of participant j at time t, $w_i \in]0, 1[$ the weight of the feature i and n the number of features, then $C(t)$ is defined by equation 1:

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

As features are evaluated through a sliding windows on the stream of raw data, they are updated each frame (for example 30 time per second with a Kinect) and so is the computation of the coupling.
There is a link between the coupling and the quality of the interaction and the goal of the model is to improve the quality of the interaction through a regulation of the coupling. Usually, decision models aim at improving the interaction, but they neither measure explicitly its quality nor consider it as an input. The difference is important. For example, for a classical model, rules describe what the agent has to perform for each intended behavior of the human. The problem is that if the actual behavior performed by the human is not envisaged in the model, the reaction of the agent could be inappropriate. As a consequence, the believability and the feeling of co-presence can be broken. Taking into account the quality of the interaction can limit this type of problems, because the rules of the model are expressed in terms of means to regain a good quality of the interaction, whatever the behavior performed by

the human. Through the generic notion of *coupling* between the human and the agent, we can define what kind of generic behavior can improve the quality of the interaction. Obviously, we do not claim that the model is so general that it is able to find the perfect behavior the agent should show for any behavior performed by the user. We will see later that some elements of the model must be explicitly specified for a given context. A context of interaction is a family of interactive activities (for instance, dance, theater, sport...). However, we claim that it allows the definition of more generic rules that produce a better adaptation of the agent behavior with few efforts of modeling for certain contexts of interaction.

2) Variables, rules and equations of the model are based on a psychological metaphor named *Aliveness metaphor*, described in [5], that considers an interaction as a living system. This metaphor is based on the observation that a communication between two people passes through three possible states: i) the *co-regulation* state during which each participant adapts their behavior to that shown by the other interactant through relatively reactive principles socially recognized (mimicry, backchannel...). ii) The *variability* state occurs when participants introduce modulations in their co-regulation. This variation is also socially accepted and aims at producing a kind of *regulation of the co-regulation*. For example, let us imagine someone who nods, for instance, to inform their partner that they agree and that there is no need to argue anymore. If the partner continues to argue, the first person can indicate anew that it is not necessary to go on, by increasing their head nods amplitude. They modulates the co-regulation and if the partner stops to argue, they will also stop to nod. It is worth noticing that variability is very subtle and never exactly the same because it is linked to the reciprocal influence between the interactants and to their own behavioral intra-variability. iii) The *innovation* state generally occurs when one of the interactants wants to offer something new and different that causes a relevant change in the interaction (often because they find that the current co-regulation has lasted long enough). In this case, the socially accepted scheme of the co-regulation (and variability) is broken. The other interactant has to adapt to this new proposition and Fogel and Garvey consider that, if innovations are accepted by the two interactants, they are progressively integrated to the new scheme of co-regulation and then, make the interaction evolve. From a psychological point of view, the evolving aspect of an interaction is part of a more global property: its developmental aspect.

Our proposition does not implement the whole developmental aspect, however, it integrates the main principles of the theory in its formalization. We use a discrete random variable S to model the three states of the interaction (c*o-regulation*, *variability* and *innovation*; **tag 4** , Fig. 1). This variable evolves in time according to the history of the sensorimotor *coupling* $C(t)$ between the two interactants.

To represent the dynamical evolution of S we use two continuous time variables $\beta_v(t)$ and $\beta_i(t)$, which represent respectively the tendency to introduce variations and the tendency to innovate (**tag 3**, Fig. 1).

How these two variables increase or decrease during time depends on the evolution of the coupling $C(t)$:

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} is a tolerant threshold used to determine when the variation of coupling is small enough to increase the tendency to introduce a variation. α_v and α_i are the rates the tendencies to introduce a variation and to introduce an innovation increase at. This pseudo-code represents the following principles: If the coupling does not change during time, the tendency to make some variations increases proportionally to the coupling. Otherwise, it is the tendency to innovate ($\beta_i(t)$) that increases (but more slowly since $\alpha_i < \alpha_v$). In other words, if the coupling is too perfect, variations will be introduced, and if the coupling is not stable, an innovation will be proposed after a while. They reproduce the principles of the aliveness metaphor of the interaction.

After the computation of the tendencies to innovate and to make variations, the state of the interaction S is updated through a random sampling among the three values *co-regulation*, *variability* and *innovation*. This random sampling is done with probabilities computed from the values of $\beta_v(t)$ and $\beta_i(t)$ normalized through a *softmax* function (see [3] for technical details). When an interaction state S changes, the tendencies $\beta_v(t)$ and $\beta_i(t)$ are reevaluated as follows:

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

In short, each time the interaction passes in *variability*, the tendency to variate is reset and the probability to innovate increases. The same principle is used for the tendency to innovate. In other words, when an innovation (or variation) is triggered, the need to reach the corresponding state is satisfied and then the tendency to reach it again is reset. After a while, the tendency to innovate or to introduce a variation will increase again, according to the evolution of the coupling. Another effect of this mechanism is that, if an interchange of co-regulation and variability appears, the tendency to innovate increases. If the coupling varies a lot and often, it could mean that the dyad has problems to interact and so the tendency to innovate increases too. As soon as the interaction passes in variability (or innovation) state, the tendency to offer variation (or innovation) is reset, to give the interaction time to reattain a potential co-regulation state.

Whenever the interaction passes through one of the three states, the behavior that the agent should perform depends strongly on the context of the interaction. For such a reason, three generic functions (*BehavCo()*, *BehavVa()* and *BehavIn()*; **tag 5**, Fig. 1) must be implemented. They describe the behavior that can be performed respectively during the states of co-regulation, variability and innovation. Each one of these functions will produce input for the gesture synthesis module (**tag 7**, Fig. 1). To avoid that, once defined, the behavioral functions produce always the same behavior for a given input, they can be parameterized. For example, the expressivity the agent should show is a parameter. To make the parameters evolve during time and according to the dynamics of the interaction, we use a classifier system which contains rules in the form (condition $\rightarrow$ action), written in a XML file, that depend on the context (**tag 6**, Fig. 1). Currently, the rules are very simple. For example, we can use the expressivity of the human to define the expressivity of the virtual agent and then, to provoke a certain kind of imitation whatever the movement performed by the interactants. Any low-level feature and variable of the model (like S or $\beta_i(t)$ for instance) can appear in the condition of the rules. Any parameter of the behavioral functions can be modified by the action of the rules.

The synthesis module is implemented using the engine Unity3D[1]. The agent can play motion captured animations, but it can also be animated using inverse kinematics. We use, for this purpose, the FinalIK library[2], provided by Root-Motion, which allows for a mixture of motion captured movements and inverse kinematics. So, the synthesis module can play a specific gesture, with variations in expressivity that can be applied at any moment. Since we want to study only body movements, we use a minimalist humanoid representation for the agent, a yellow skeleton created with Bezier curves with cylindric meshes dynamically built on. This representation minimizes some problems, such as collision management and co-articulation between two different gestures, which are not solved in our synthesis module. The whole code of the presented architecture will be at disposal on the project website[3].

3.1 Exergame application

We applied our model to a fitness context. Exergames for sport are flourishing thanks to the more and more affordable body sensors used in edutainment. However, the coupling between a virtual agent and a human during a training session is far from being simple and correctly simulated. We use this application to evaluate our decision model, as described in the next section. We chose this type of context because it reposes exclusively on bodily interaction and mutual influence is an essential factor. The exergame consists of a virtual agent able to perform fitness movements with a human. Both the agent and the user can play the roles of the coach or the student, that is the coach shows movements that the student must imitate. The two interactants can vary their speed, for

[1] `https://unity3d.com`
[2] `http://www.root-motion.com/final-ik.html`
[3] `http://www.ingredible.fr`

example, by accelerating to motivate each other, or they can adapt to the speed of the other, for instance, by slowing down if the partner cannot keep the pace. Examples of this type of interaction can be seen in the videos uploaded on the project website[4].

Mutual influence in human-human interaction can be very subtle and then very hard to consciously feel and evaluate; in human-virtual agent interaction it can be even harder since virtual agents are just an approximation of humans, however, in this context people have necessarily to pay attention to it. We opted for these two asymmetric interactions, that is one in which the users are mainly passive (they have principally to follow the agent lead) and one in which the users are mainly active (they have to make the agent follow them), because of a previous research study we conducted [2]. In this study we investigated people's perception of coupling during an interaction with a virtual agent and we noticed that the effort made by subjects to maintain the interaction could be enough to make them feel that they were coupled. In that evaluation the participants may behave equally actively and passively and it was hard to discriminate when the adaptive agent behavior had a real influence on the feeling of the interaction.

Currently, our model is parametrized in a rather pragmatic fashion: We observed videos of real humans making fitness and tried to obtain similar dynamics. Then, our model is able to generate the agent behavior in both roles. For the exergame application we parameterized the decision model as follows.

Firstly we defined coupling as a combination of the correlation between the speed and the movement performed by the two interactants (agent and human). In other words, in equation 1 we associate weights to all the low-level features which correspond to the position and the speed of each joint and the label of the recognized gesture (if any) and 0 to all the others. For example, if the players perform the same fitness movement at the same speed, the coupling is close to 1, whereas if they perform different movements with different speed, the coupling is close to 0. If they perform the same gesture with very different speed, the coupling is close to 0.5.

Secondly, we define the three behavioral functions ($BehavCo(), BehavVa()$ and $BehavIn()$) for each state of the interaction. When the agent is the coach, the behavioral function for the co-regulation state consists in adapting the agent speed to the human's. In variability state the behavioral function makes the agent offer a variation in speed or change the movement for another known one. A movement is *known* if it has been performed by both partners at least once. At the beginning the list of the known movements is empty and it is filled every time the interaction passes in innovation state. Indeed, the task of the behavioral function associated to this state consists in making the agent offer an *unknown* movement. Currently, this function can add up to 9 fitness movements, which we captured with a motion capture system. When the agent is the student, the function associated to the co-regulation state makes the agent perform the movement offered by the human coach and adapt its speed to the

[4] `http://www.ingredible.fr/?page_id=201`

human's. In variability state, the behavioral function offers variation in speed (to simulate, for example, tiredness or eagerness to move more). Innovation arises in case of bad coupling, for example, when the human's movement is not recognized. The behavioral function, associated to this state, triggers a gesture of non understanding: the agent shakes the hands and the head.

Finally, we write two rules in the classifier. One specifies that if the coupling is good for a while, the range of speed variation will grow. Reciprocally, another rule indicates that if the coupling is low, this range will decrease.

4 Evaluation

As explained in section 3.1, we use the exergame application to evaluate our decision model. We define two scenarii each one composed of three conditions:

1. ***CoachScenario:*** the agent plays the role of the coach and the human is the student who has to follow the coach and imitate its fitness movements.

- *InsensitiveCoach*: the agent does not take into account the human, it performs its lesson, one fitness movement after the other, without caring for the human performance. Fitness movements are chosen randomly and repeated a random number of times.

- *SensitiveCoach*: the agent is driven by our decision model. It is autonomous and it is aware of the human, it can introduce innovation by performing an unknown movement and it can offer changes in speed. If the human does not keep the pace, it can adapt again its behavior to that shown by the human, that is the agent respects its student's physical limits.

- *WOzCoach*: the agent behavior is controlled through a Wizard of Oz technique.

2. ***StudentScenario:*** in this scenario the roles are switched, the agent plays the role of the student while the human has to teach it a fitness lesson. Similarly to the first scenario, the second one is divided in three conditions, as well:

- *LazyStudent*: the agent capability to take into account the human behavior is limited. Each time the human offers another fitness movement, the agent has a 50% chance of performing the right movement, otherwise it performs a randomly selected one. The agent never adapts to the human's speed.

- *AttentiveStudent*: the agent is autonomous and driven by our decision model: it is aware of the human and performs all the movements offered by the coach that it can recognize, otherwise it shows a gesture of refusal. It can adapt to the speed shown by the human, but it can also decide to vary its expressivity (slowing down, speeding up), to simulate for example a student who is getting tired or who wants to do more than what is offered by the coach.

- *WOzStudent*: the agent behavior is controlled through a Wizard of Oz technique.

The WOz, who operates the agent, is managed by somebody used to fitness sports. In another room, she stands in front of a capture device (Microsoft Kinect) and can see the participant through a webcam. Each movement that she performs is analyzed to be recognized and then played by the agent who is projected in front of the participant. The agent speed is modified according to the speed shown by the WOz. The analysis module described in [9] is used for movement and speed recognition. The recognition delay is about 0.41 seconds, and the recognition rate in real-time is 96.18%. Obviously, when the WOz plays the role of the student, a slight delay is introduced: The time needed by the WOz to recognize the participant's movement (about 0.48 seconds [9]) plus the time needed by the analysis module to recognize that performed by the WOz. However, this delay remains under 1 second which is still acceptable. We choose to manage the WOz in this way for two reasons. Firstly, keyboard control is not a good solution, it would introduce a delay of variable duration. This duration would depend on the time needed by the controller to find the right key to click, particularly when the WOz plays the role of the student and the right movement (among nine) must be selected to copy that performed by the human coach. Secondly, we do not want to map directly the WOz's behavior on the virtual agent since we would map also parasite movements (such as scratching one's nose, hesitations...) that our system cannot generate. We fear that the subjects would recognize the WOz immediately. Besides, the WOz's motivation does not remain the same all along the experiment (which lasted two weeks): in the 50^{th} interaction the fitness movements are not performed with the same enthusiasm as in the first one and we do not want to introduce any bias due to the WOz's boredom and tiredness. To avoid the latter problem and to give the WOz away too easily, we decide that, disregarding the condition, the agent has to perform the nine fitness movements always in the same way, that is by playing the corresponding motion captured file. The WOz can decide only for the movement to perform and its speed, but not how this movement is actually displayed.

The aim of this evaluation is twofold: On the one hand, we want to show that the behavior generated by our model is similar to that performed by a real person in the same situation. On the other hand, we aim to investigate the relationship between our decision model and the agent behavior believability, the feeling of co-presence and the game experience. We believe that these three dimensions improve when the user plays with an agent that appears aware of human behavior and able to adapt to it. Thus, we formulate the following hypotheses:

- **Hypothesis 1.** The dynamic evolution of the coupling generated by our model is more similar to that generated by a WOz than that obtained by an agent who does not take into account the user's behavior.

- **Hypothesis 2.** Subjects perceive when the agent takes into account their behavior and shows adaptation. Especially, we expect that the type of scenario influences the subjects' perception of the agent capacity of adaptation. Since in *StudentScenario*, they are more active, having to make the agent follow them, participants will be more sensitive to the agent adaptive behavior.

- **Hypothesis 3.** We expect higher results for the believability of the agent behavior, the feeling of co-presence and the game experience in conditions *SensitiveCoach* and *WOzCoach* than in condition *InsensitiveCoach* and in conditions *AttentiveStudent* and *WOzStudent* than in condition *LazyStudent*; that is when the agent behavior is determined by our decision model or by the WOz. We hope to find no significant differences between conditions *SensitiveCoach* and *WOzCoach* and between conditions *AttentiveStudent* and *WOzStudent*; that is, we hope that the agent behavior piloted by our model or by the WOz is perceived similarly.

4.1 Method

Fifty-two subjects (14% women, 86% men) took part in the experiment and each one participated in both scenarii but in just one condition (that is they were once coach and once student). Subjects participated in pairs and an experiment session was conducted as follows. The two subjects were explained that they were going to do sport with a virtual agent and that the aim of the interaction was to perform the fitness lesson together the best one can. We told them that they would have participated twice, once as coach and once as student. The subjects did not know in which condition they were participating, we just said them that when one was the coach the other participant was the student and vice versa. They were told also that the agent in front of them could be driven by our system or by the other participant or by a third person.

In order to help the participants to learn the nine movements before starting the interaction with the agent, we showed them a video. This video was divided in three parts. In the first part a girl presented the fitness movements; then she invited the two subjects to perform the movements with her. Finally the participants were told that, while being coach, they could do any of the nine movements, repeat them as many times as they wanted and vary their pace as they will.

As soon as the subjects felt ready, they were placed in two separates rooms (see Figure 2), each one equipped with an equal size TV screen, a Kinect and a green fitness mat which defined the area where they could perform their movements. The participants were assigned roles and we gave to the coach two minutes to think about the sequence of fitness movements they could show to the agent. To help the subjects to remind the movements, we attached pictures of them, labeled with their name, on the wall next to the TV screen. In the

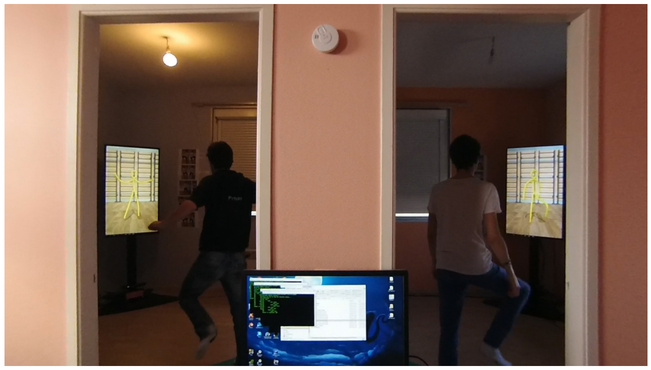

Figure 2: Experiment setting.

meantime all the system was launched and the virtual agents appeared on the screens. As explained in section 3, the agent was a yellow stickman. To start the interaction the virtual agent waved its arm to signal that it was ready. Both the participant and the agent could stop the fitness session at any time by performing a bow. Then the subjects' roles were switched and the second part of the experience could begin. At the end of each interaction, the participants were asked to fill in a questionnaire to judge their experience and the agent behavior. The questionnaire contained thirty statements, showed in a random order, each used a 6-points Likert scale (1 = disagree strongly; 6 = agree strongly). Seven of them assessed the perceived believability of the agent behavior, by addressing the closeness of the agent behavior to that a human could show. Our evaluation of co-presence was inspired from a questionnaire proposed in [1] and it consisted of 8 questions. Finally, to assess the game experience we used the 14 questions in the version of GEQ In-Game proposed in [11] plus another question about the feeling of engagement in the game. All questions are listed in Table 1. For each interaction we logged also the level of coupling every 0.5 seconds.

We recorded 104 interactions, but we had to delete five of them: three because of tracking problems and two because the participants were really not at ease with fitness sports. So, in the end, we collected data from:

- eighteen subjects, age from 17 to 36 ($Median$ = 19.5), for condition *InsensitiveCoach*;
- fifteen subjects, age from 18 to 23 ($Median$ = 18), for condition *SensitiveCoach*;
- seventeen subjects, age from 15 to 45 ($Median = 21$), for condition *WOzCoach*;
- fifteen subjects, age from 17 to 42 ($Median = 21$), for condition *LazyStudent*;

Table 1: The thirty statements in our questionnaire.

Dimension	Question
Believability	**q1.** The agent was not controlled by someone else. **q2.** The coach/student's behavior was believable. **q3.** The agent was behaving like a real person. **q4.** The agent was not behaving like a real person. **q5.** The agent was controlled by somebody else. **q6.** The idea that the agent did not behave like a real person crossed my mind. **q7.** I perceived the agent as a simple computer program.
Co-presence	**q8.** The agent was aware of my behavior. **q9.** The agent was considering my behavior. **q10.** The agent was adapting its behavior to mine. **q11.** The agent did not care at all about my behavior. **q12.** I thought I was in the presence of another being. **q13.** The agent was paying attention to me. **q14.** I felt like being with the agent, in the same room. **q15.** I felt as I was playing with the agent.
Game Experience / Engagement	**q16.** I had the feeling to be engaged in the game. **q17.** I felt successful. **q18.** I felt skillful. **q19.** I was interested in the game's story. **q20.** I found it impressive. **q21.** I forgot everything around me. **q22.** I felt completely absorbed. **q23.** I felt frustrated. **q24.** I felt irritable. **q25.** I felt challenged. **q26.** I had to put a lot of effort into it. **q27.** I felt bored. **q28.** I found it tiresome. **q29.** I felt content. **q30.** I felt good.

Table 2: One-tailed Wilcoxon test results. **L** stands for alternative hypothesis *less* and **G** stands for alternative hypothesis *greater*.

Questions	***LazyStudent–AttentiveStudent***	***LazyStudent–WOzStudent***	***AttentiveStudent–WOzStudent***
q2		**L** W=55.5, p<.05	
q3		**L** W=57.5, p<.05	**L** W=91.5, p<.05
q6			**G** W=204, p<.05
q8	**L** W=46, p<.05	**L** W=24.5, p<.05	
q9	**L** W=75, p<.05	**L** W=33, p<.05	
q10	**L** W=60, p<.05	**L** W=50.5, p<.05	
q11	**G** W=228.5, p<.05	**G** W=204, p<.05	
q13	**L** W=81, p<.05	**L** W=27.5, p<.05	**L** W=89.5, p<.05
q14	**L** W=74.5, p<.05	**L** W=51, p<.05	
q15	**L** W=73.5, p<.05	**L** W=37, p<.05	
q16	**L** W=65, p<.05	**L** W=49, p<.05	
q17	**L** W=76, p<.05	**L** W=45.5, p<.05	
q18	**L** W=86, p<.05	**L** W=55.5, p<.05	
q21		**L** W=40.5, p<.05	**L** W=72.5, p<.05
q22			**L** W=90.5, p<.05
q23	**G** W=215.5, p<.05	**G** W=184.5, p<.05	
q24	**G** W=196, p<.05	**G** W=163, p<.05	
q25		**G** W=154.5, p<.05	
q26			**G** W=190.5, p<.05
q29	**L** W=81.5, p<.05	**L** W=72.5, p<.05	
q30	**L** W=70.5, p<.05	**L** W=37.5, p<.05	

Table 3: One-tailed Wilcoxon test results. **L** stands for alternative hypothesis *less* and **G** stands for alternative hypothesis *greater*.

Questions	***InsensitiveCoach–SensitiveCoach***	***InsensitiveCoach–WOzCoach***	***SensitiveCoach–WOzCoach***
q8	**L** W=65, p<.05		**G** W=191, p<.05
q9	**L** W=57, p<.05		**G** W=196, p<.05
q13	**L** W=86, p<.05		**G** W=177.5, p<.05
q17			**L** W=88, p<.05
q18	**G** W=197.5, p<.05		**L** W=82, p<.05
q21	**G** W=196.5, p<.05	**G** W=230, p<.05	
q22	**G** W=181.5, p<.05	**G** W=229.5, p<.05	
q23	**L** W=75.5, p<.05		
q25	**L** W=81, p<.05		
q28	**L** W=91.5, p<.05		
q29	**G** W=212.5, p<.05		
q30	**G** W=201, p<.05		

- nineteen subjects, age from 15 to 45 (*Median = 19*), for condition *AttentiveStudent*;
- fifteen subjects, age from 18 to 36 (*Median = 20*), for condition *WOzStudent*.

We tried to have the same number of participants for each condition, however it was not possible since it depended on the students' good willingness to participate and sometimes it was very hard to convince them. So we decided simply to collect as many interactions as possible.

4.2 Results

We evaluate the collected data in the two scenarii separately.

Firstly, we compute two histograms of the coupling for the three conditions of each scenario. They represent the amount of time that each interval of values the coupling takes during all the interactions. The result is shown in Figure 3. For the *CoachScenario*, the histogram shows that the distribution of the values of the coupling in *SensitiveCoach* (our model) is more similar to that in *WOzCoach* than that in *InsensitiveCoach*. The histogram for *StudentScenario* shows that, for high values, the distribution of the values of the coupling in *AttentiveStudent* (our model) is more similar to that in *WOzStudent* than that in *LazyStudent*. However, it is not the same for low level values which are less frequent for *AttentiveStudent* than for *LazyStudent* and *WOzStudent*.

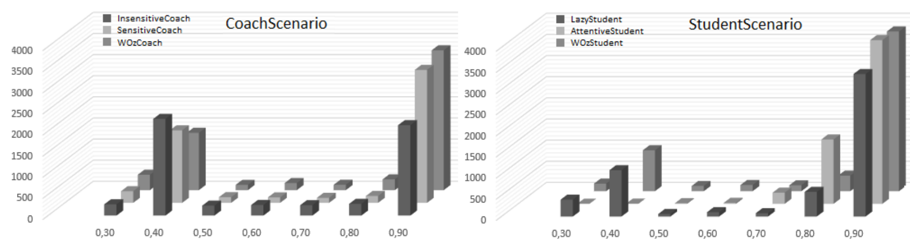

Figure 3: Histogram of coupling for the 3 conditions in *CoachScenario* and in *StudentScenario*.

Secondly, the questionnaire is analyzed by evaluating each statement within the context of the three conditions (*InsensitiveCoach*, *SensitiveCoach* and *WOz-Coach*) and within the context of the three conditions (*LazyStudent*, *AttentiveS-tudent* and *WOzStudent*). We compare the answers to each question pairwise, between each couple of conditions. Since each participant took part in just one condition for each scenario and since we avoided any bias related to the order of the participation to the scenarii (for a subject who started the evaluation by being coach there was another subject who started by being student), we consider that we are in an independent-measures design. For such a reason, we utilize the Wilcoxon test (which is a non-parametric equivalent of the t-test) and we chose the one-tailed version of it to know more about the direction of the change. The test was performed using **R** and all significant results (which reject the null hypothesis ($p<.05$)) are reported in Tables 2 and 3.

Within the context of the *StudentScenario*, when the agent played the role of the student, we find significant results for several questions. The agent behavior expressed in condition *LazyStudent* is perceived very differently from that expressed in conditions *AttentiveStudent* and *WOzStudent*, in terms of the feeling of co-presence, engagement and game experience. People feel a weaker feeling of co-presence with the agent in condition *LazyStudent* than in conditions *AttentiveStudent* and *WOzStudent*. For example, in these two conditions, the agent is judged more aware of the human behavior (*q8*, *q9*, *q13*), more capable of adapting its own behavior to that shown by the human (*q10*), more present (*q12*, *q14*), less uncaring of the human (*q11*). Participants feel also more engaged in the game (*q16*) in conditions *AttentiveStudent* and *WOzS-tudent* than in condition *LazyStudent*. As for the game experience, subjects feel more successful (*q17*), skillful (*q18*), and satisfied (*q29*, *q30*) in conditions *AttentiveStudent* and *WOzStudent* than in condition *LazyStudent*. They are also less frustrated (*q23*) and irritable (*q24*). As we hoped, we have very few significant differences between condition *AttentiveStudent* and *WOzStudent*. In general, the agent behavior in condition *WOzStudent* scores a little better and the agent was judged more believable (*q3*, *q6*), more attentive to the human behavior (*q13*) and people feel more absorbed in the interaction (*q21*, *q22*).

Within the context of the *CoachScenario*, results are poorer and a little unexpected. Condition *WOzCoach* is almost never evaluated better than the other two conditions. People feel simply more successful (*q17*) and skillful (*q18*) in condition *WOzCoach* than in condition *SensitiveCoach*, but that is all. We find that 3 questions (*q8*, *q9*, *q13*) out of the 8 used to evaluate the co-presence score better in condition *SensitiveCoach* than in conditions *InsensitiveCoach* and *WOzCoach* (that is, the agent appears more aware of the human behavior). The game experience is judged better in condition *InsensitiveCoach* than in condition *SensitiveCoach* (in 8 questions out of 15): People feel less challenged (*q25*) and tired (*q28*) and more skillful (*q18*), absorbed in the game (*q21*, *q22*) and satisfied (*q29*, *q30*) in condition *InsensitiveCoach* than in condition *SensitiveCoach*.

4.3 Discussion

In [3], we described how our model reproduces the aliveness metaphor and that it can make the interaction pass through the three states of co-regulation, ordinary variability and innovation. Here, through the experiment we conducted, we compare the dynamic evolution of the coupling generated by our model to that generated by the interaction with a non-adaptive agent and an agent driven by a human. The histograms in Figure 3 present objective measures which show that our model is able to reproduce behavioral properties of the interactions that are close to those obtained with the WOz. Indeed, in *CoachScenario*, conditions *SensitiveCoach* and *WOzCoach* share the same pattern since coupling values are closely distributed. As a consequence, the dynamic evolution of the coupling generated by our model is more similar to that generated by a WOz than that obtained by an non-adaptive agent. In *StudentScenario*, coupling values for condition *AttentiveStudent* are higher than for conditions *LazyStudent* and *WOzStudent* excepted for low levels of coupling. This is probably due to the fact that, in condition *AttentiveStudent*, the agent is programmed to follow closely the fitness movements and the speed proposed by the human coach. It plays its role of student too perfectly, keeping a high level of coupling all the time. As we will see later, such a strategy is not a good idea, since it does not improve the agent believability. The data collected during this experiment will allow us to find a better parameterization, for example by reducing the time needed to enter in variability state. In conclusion, the histograms in Figure 3 show that we obtained an overall satisfactory result, which sustains our first hypothesis.

Let us discuss the results of the questionnaire analysis. First of all, it is important to notice that in both scenarii we obtained essentially no result about the believability of the agent behavior. Our autonomous agent is never judged more believable than the other agents. This disappointing result could be due to several factors: Firstly, in all conditions the agent movements are performed from the same motion captured data, so all fitness gestures are performed always in the same way. Make the agent adapt to just one quality of the movement is probably not enough to influence the human perception of the agent believability. Secondly, even if we asked participants to pay attention only to the agent behavior and not to its physical aspect, the agent representation could have had a negative influence on its believability, since it is just a rough representation of a human. From now on we will discuss only the results on the other dimensions we took into account: co-presence and game experience.

With regard to the first scenario, the results of the Wilcoxon test show that participants feel a slight difference between the non-adaptive insensitive coach and the virtual coach driven by our system. The autonomous agent is judged more present, however this difference is not very eloquent since it is statistically significant only for a small number of questions (3 out of 8). Moreover, the sensitive coach, with respect to the insensitive one, was perceived quite differently in terms of game experience but the results are not always in favor of the coach driven by our model. If people are less frustrated interacting with the sensitive coach, they also feel more absorbed in the game and more skilled and content with the insensitive coach. What surprises us even more is that the WOz never scores better than the non-adaptive insensitive agent. So, even if, as we hoped for, we do not find significant differences between the autonomous agent and the agent piloted by the WOz, this result is not very interesting and does not allow us to say that the agent behavior is closer to the behavior a human would show in the same context. Besides, the box plot diagram of the mean level of coupling (see Figure 4, diagram on the left) shows that the level of coupling is significantly higher in conditions *SensitiveCoach* and *WOzCoach* than in condition *InsensitiveCoach* and that it is a little better in condition *WOzCoach* than in *SensitiveCoach*. That means that both our system and the WOz attain more coupled interactions with the participants but it means also that this is not enough to improve their feeling about the interaction, as the answers to the questionnaire show. We suppose that one reason for such poor results could be the aim of the game. In this scenario people had to follow the best they can the virtual coach, so the effort they produced to follow the agent lead made them recover coupling all the time and gave them the impression of being in the interaction disregarding the condition. The efforts made by both the autonomous agent and the WOz to adapt their behavior to that shown by the subject were not easily perceived and improved just a little the feeling of the interaction in condition *SensitiveCoach*. This confirms a similar result we obtained in the previous evaluation [2].

With regard to the second scenario, the results of the Wilcoxon test are more conclusive. They show that participants clearly feel a difference between the less-adaptive lazy agent and the agents piloted respectively by our system and by the WOz. The feeling of co-presence is stronger when people interact with them than when they interact with the agent in condition *LazyStudent*. Moreover, the game experience scores better for the autonomous agent and the WOz than for the less adaptive agent. Let us notice that, in this scenario, the agent in condition *LazyStudent* is not completely indifferent to the human and, even in this case, our model is judged significantly better. As shown in Figure 4 (diagram on the right), the autonomous agent is highly more coupled with the participants than the agent in the other conditions, even more than the agent piloted by the WOz. This could be due to different facts: Firstly the agent in condition *WOzStudent* suffers of a slight delay with respect to the

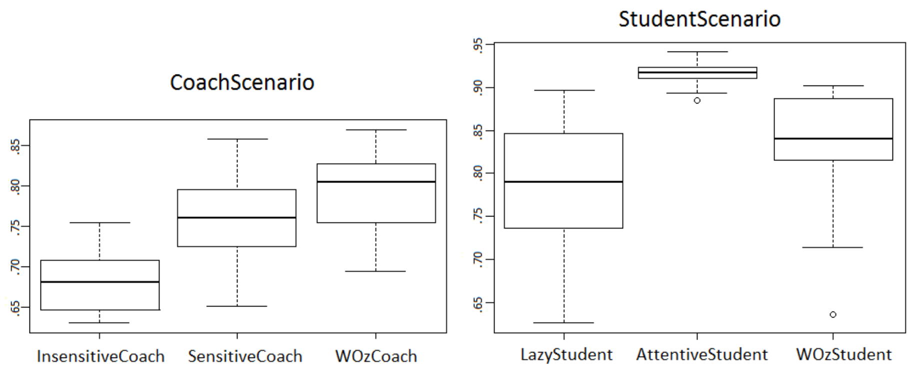

Figure 4: Box plot diagrams of the mean level of coupling.

autonomous agent. Secondly, maybe, because of a human factor that we cannot control, the WOz adapts her behavior to that performed by the subject less often than the agent in condition *AttentiveStudent* which was parameterized to adapt its behavior quite often. What is really interesting is that, even though the autonomous agent is better coupled with participants, the answers to the questionnaire reveal almost no difference between this agent and that piloted by the WOz. Actually, the few significant results we found are in favor of the WOz who, for example, is judged more realistic. So, too much adaptation is not a good strategy to improve the virtual agents behavior.

These results sustain just part of our third hypothesis: In *StudentScenario*, an adaptive agent behavior has a positive influence on the perception of the feeling of co-presence and improves the game experience. Together with the poor results we obtain for the first scenario, they confirm our second hypothesis and show an effect due to the type of scenario. The adaptive agent behavior has a strong influence on the perception of the interaction mainly when the user exerts an active behavior, that is when the user is the coach in our setting. In this case they are more attentive to the agent behavior and to its capacity to adapt. On the other hand, when they are passively following the agent, their efforts exert a stronger influence on the feeling of the interaction than the adaptive agent behavior.

5 Conclusion and Perspectives

Through our experiment we showed that the proposed model can generate an adaptive body behavior comparable to that a human would perform. Subjective evaluations reveals that, in context where the human is particularly attentive to the agent, users perceive clearly the adaptive behavior performed by the agent and that it improves their judgment of the agent in terms of co-presence and game experience. On the other hand, when users are mainly passive, their effort to follow the agent seem to be more important than the agent adaptive behavior, which goes almost unnoticed. This result makes us believe that taking coupling into account is an interesting and promising approach and it comforts us in the idea that we should continue on this path by improving the model. For example, we think that better results could be attained by increasing the number of qualities of movement (such as amplitude, energy, acceleration...) the agent should be able to adapt to and that should used to compute the level of coupling.

From the experiment, we conclude also that a too well coupled agent is not well judged in terms of believability, feeling of co-presence and game-experience, maybe because, to simulate human behavior, the agent should be a little more unpredictable. We aim at using the data collected in this evaluation to better parameterize our model. We are also defining a new application of the model to test its genericity and, in the future, we would like to improve the rules of the classifier by implementing a learning mechanism.

Acknowledgment

This work was funded by the ANR INGREDIBLE project: ANR-12-CORD-001 (`http://www.ingredible.fr`).

References

[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 International Workshop on Precence*, pages 216–223, 2004.

[2] 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.

[3] 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.

[4] 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.

[5] A. Fogel and A. Garvey. Alive communication. *Infant Behav Dev*, 30(2):251–7, 2007.

[6] J. Gratch, A. Hartholt, M. Dehghani, and S.C. Marsella. Virtual Humans: A New Toolkit for Cognitive Science Research. In *Cognitive Science*, 2013.

[7] 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.

[8] C Heeter. Being there: the subjective experience of presence. *Teleoperators and Virtual Environments*, 1(2):262–271, 1992.

[9] 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.

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

[11] 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.

[12] 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.