# INGREDIBLE: A platform for full body interaction between human and virtual agent that improves co-presence

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

# INGREDIBLE : A platform for full body interaction between human and virtual agent that improves co-presence


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

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

**Julien Soler**
Virtualys
julien.soler@virtualys.com

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



## ABSTRACT
This paper presents a platform dedicated to a full body interaction between a virtual agent and human or between two virtual agents. It is based on the notion of coupling and the metaphor of the alive communication that come from studies in psychology. The platform, based on a modular architecture, is composed of modules that communicate through messages. Four modules have been implemented for human tracking, motion analysis, decision computation and rendering. The paper describes all of them. Part of the decision module is generic, that is it could be used for different interactions based on sensorimotor, while part of it is strictly dependent on the type of scenario one wants to obtain. An application example for a fitness exergame scenario is also presented in this work.



## Author Keywords
virtual agent, embodied cognition; coupling, co-presence, architecture, platform

## ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous


## INTRODUCTION
When designing virtual interactive agents, a main research problem consists in making the interaction with these virtual entities as human-like as possible. Some properties like the agent realism and behavioral believability, and human's sense of co-presence contribute to such an illusion. Co-presence is the more subjective and complex to address because it corresponds to the feeling to be in presence of another person (*being with*) [1, 5]. Within the national project called INGREDIBLE, we tackle this subject by focusing on the sensorimotor aspect of human-human interactions and its role on people's feeling of mutual presence. Indeed, researchers in the recent field of embodied cognition asses the crucial role of the body for the construction of the meaning of our own-world [21, 24]. As the body is our only mean to understand the world, we use it at the very beginning of our life and discover some regularities between our actions and our perceptions. These regularities provide the basis of our representations and sense making. It is also suggested by [12, 15] that this principle can be extended to social cognition. Neuroscientists confirm the link between the body capabilities and social cognition [18]. Research studies in psychology and computer science have shown that, in human-human interaction, people continuously adapt and influence each others [25, 3, 9, 7]. In fact, our first interactions with other people (often our parents) induce sensorimotor phenomena and mutual influences that contribute to the feeling of their presence. Our hypothesis is that, as these phenomena are rooted in our memory, they are very important to be produced during an interaction with a virtual character in order to afford the feeling of its presence. For that, we propose here a whole platform for full body human-virtual agent interaction. An important goal of the INGREDIBLE project[1] is that the resulting platform will be provided to the theatrical company[2] we collaborate with, so that the artists will use it in artistic performances of mixed realities. From the artist's point of view, their real world is expanded through the agent's virtual one and the performance is the result of their interaction. To create the illusion of a natural interaction the virtual character must be capable of maintaining a gestural coupling with the human.

Next section describes the current context and what differentiates our proposition from previous similar works. A general description of the platform and some of its modules are explained in the third section. The fourth section illustrates


MOCO'16, July 05 - 07, 2016, Thessaloniki, GA, Greece
ISBN/978-1-4503-4307-7/16/07
http://dx.doi.org/10.1145/2948910.2948943


[1] `www.ingredible.fr`
[2] Web site of the theatrical company: `www.derezo.com`

some examples of its application in different situations of bodily interaction. The final section describes some of our perspectives.

## CONTEXT

In the field of virtual agents, some propositions can be considered close to ours. In [17], Kopp and colleagues work on the notion of *resonance* (from motor resonance to social resonance). Social resonance gathers different phenomena which produce a connection between two people. It is obtained through embodied coordination and mutual adaptation and produces some interactive patterns like mimicry, imitation or synchrony. To favor these phenomena, they use virtual agents who compare a model of their possible actions and gestures with the action of a human [22]. The model is able to interpret the behavior of a human in front of the agent in term of the agent capabilities. The authors argue that this can favor the expectancies of the agent and then the resonances between it and the human. Gratch et al. [14] propose that virtual humans serve as tools for cognitive science. Their toolkit includes a generator of nonverbal behavior based on social psychology [19]. Behaviors are observed during different interactions between humans and are used as a basis to define a rule based system. It generates a relevant nonverbal behaviors according to the sentence that the agent must produce and its emotional state. In contrast to these previous approaches, our work is less general in the sense that it address only gestural interactions. Of course, we agree on the fact that, on the long-term, it is necessary to mix the high level with the low-level part of the behavior but for the moment, we prefer to focus our effort on the body to better fit with the embodied cognition field. Moreover, our approach makes original propositions with respect to the feeling of the co-presence of a virtual character. It addresses a full body interaction in real time between a human and a virtual character and proposes a decision model that is based on the notion of sensorimotor coupling integrated with a dynamical system inspired by the aliveness metaphor of communication proposed by [9]. Coupling is defined as the continuous mutual influence between two individuals [2]. It possesses the capability to resist to disturbances, and compensates them by making the interaction evolve. Disturbances come both from the environment and from within the individuals. This definition is recursive since coupling exists because of the interactants' effort to "recover" it as its quality decreases. In [2] this effort is considered as an explanation of the improvement of the feeling of co-presence, engagement and believability. Interactants measure coupling unconsciously and unconsciously they regulate it. The aliveness metaphor considers that an interaction is like a living system and possesses some mechanisms of regulation. These mechanisms guarantee that if the coupling between two interactants decreases, they will make an effort to recover it. The result is a dynamical evolution of the interaction such that the reaction of the virtual agent to a gesture of the human will change time after time, according to the measurement and the variation of their mutual coupling. This decision model can make appear a regulated development of the interaction that is never exactly the same (as we can hardly obtain with rules based systems). We consider these principles important to favor the feeling of co-presence. Moreover, the virtual agent is able to react to unexpected behavior of the human. Indeed, even if this behavior is *nonsense* for a given scenario, it is possible to evaluate the coupling between the body of the human and the body of the agent, and then, to propose a response of the virtual agent which aims at increasing the level of coupling. Then, the interaction is not *broken* by a sudden dysfunction of the virtual agent. Figure 1 illustrates the main difference between rules based approaches and our approach.

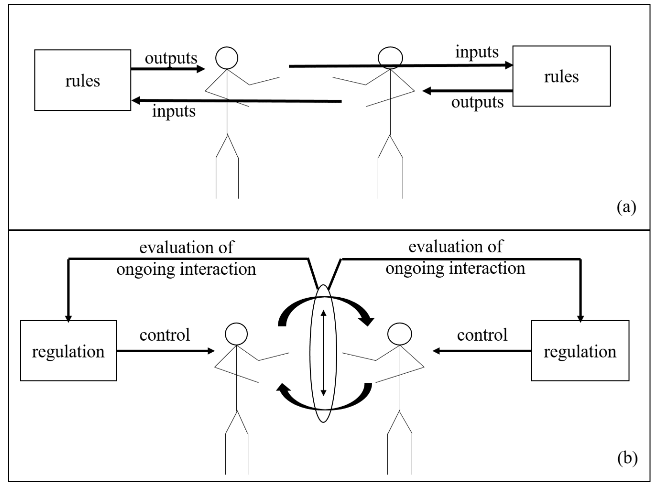


**Figure 1. a) input/output rules based approaches: the behavior of the virtual agent is described in term of human actions that are considered as input. b) Our approach: the behavior of the agent is a regulation of the coupling between the human and the agent itself.**

## THE INGREDIBLE PLATFORM

The INGREDIBLE Platform, based on a modular architecture, is composed of several modules which offer different functionalities. The modules are: *capture*, *analysis*, *decision*, *synthesis* and *rendering* which are linked together through a *communication framework*. Each element of the platform is described in the following subsections.

### Communication framework

Our platform relies on a communication framework based on the serializing structured data protocol buffer protobuf[3] This protocol is language and platform neutral and optimized to accelerate the communication of messages. This real-time aspect is important when numerous modules interact together, including motion capture or animation data. Our framework provides API for `C#`, `C++`, `JAVA` and `PYTHON`. When a new module is created, the first thing to do is to define the modules it will exchange data with and then some predefined callback functions, dedicated to the treatment of a received message, must be implemented. The IP address of each modules and the information about which modules communicates with each other are not hard coded, but they are collected in a configuration file. The messages that the modules can exchange must respect specific formats. Ten predefined types of message can be

[3] `https://developers.google.com/protocol-buffers/`

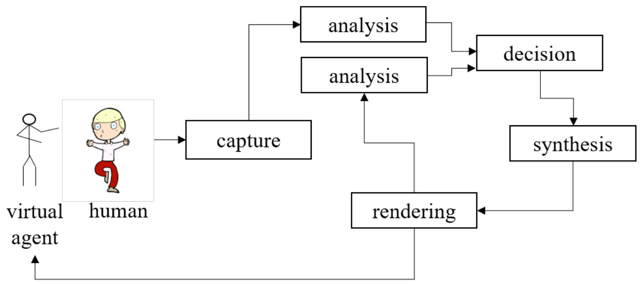


**Figure 2. Configuration of the framework for an interaction between a human and a virtual agent.**

shared in our platform. Skeleton, Bone, Rotation and Position are relative to the bodies (of the human as well as the virtual agent); SkeletonDetected and SkeletonLost provide information about the input devices state; GestureRecognition and Features contain important data used to evaluate the coupling; Decision and GestureControl are used to inform the module, which take care of the agent behavior synthesis, about what the agent should do. Finally Generic messages are used to transmit freshly introduced information which were not envisaged from the beginning. For example, we use generic messages to test different theoretical propositions and we transform these messages into a specific format when we consider that the proposition is mature. This modular approach allows us to configure different setups with different capture devices or rendering engines.

Figure 2 illustrates a configuration dedicated to the interaction between an agent and a human while figure 3 illustrates a configuration dedicated to the interaction between two virtual agents. A part from the capture module, which receive input data (see next section), all other modules ignore whether data come from a real human, a motion capture file or a virtual agent, for such a reason it is easy to combine them for different configurations. For example, we can develop a decision module able to control a virtual agent that interact with a human and then use it to make the agent interact with another virtual character. In this case both agents are controlled by their own copy of the decision module (this is the case shown in figure 3. The reader should notice that in both configurations the decision module of an agent takes as input the data which come from both analysis modules, that is the one which analyses the interacting (real or virtual) partner's behavior and the one that analyses the agent behavior itself. This is one of the originality of our platform: to determine what the agent should do, the decision module needs to evaluate the behavioral coupling (see section **Decision module**). More information on this framework is available on the web site *anonymous* [4].

### Capture module

The capture module carries out two important tasks:

[4]For the moment, the API is available and, if interested, the reader can send an email to the authors to access to the source code.

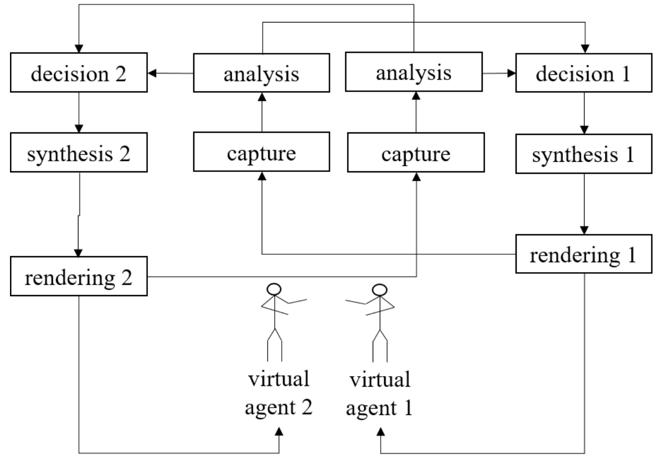


**Figure 3. Configuration of the platform for an interaction between two virtual agents.**

- It creates the connection with the input device or with the rendering module to receive the tracked skeleton. We consider a rendering module as a device like another; this enables us to replace a human with a virtual agent.
- It transforms this skeleton in the specific format used in our platform. In this way, when a new input device must be connected, the only effort required is in the capture module; all the other modules in the platform are not concerned since they will always receive the same type of data.

The internal skeleton format possesses 15 joints, this is usually a subset of the joints provided by modern capture devices. Our choice comes from the fact that we wanted to limit the quantity of data circulating on the communication framework and that 15 joints are enough for computing gesture recognition, motion quality, and coupling between the two interactants. Currently, five capture modules are available, one for each input device we can connect. Thus, our platform can receive data from Microsoft Kinect[5], Moven suit[6], OptiTrack[7], ARTTrack[8] and also our rendering module.

### Analysis module

The analysis module works in real-time. It takes a skeleton as input from the capture module and computes two types of information: 1) a set of motion features extracted from the temporal succession of the posture of the skeleton, 2) the current recognized gesture or movement. Features, which are provided for each part of the skeleton (for instance for the left hand, for the right foot, etc...), are real values extracted from the movement and represent the expressivity of the analyzed behavior, that is the manner in which the behavior (gesture or movement) is performed. Overall, 233 features are computed and among them there are, for example, the speed, the density, the kinetic energy, the fluidity, etcetera. The formulae used to compute them come from the literature, like [13] and

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

[23]. These features are used by the gesture recognition process, and also by the decision module to evaluate the coupling between two skeletons, as explained in the next session.

The gesture recognition process is divided in two steps, one executed off-line and one on-line. First of all the set of gestures we want the agent able to recognize, is collected with a motion capture device.

From the collected set of gestures, the off-line process creates what we call the database of reference gestures. A sliding window of 60 frames, with a shift of 30 frames, cuts each repetition of a gesture in chunks. For each chunk, motion features are computed and their mean, minimum, maximum, and standard deviation are calculated and stocked as a line of the database, together with the corresponding gesture class. To determine the minimal subset of discriminant features, a Principal Component Analysis is computed. Once the database of reference gestures and the discriminant features are determined, the on-line process can run during the interaction to recognize the (real or virtual) human's gestures and expressivity in realtime. The recognition algorithm computes and stores features in a history (that is an array of 60 elements). This history represents the sliding window in which all the discriminant features are computed. The sliding window shift is one frame, that means that the discriminant features are computed at each frame, which enables the system to recognize gestures continuously. More details and an evaluation of the whole process can be found in [16].

This algorithm has the characteristic to work on a continuous flow of data. No needs to indicate a start or a stop, it is able to guess a gesture before its end. Indeed, the evaluation tests that we conducted show that a gesture is recognized during its stroke [16]. This capability to recognize a gesture before its end is very helpful when implementing a virtual agent that can reactively regulate its behavior according to that performed by the (human or virtual) partner. It shows an accuracy recognition of $96,18\%$ on a ten gestures database (5 coming from our capture system, 5 coming from the MSRC-12 Kinect gesture data set from Microsoft [10]). Another important property of our algorithm is its capability to be morphology-independent, since it works on skeletons which are normalized beforehand. Moreover, to be trained, it needs very few repetitions of a gesture to reach a good recognition rate. Typically, 5 repetitions of one gesture, performed with different motion qualities, are enough to make the system recognize it quite accurately. While running, during an interaction between the agent and a (real or virtual) human, this algorithm tries to recognize the partner's gestures continuously and the analysis module sends, in a `GestureRecognition` message, the most probable recognized gesture, as well as some of the motion features (in a `Feature` message).

**Decision module**

The decision module relies on a generic model of the alive communication principle [9]. This principle is inspired by the communication between children and mothers and considers that communication is an evolving process that includes three modes: *co-regulation*, *ordinary variability* and *innovation*. During *co-regulation*, people continuously adjust their behavior to that showed by others through postures or gestures. They do that to maintain a connection, they implicitly say *we are in-line*. It is for example the reason for using behaviors such as mimicry and backchanneling. This co-regulation is never exactly the same because there is an intrinsic variability in human behavior, however it never exceeds some socially admitted rules. The *ordinary variability* arises after a certain amount of time of co-regulation. Participants dare to introduce some subtle but noticeable changes which suggest an adapted reaction. This variability can change progressively the goal, the shape or the intent of the interaction. The *innovation* is an extraordinary variability. It is a strong proposition that suggests generally a new type of behavior. In the long term, an innovation could be accepted by both interactants and become an ordinary variability. This succession of modes produces a developmental evolution of the interaction. People feel variability and innovation and they adapt their behavior according to them. This adaptation is perceived by each interactant and we assume that it favors the feeling of co-presence.

The decision module is generic in the sense that the key mechanisms of the alive communication metaphor, applicable to any type of interaction, are implemented. However, according to the context, every interaction is different and could need different information. That is, the decision module cannot be used as it is for any type of interaction, some information relative to the context must be specified. As we will see, this information, which is how coupling is computed, what gestures can be used and how they can be modified, depends on the context. For example, in the artistic imitation game, where two people have to imitate each other movements, coupling is computed as the correlation of their respective joints position, while in an exergame, where the agent plays the role of a fitness coach and the human is the student, coupling computation should take into account the performed movement and some motion features, such as speed and amplitude.

Figure 4 summarizes the decision module. It takes as input the messages coming from the analysis module that analyses the partner's behavior and the messages coming from another instance of the analysis module that analyses the agent's own behavior. As explained in the previous section, these messages can contain the identifier of the recognized gesture and/or the features describing the behavioral expressivity. The module needs these data from both interactants in order to compute the sensorimotor coupling between them. For a detailed and formalized description of its functioning, one can refer to [8]. In short, a discrete random variable is used to represent the mode of interaction of the participants. The probability to be in each mode depends on the temporal evolution of the coupling between them. The coupling is the sum of the correlation of features and gestures/movements that are weighted according to their relevance in the specific context the interaction takes place in. The probabilities evolve according to the temporal variation of the coupling: If the coupling is strong enough for awhile, the probability to offer a variation increases. If not, it is the probability to innovate that increases. This later increases also after each cycle be-

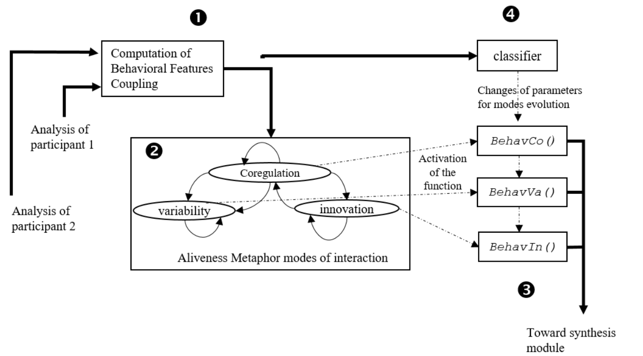


**Figure 4. Generic part of the cognitive module: (1) The coupling between the two interactants is computed, it is used to estimate the mode of the interaction through a discrete random variable (2). (3), Behavioral functions dedicated to each mode of interaction. (4), modulation of the behavioral functions through some modification of their parameters according to the coupling.**

tween co-regulation and variability modes. Any time variability and innovation modes are attained, their respective probability is reset. As a result, the succession of modes of an interaction follows the aliveness metaphor: At the beginning, the interaction is in co-regulation mode. If it progresses well, it is followed by some variations and the interaction enters in variability mode. Such variations can decrease the coupling. However, if the interactants make efforts to recover it, the interaction returns in co-regulation and after awhile, new variations can appear. After some cycles of co-regulation and variability, the interaction can enter in innovation. It also enters in innovation if the coupling remains too low. This later case corresponds to a lack of understanding between the interactants and in such a situation, a good solution is to try an innovation, even a drastic one, such as ending the interaction.

If the other participant is a human, they can make variability or innovation at any time and this change is detected from the current values of the coupling. At the opposite, the model can propose to change the mode of the interaction according to the value of the probabilities. In this case, it is the controlled agent that modify its behavior. For that, three abstract behavioral functions (`BehavCo()`, `BehavVar()` and `BehavIn()`) are performed according to the mode of the interaction (*co-regulation*, *variability* and *innovation*). In practice, these functions must be defined and describe the behavior the virtual character should perform when the interaction is in the mode that they represent.

The agent behavior computed by these functions could be different according to which interactant causes a change of mode. For example if it is the human who offers an innovation, the agent can act differently from the way it would have done if the innovation came from the agent itself. This information, that is the initiator of the mode change, is computed by the decision module: the initiator is the interactant whose relevant features have evolved the most in the last time window.

Another important characteristic of the three behavioral functions is that they are parameterizable. We want that when the agent has to offer a behavior variation or an innovation, the resulting animation will not always be exactly the same. For example if the agent enters in variability mode and decides to vary its speed, the amount of such variation will not be always exactly the same. This amount is a parameter to the *BehavVar()* function and its value varies during the interaction according to a set of rules defined in a system classifier. Remember that here we focus on the sensorimotor aspect of interactions, so the classifier can contain rules that modify the agent behavioral expressivity according to the level of coupling. For example, a rule could say that if the coupling is higher than 80%, the speed can increase randomly in a range from 0 to 5% of the initial speed. At the opposite, another rule can specifies that if the coupling is lower than 80%, the speed of the virtual agent varies to get closer to the human's speed. This will exhibit a co-regulation behavior. There is no limit to the number of rules that the classifier could contain. The creation and the manipulation of these rules are relatively easy because the rules are written together in an `XML` file with dedicated tags for conditions and actions. By default, the system chooses randomly a rule among those which have their condition equal to true. However, it is possible to add a weight to the rules, which is taken into account by the random choice process. Rules with higher weights are obviously favored.

The messages that the decision module can send are written in the `Decision` format. This format allows to specify the identifier of the gesture the agent has to perform, the expressivity that the agent has to show, and a set of key positions that some skeleton joints have to reach. At present, just the values of speed and amplitude are actually sent by the decision module, other features could be easily added, but for the moment our rendering can manage only variations in speed and amplitude.

### Synthesis and Rendering module

Since in this research work our main concerns were the gestures recognition, the motion features analysis and the decision algorithm, we did not concentrate on the synthesis and the rendering and no important research progress have been done on this side. Currently, we have a unique module for synthesis and rendering which has been created using the game engine Unity3D[9]. This engine provides a virtual scene where an agent can be displayed. The agent can play motion captured animations, but it can also be animated using inverse kinematics. For this purpose, Root-Motion provides a library called FinalIK[10], that one can import in Unity. So, the synthesis and rendering module can play a specific gesture, with variations in expressivity such as speed and amplitude (thanks to FinalIK), or it can animate the agent according to received key-frames. These key-frames contain the position to attain for a sub-set of the skeleton joints. Again it is the FinalIK library which takes care of the computation of the animation. This library is a useful and fast solution, the drawback is that it is not possible to use specific algorithms like, for example, data low-dimensional parameterization models [6, 20, 4].

[9] **https://unity3d.com**

[10] **http://www.root-motion.com/final-ik.html**

The synthesis and rendering module can receive as input the identifier of a gesture to play, some values that specify the expressivity the agent should show and some key-frames. Any other synthesis and/or rendering tool able to receive and use this type of data could be connected to our platform.

## APPLICATION EXAMPLE

We uses our platform in a fitness exergame scenario. The goal of this context of interaction consists in making the virtual agent capable of performing fitness exercises with a human or another agent. The interaction can take place in three different situations: 1) the agent plays the role of the coach and its partner is the student who has to reproduce the movements proposed by the coach; 2) the agent is the student and its partner plays the role of the fitness coach; 3) there is neither coach nor student, both interactants perform fitness exercises together, both can offer a movement or follow that offered by the other, like two companions. The agent has a limited number of movements that it can perform and recognize: Twelve fitness movements were captured through a motion capture system. They are used by both the analysis module, to enables the agent to recognize them, and the synthesis/rendering module, to play them on the virtual character. At the beginning of the interaction, independently of the situation the interaction takes place in, any movement offered for the first time (by the coach or by the companion) is a novelty (that is an innovation). We call it an "unknown movement" in the sense that it is not known by both the interactants. As soon as a movement is performed by both participants, it becomes a "known movement" and it is not considered as a innovation anymore. Offering an already known movement is considered as a variability. The variability mode is attained also when an important change in the manner the movements are performed is offered. For example when an interactant accelerates.

Since fitness practice relies on specific body movements and on the manner they are performed, to compute the coupling we consider as relevant features the speed and the amplitude of the fitness movements, as well as the identifier of the movement performed by both interactants. Thus, the more similar are the performed movements (in shape and expressivity), the better is the coupling. To distinguish the three situations, we have to specify for each of them the three behavioral functions: `BehavCo()`, `BehavVar()`, and `BehavIn()`. The (`BehavCo()`) function is the easiest to define since it is the same for all situations: when the interaction is in co-regulation mode, the virtual agent modulates its speed and amplitude to be as similar as possible to those showed by the human.

Whenever the interaction enters in variability mode, the `BehavVar()` depends on both the situation and the interactant who has caused the change, that is, the initiator.

- *Situation 1, the agent is the coach.* When the agent is the initiator, it can offer another known fitness movement, or modify its behavioral expressivity by changing speed or amplitude. As explained in section **Decision module**, the amount of the expressivity change is not a fixed value but is a parameter which can vary during the interaction according to the classifier system. Anyway, since the agent is the coach, it has the tendency to increase speed and amplitude, in order to push the student to make efforts. On the other hand, if it is the student who causes the interaction to enter in variability mode, the agent coach reacts differently according to what the partner has offered. If the student has varied the expressivity, the agent adapts its own manner of performing the movement; if the student has offered a known movement, the agent shows a gesture of refusal, it is the coach after all!
- *Situation 2, the agent is the student.* The agent can cause the change in the interaction mode only by modifying its speed or amplitude, it cannot offer another movement since it is just the student. On the other hand, if the coach is the initiator, then the agent student adapts its own behavioral expressivity to that shown by the partner and, if it can recognize the offered known fitness movement, the agent performs it. Otherwise it shakes its head to show that it has not understood.
- *Situation 3, the agent is a fitness companion.* When the agent is the initiator, it can offer another known fitness movement, or modify its behavioral expressivity by changing speed or amplitude. If it is its partner who causes the interaction to enter in variability mode, then the agent adapts its behavior to follow their lead. It shakes its head only if it cannot recognize the movement offered by the companion.

Similarly, Whenever the interaction enters in innovation mode, the `BehavIn()` function depends, too, on both the situation and the initiator.

- *Situation 1, the agent is the coach.* The agent can cause the change in the interaction mode by offering a new unknown fitness movement or by performing a bow gesture to signal that the sport session is over. On the other hand, if it is the student who causes the interaction to enter in innovation mode, the agent coach reacts differently according to what the partner has offered. If the student tries to offer an unknown movement, the agent shows a gesture of refusal. If the student bows to stop the interaction, the agent respects their choice and bows back to conclude the sport session.
- *Situation 2, the agent is the student.* The agent can cause the change in the interaction mode only by bowing to stop the fitness session. It is interesting to notice that, according to the decision module, the agent becomes the initiator of an innovation after several cycles between co-regulation and variability modes, that means that the coach is performing always the same known movements and that, as a consequence, the agent student is getting bored. That is why it decides to stop the session.

  In this situation, also the coach can be the initiator of an innovation. In this case the coach can offer an unknown movement or perform a bow to stop the interaction. In the first case, the agent starts performing the same movement (if it can recognize it) or shakes its head (if it cannot recognize the movement). In the second case, the agent bows back.

- *Situation 3, the agent is a fitness companion.* When the two interactants are just two companions who practice sport together, they can both cause the interaction to move in innovation mode by proposing an unknown fitness movement or by performing a bow to stop the sport session. When it is the partner who causes such a change in the interaction, the agent just follows the lead.

To complete the fitness exergame scenario, we define four basic classifier rules: one specifies that if the coupling is higher than a threshold for awhile, the range of the possible values for the random change of speed increases. Another does the same but for the amplitude. Reciprocally, two rules indicate that if the coupling is lower than a threshold, these two ranges decrease. Of course, numerous other rules could be imagined. The purpose here is to illustrate the possibility to introduce some evolution of the behavior during the interaction.

Figure 5 shows the interaction set-up. The agent appears on a big TV screen. It has a humanoid shape which allows us to avoid all facial expressions and to make all animation flaws less visible. A video associated to this paper provides an example of a whole interaction. This video shows parts of two interactions which took place in situation 3, that is when the agent and the girls were fitness companions.

It is important to highlight that our platform can be used with other scenarii. Part of the platform is generic but some elements are obviously strictly dependent on the type of interaction we want to obtain. To sum up, for each new scenario, one has to determine which features are relevant to compute the sensorimotor coupling, the rules of the classifier, and the three behavioral functions. This last point requires a minimal programming knowledge. Instead, the rules of the classifier system are quite easy to write, since they are collected in an file written in `XML`.

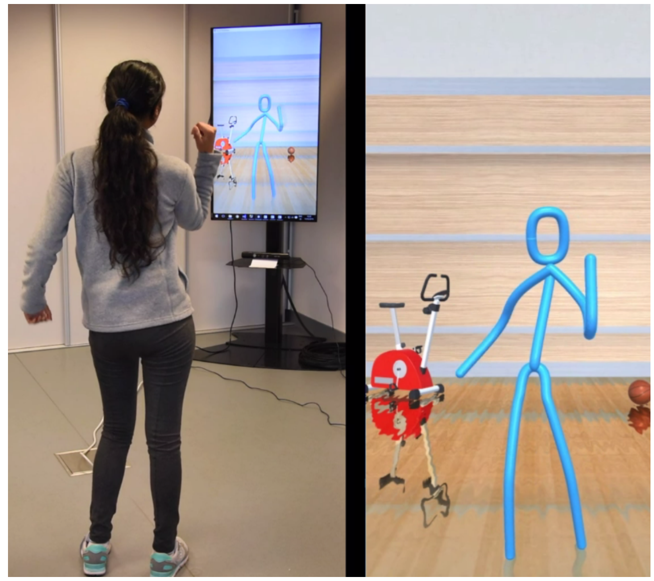

**Figure 5. Snapshot of the video that illustrates the example of fitness exergame. On the left, the full setup, the human is tracked by a kinect. On the right a zoom on the virtual agent. As we want to investigate the role of the full body in an interaction, we choose an agent that has a vague humanoid shape.**

## CONCLUSION AND PERSPECTIVES

In this paper, we have presented an approach for the development of virtual agents able to perform full body interactions with a human or another virtual character. We have shown its full implementation in an interactive platform which we aim to provide to a theatrical company for artistic performances purpose. The platform relies on a communication framework and a set of modules that takes care of different functionalities, such as the analysis of the human's (or the other agent) behavior, the rendering of the agent non verbal behavior, and the decision computation which follows the alive communication metaphor. This metaphor emphasizes the role of the regulation of the coupling between people during natural interaction. Our aim is to integrate such a metaphor in our virtual agent platform in order to improve the human's feeling of co-presence and the agent behavior believability. An evaluation of this assumption is underway: Subjects have to interact with an agent that plays once the role of the coach and once that of the student. We plane to make a comparison between objective and subjective measures, as it is made by [26] to asses the acceptability of a human robot interaction. The platform is free for testing and research purposes, the reader has just to ask the authors for access. All contributions or new modules will be very welcomed. Currently, another specialization of the generic decision module is underway. It will be used by the theatrical company we collaborate with, to prepare an artistic performance. During this performance, a real artist interact with the virtual agent. The aim of this interaction is to play the theatrical imitation game: Two players facing each other imitate the other person's movements but, from time to time, each player can introduce subtle changes by proposing variation in the hands and feet direction. This dyadic imitation game causes dynamic notions of coupling and interaction to emerge naturally from both players. No motion captured gestures can be used because it is not possible to define in advance the evolution of the posture that the participants will propose. In such a scenario, coupling will be computed using as relevant features the 3D positions of hands and feet. We plane also to use this artistic scenario in the platform configuration showed in figure 3. Our goal is to study the emergency of evolving interactive patterns between two agents controlled by two decision modules parameterized with different values. We hope that the two interactive agents will give the illusion of being autonomous and *in line* with each other. A module dedicated to the connection with artistic devices is also under development. It will allow to convert Open Sound Control [11] messages in a format of the communication framework (and vice versa) and then to inject in the platform other information that can be considered as expressive (music, sound). The goal is to increase the possible artistic scenarii in which the alive communication metaphor can still be applied.

## ACKNOWLEDGMENTS

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

## REFERENCES


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

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

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

4. P. Carreno-Medrano and S. Gibet. End-Effectors Trajectories: an Efficient Low-Dimensional Characterization of Affective-Expressive Body Motions. In *International Conference on Affective Computing and Intelligent Interaction*, pages 435–441, 2015.

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

6. J. Chai and J. K. Hodgins. Performance animation from low-dimensional control signals. *ACM Transactions on Graphics*, 24(3):686–696, 2005.

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

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, CASA, to appear*, May 2016.

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

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

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

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

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

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

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

16. C. Jost, I. Stankovic, P. D. Loor, A. Nédélec, and E. Bevacqua. Real-time gesture recognition based on motion quality analysis. *ICST Trans. e-Education e-Learning*, 2(8):e5, 2015.

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

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

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

20. H. Liu, X. Wei, J. Chai, I. Ha, and T. Rhee. Realtime human motion control with a small number of inertial sensors. *Symposium on Interactive 3D Graphics and Games on - I3D '11*, pages 133–140, 2011.

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

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

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

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

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

26. V. Weistroffer, A. Paljic, L. Callebert, and P. Fuchs. A methodology to assess the acceptability of human-robot collaboration using virtual reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, pages 39–48, 2013.