Directing Autonomous Digital Actors

21 avril 2015

Table des matières

1 Résume de la proposition de projet / Executive summary

2 Introduction et positionnement / Introduction and positioning

2.1 Contexte et enjeux économiques et sociétaux / Context, social and economic issues

Animating virtual agents with expressivity is a grand challenge for the entertainment industries (video games, movie industry) that rely mainly on motion capture data which allows them to produce rich and subtle motion but with a high cost in time and finance. On the other hand, technology for interactive agents uses mainly procedural approach. While such approach allows modulating in real-time agents' motion and its quality, the results are still far from being natural and realistic. Lately statistical approaches have been developed. They are promising as they produce animations captur-ing naturalness and richness of human motion. However the control of such animation technique is still an issue and its extension to a large range of motion activities is also an important challenge.

DADA aims to bridge the gap between those previous techniques by proposing a general framework for combining them into a unified interface. A desirable outcome of the project will be a completely novel interaction model for rehearsing with virtual actors and incrementally building complex multi-actor performances with multiple layers of 3D animation. Thus DADA fits the component Information and Communication Society ; it is also in line with at least two axes of the ANR call.

First axis : Le numérique au service des arts, du patrimoine, des industries culturelles et éditoriales (3.7.1.3). There are several potential applications of DADA on top of the proposed virtual theater. The creation of expressive virtual characters can be used in video games, especially for NPC (non-player characters), and in serious games. Indeed being able to simulate the motion of a virtual actor with different expressivities and for different morphologies while maintaining a high level of naturalness and lifelikeness will be a big benefit in time and money.

Second axis : Interactions des mondes physiques, de l'humain et du monde numérique (3.7.2.4). The outcome of DADA will benefit the creation of virtual agents, either autonomous or controlled by humans. These agents ought to display a large variety of communicative and emotional expressions toward human interactants as well as to perform many actions with objects in their virtual environment. Enhancing, in quantity and expressivity, the behaviors of virtual agents is one of the challenges of DADA that falls under the axis.

Related projects : There exist several large European projects that are related to DADA research themes. However, to our knowledge, none covers our research question of building expressive animation with different levels of control. We can name the NoE IRIS on story-telling, the IP Companions on dialog virtual agents, the IP REVERIE on modeling virtual characters in highly immersive virtual environment, and the STREP Ilhaire aims to simulate laughing agent using data-driven and motion graph approaches. On the National side, we can name the Feder project Anipev in which the database Emilya has been captured.

Several recent projects have been devoted to the interface between computers and theatre, e.g. ANR VIRAGE (a generic architecture for controlling lighting and music during theatre production), ANR OSSIA (authoring tools for writing interactive, multimedia scenarios) and ANR INEDIT (INteractivité dans l'Ecriture De l'Interaction et du Temps). ANR Spectacle-en-ligne(s) was a SHS CORPUS project dedicated to capturing, indexing and annotating 200 hours of theatre rehearsals recorded in high-definition video. Those projects focus on the interaction of computer systems with real actors. In contrast, DADA will focus on the core issue of directing virtual actors.

Despite considerable academic research, few procedural animation systems have become commercially available in recent years. Euphoria by Natural Motion is a real-time procedural animation engine, which has been used in Grand Theft Auto 4 and other games. However, actions and expressions are difficult to control. Xtranormal Technologies was an online service for quickly creating 3-D animations from dialogues decorated with stage directions, using a proprietary procedural animation engine limited to non-expressive behaviors. Actor Machines is a company created by Ken Perlin to commercialize packages of trained virtual actors with a large range of actions and expressions, which has not delivered any product yet.

2.2 Positionnement du projet / Position of the project

Creating believable, human-like performances by virtual actors is an important problem in many digital storytelling applications, e.g. creating non-player characters (NPC) for video games, creating expressive avatars in next-generation virtual worlds, populating movies and architectural simulations with background characters and crowds, creating be-lievable virtual tutors and coaches in educational serious games, and creating believable characters for interactive fiction and interactive drama (Tannenbaum 2014).

A desirable feature for such applications is the ability to create virtual actor performances which are both expressive and controllable. Motion capture actors are expressive, but once recorded, their performances cannot easily be controlled, edited or modified. As a result, game companies ought to get engaged in extensive motion capture sessions of all actions and moods of all characters in every new game they create. On the other end of the spectrum, procedural 3D animation can be controlled in every detail using sophisticated programming techniques, but they fall short of providing the level of expression required for conveying the subtle inflexions of human-like performances.

Character animation has been tackled through various approaches in the past. To name a few, chosen among those that are directly related to DADA, we can cite : embodied conversational agents (ECA), ie autonomous virtual characters [14] ; statistical models learned from motion capture examples [58] ; physically-based animation [70] ; and speech-driven animation [23]. Very few attempts have tried to merge these various approaches into a single model offering on one hand expressive animation and on the other hand high control over the animation.

In order to make progress in the field, we propose to shift the focus from autonomous characters to autonomous actors. Autonomous characters (such as The Sims) make decisions based on AI models of their personality and goals. In contrast, autonomous actors follow a precise script, written by the director. Their autonomy is therefore limited to perform-ing a precise sequence of actions as a result of various cues written in the script. Creating such performances procedurally using autonomous actors is a valuable goal because it would make it possible for each performance to be unique, which is widely regarded as an important quality to ensure liveliness and immersion, while maintaining a high level of directorial control. Merging both approaches would allow creating autonomous actors able to follow a script (specified in high-level command-like language) that give the main directions the actors ought to follow while adapting their behaviors autonomously to the virtual environment they are placed in that includes objects and other actors.

The goal of the DADA project is to design, implement and evaluate novel interfaces for directing expressive, autonomous virtual actors, borrowing from established theatre practices. We will combine fundamental re-search in 3D animation, machine learning and intelligent agent programming to leverage motion capture data sets of professional actors into a virtual theatre company of synthetic actors with acting skills, i.e. ability to respond to a director's instructions and to perform together on a virtual stage. Virtual theatre will be used as a test application for obvious extensions to other digital storytelling applications.

To reach this ambitious goal, DADA will learn parameterized models of actor's movements and gestures from existing annotated motion capture databases of actor performances ; and create intuitive authoring tools for creating a script of actions and cues in a machine-readable format suitable to real-time control of the virtual actors. More precisely, the academic partners of the project will engage fundamental research along two main directions :

1. Animating autonomous actors procedurally. A key idea in DADA is to separate the animation model into a proxe-mic component regulating how actors interact with each other and the audience, and a kinesic component regulating how actors use their body language to communicate moods and expressions [101]. The proxemic component of animation will drive the positions and orientations of actors on the stage as well as their gaze directions. This component will be driven by a model encompassing the social relations between and the emotional states of the autonomous actors. The kinesic component of animation will drive all other degrees of freedom of the virtual actors. This component will be driven by parametric statistical models trained from an existing motion capture data-set. The separation between the two components is expected to yield important benefits in terms of expressivity and composability.
2. Synchronizing virtual actors to a single story-line using a story-driven architecture of actors following a scripted sequence of instructions [84]. In contrast to previous works, which used programming languages [76], we will investigate multimodal interfaces offering directorial control in a high-level, pseudo-natural language familiar to the director. The language will be compiled internally to a finite-state machine representation controlling the real-time execution of the autonomous actors.

2.3 Etat de l'art / State of the art

A large body of theoretical research work relates to acting and directing in the theatre, especially from a cognitive science perspective [51]. Its applicability to virtual actors is limited because of the huge gap between virtual and human acting skills. An extensive survey of acting techniques used in 3D animation and virtual worlds can be found in the excellent collection edited by Tanenbaum, et al. [101]. The priomises and limitations of virtual actors in contemporary theatre has been surveyed by Dixon [29], Salter [94] and Bourassa and Poissant [**?**].

2.3.1 Single-character animation

Animation of an avatar is usually tackled by working separately on the full body animation model on the one hand and on the face (and gesture) animation model on the other hand (since the latter animation strongly depends on the dialogue the avatar is engaged in), where the animation produced by the two models are merged to produce a final complete animation [98].

Full body kinematic animation (or control) consists in animating the full body of an avatar while he is performing actions such as walking, dancing, sitting etc. Although there has been lots of work on this subject it is still a challenging problem due to the high dimensionality of the character's configuration. Data-driven approaches are very popular here and make of use motion-capture data to learn animation models which, once learned, may be used to animate a virtual character to perform a given task. Many systems have been proposed for producing animation models and controlers, they usually are based on statistical models such as Hidden Markov Models (HMMs) [64] and Conditional Random Fields (CRFs) [62, 17]. Most accurate methods exploit a large dataset of motions where one can synthesize a complete motion sequence corresponding to a particular task by using warping or blending strategies of motions in the training set [113]. Locomotion controllers have been proposed that concatenate motion clips from a motion capture dataset to produce an animation that is smooth [104, 77]. High-quality kinematic controllers have been built from this idea by using a *motion graph*, which is a graph structure that describe how clips from a dataset can be reordered into new motions [59].

While locomotion controllers are driven by direct high-level commands (such as desired movement direction), no such clear control signal is available for body language. To animate the face, and accompanying arm gestures, many works have focused on developing specific animation models based on a dialogue related input, either speech, text or prosody features [64, 62, 17, 24]. At the end, recent work has demonstrated such models for the case of locomotion believable controllers, gesture controllers [62] and face controllers [28].

Yet all these statistical approaches require large annotated datasets to work well.

Thereby these approaches do not easily work with small training sets which is a key issue, as stressed for instance in [65], since first it requires considerable effort and time to build large datasets, and second because many applications demand unique motion styles and require their own datasets. This has led a number of researchers to put the effort on designing models that may be easily learned from a few samples. One main approach for doing so lies in the use (or learning) of a continuous state space to represent the data, making learning in this low dimensional space much easier [65, 17]. A relevant technology for this are Gaussian Process which have been extended for dealing wiuth dynamic data in [109].

These latter models are not far from recurrent neural networks, and to Long Short Term Memory neural networks in particular [48, 41], that have been shown recently to work well for complex signals such as speech and handwriting, for recognition tasks [40] as well as for synthesis tasks [38]. These models are part of a current trend in machine learning called representation learning (see the recently born conference ICLR at http ://www.iclr.cc/) which aims at discovering relevant and usually low dimensional representation of the data under investigation (the pionner work of this domaine is the one by G. Hinton in Science [47].

2.3.2 Multiple-character animation

ECAs gathering in groups : Prada and Paiva [86] modeled groups of autonomous synthetic virtual agents that collaborated with the user in the resolution of collaborative tasks within a 3D virtual environment. Rehm and Endrass [90] implemented a toolbox for modeling the behavior of multi-agent systems. In [89], Ravenet and colleagues combined a number of reactive social behaviors, including those reflecting personal space [44] and the F-formation system [52], in a general steering framework inspired by [91]. This complete management of position and orientation is the foundation of the Anonymous Engine used in the work presented here. All these models did not take into account the expression of attitudes while exhibiting the behaviors of the agents.

Forward Kinematics (FK) and Inverse Kinematics (IK) : FK and IK [10, 105] have been used to generate animation sequence procedurally. Different constraints and unconstrained numerical equation solvers are used to model the motion postures and their transitions [115]. The target based reaching models, such as Jacobian methods [114] [13], are proposed to use linear approximation to bring the end-effector close to the target ; it can be used to build a posture configuration for virtual characters. These methods are flexible and widely used for real time lost cost procedural generation and motion editing, but the resulting motions may not be as realistic as human motion. They can show quite some stiffness.

Data-driven animation : Another animation technique that is being used more and more is motion capture data. Many animation techniques based on concatenation of motion capture data and clips selection [60, 56, 45] have been proposed. These methods focus on how to improve the clips searching performance or the transition quality between successive clips. These methods have been applied very successfully to locomotion synthesis. Feng et al. [31] have extended motion example method for gesture synthesis. Since they use directly motion capture, their synthesized result is quite natural, but do suffer from some limitations : new data needs to be recorded for any new required animation.

Style Machines [11] achieve the stylistic motion synthesis by learning motion patterns from a highly varied set of motions, as a distinct choreography can perform motions with a distinctive style. Stylistic Hidden Markov models (SHMM) are proposed to train different stylized behaviors. A new style motion can also be obtained from the interpolation of SHMM sub-spaces. This approach treats animations as a pure data modeling task. Later, such types of work have been extended [110] [83], but, so far, the focus of these synthesis works is more on cyclic motion ; cyclic motions are somehow easier to treat for clustering, blending or filling missing data.

Systems embedding different animation techniques : There exist several systems embedded in large platform that combine different animation synthesis solutions. The Smartbody system [96] [102] includes a schedule controller and a realization controller. Smartbody includes different motion synthesis methods such as motion capture data and dynamic system. ETENDRE : DONNER PLUS D INFORMATION. IL FAUT AUSSI EXPLIQUER EN QUOI NOTRE SYSTEME DIFFERE DU LEUR. The behavior engine developed by Marsella et al. [75] makes use of Smartbody. This engine embeds a rule-based system that can determine facial expressions and gestures by extracting semantic, pragmatic and rhetorical content from an input utterance. Some behaviors (eg head motion) are also obtained through data-driven model. Luo et al. [71] proposed a procedural arm gesture model. They improve the quality of the procedural animation by introducing motion capture data for full body animation. ADAPT [97] is also a flexible platform for virtual human characters. This system has been designed as a gaming system for physical reactions. They use a behavior tree to model human-like behaviors for multi-characters and a Level of Detail character shadows solution is proposed to generate body parts sequentially. Blending technique is applied to compute the movements of different body parts from different Choreographers. While some motion artifacts may appear when doing blending or character motion transition, this animation platform combines various animation techniques. However it has not been focused on communicative and emotional behaviors that are defined by specific temporal patterns.

Beyond State of the Art : Both animation techniques, procedural and data-driven, have pros and cons. With the procedural animation, gestures are described using the symbolic language BML [**?**, **?**]. New behaviors can be created very rapidly ; but, most of the time, the obtained animation lacks naturalness and fluidity. On the other hand animation from motion-capture data reproduces all human motion subtleties and dynamics ; however creating new gesture is more cumbersome as it requires new recording of data and new training.

Our aim is to develop an embodied conversational agent system that embeds both animation approaches. Communicative behaviors are computed procedurally while socio-emotional behaviors such as emotions are driven from machine-learning techniques (Task 1.X). Both computations involve the same body parts. While the two animation streams are computed separately they need to be merged to produce the final animation output. This work is part of Task 2.3.

2.3.3 Authoring tools and metaphors

There has been previous work on story-driven architectures and directing tools specifically dedicated for virtual theatre. The desktop theatre, the story-driven architecture and Improv are seminal works in building virtual actors that can be scripted. Other early work includes Pinnochio [73] and GEIST [99] which targets players of interactive games by letting them "play director". Both projects were left unfinished.

Xtranormal State and Notion are based on the paradigm of "text-to-scene" animation. The DRAMA project at the University of Toulouse has also investigated this paradigm. At this point, using natural language processing appears to be out of reach.

MIRAGE [30] is an interactive story generation engine featuring 3D animation of two virtual actors playing the roles of Electra and Archemedis in a tragic style. Actions are represented by a verb, and adverb and an actor ; and are either controled by the player or generated by the system. Actor's behaviours are organized into dramatic beats with multiple actions and reactions based on communicative goals. While the focus of MIRAGE is different from DADA, it offers the important insight that actions performed by the virtual actors must be appraised from the point of view of the user (director, audience or player). This will be used as a general guideline in DADA as well.

The body action and posture coding system (BAP) [19] is an extensive description language for body movement on the anatomical level (body parts), the form level (directions and orientations of movements) and the function level (communicative and goal-driven actions). Their work is important for DADA because it offers a catalogue of expressive gestures used by professional actors in depicting various emotions [20]. The language makes it possible to precisely annotate body actions in video using the ANVIL annotation tool [53, 54].

The movie script markup language (MSML) [107] has been proposed to encode the stucture of movie and play scripts into scenes, actions and dialogues. One interesting feature of the language is that it includes an animation layer, making it possible to compute a realization of the play script in 2D animation. The synchronization of the various actors and actions in the play script is performed by generating an Object Composition Petri Net (OCPN) [69] for the entire scene. This feature has not been demonstrated with 3D animation, and it will be one of the objective of the DADA project to implement, test and validate the MSML animation layer with multiple autonomous 3D actors in a real-time game engine.

Petri nets have also been proposed as a high-level specification for virtual actor behaviours [72, 9], including the important case of turn-taking in dialogue and imitation games [16, 15].

They are also the basis for the interactive musical score system developed by the INEDIT project [6, 74], which is being extended to multimedia events [103]. Petri nets are likely candidates to become an internal, intermediate representation of the dramatic score in DADA, between the high level commands of the director and the low-level executable finite-state machine of the game engine. In previous work, the composer or director manipulates the Petri nets directly. In DADA, we will instead offer direct manipulation of cue sheet and prompt books, which offer more natural interaction for theatre directors than Petri net places and transitions.

The Q language [**?**] is an authoring language for writing scenarios involving multiple autonomous agents. While not targeting theatre as an application, the language borrows heavily from theatre practices and is based on defining cues that synchronize the actions and reactions of the agents. Q is a dialect of the scheme language and assumes that the scenario writer is familiar with programming. The directing language for DADA will also be based on cues and actions, but will be extended to include direct user manipulation using a graphical user interface, rather than a programming interface. Furthermore, animation generated with the Q language was fairly primitive and we plan to offer a more expressive and extensive description of actor's full body animation.

State of the art in France

International state of the art

2.3.4 Autonomous agents and non-player characters (NPC)

Our project is also related to work on NPCs - behavior trees,

2.3.5 Robotic actors

Our project is also related to the emerging field of robotic actors performing theatre on a live stage [67]. An international workshop was dedicated to robotics and theatre at ICRA 2012[1]. Our research on directing autonomous digital actors will likely be applicable to the case of robotic actors as well.

2.3.6 Prior art at Inria

Previous work in the IMAGINE team can be useful to the DADA project, including work on retargeting costumes to different actor body shapes [12] ; representing actor's anatomy with high precision ontologies [**?**] ; sketch-based modeling of actor's movements using motion brushes [78], lines of action [42] and spatio-temporal curves [43] ; steering behaviors coordinated with finite state machines [66] and extended to the case of actors (or cameras) looking in other directions than their target destination [36] ; implicit skinning of actor's body shapes for improved rendering of character animation [106] ; and audio-visual prosody modeling for expressive facial animations of actors [7, 21, 8]. All previous work will be made available to the DADA project.

2.3.7 Software foundation

Our developments will be based on the GRETA platform and use the EMILYA database.

[1] Robotics and Performing Arts : Reciprocal Influences, http ://www.robotics-and-performing-arts.sssup.it/

GRETA The Greta platform simulates virtual agents able to communicate verbally and nonverbally with human users and/or other virtual agents. Given a set of intentions and emotions to be communicated, the platform instantiates them into sequences of synchronized nonverbal behaviours. It can be used to compute these multimodal behaviours when the virtual agent acts as a speaker or as a listener.

The Greta system allows a virtual or physical (e.g. robotic) embodied conversational agent to communicate with a human user [82, 81]. It is a SAIBA compliant architecture (SAIBA is a common framework for the autonomous generation of multimodal communicative behavior in Embodied conversational agents [55]). The main three components are : (1) an *Intent Planner* that produces the communicative intentions and handles the emotional state of the agent ; (2) a *Behavior Planner* that transforms the communicative intents received in input into multimodal signals and (3) a *Behavior Realizer* that produces the movements and rotations for the joints of the ECA.

A *Behavior Lexicon* contains pairs of mappings from communicative intentions to multimodal signals. The Behavior Realizer instantiates the multimodal behaviors, it handles the synchronization with speech and generates the animations for the ECA.

The information exchanged by these components is encoded in specific representation languages defined by SAIBA. The representation of communicative intents is done with the Function Markup Language (FML) [46]. FML describes communicative and expressive functions without any reference to physical behavior, representing in essence what the agent's mind decides. It is meant to provide a semantic description that accounts for the aspects that are relevant and influential in the planning of verbal and nonverbal behavior. Greta uses an FML specification named *FML-APML* and based on the Affective Presentation Markup Language (APML) introduced by [22]. FML-APML tags encode the communicative intentions following the taxonomy defined by [85], where a communicative function corresponds to a pair (*meaning*,*signal*). The meaning element is the communicative intent that the ECA aims to accomplish, whereas the signal element indicates the multimodal behavior exhibited in order to achieve the desired communicative intent. The multimodal behaviors to express a given communicative function to achieve (e.g. facial expressions, gestures and postures) are described by the Behavior Markup Language (BML) [**?**].

Lately we have been developing data-driven approach to capture the link between acoustic features (speech [25], laughter [24, 26]) and multimodal behaviors. The obtained animations show more subtle motions than those generated with the procedural model.

EMILYA In this section we briefly present the Emilya database (Emotional body expression in daily actions database). The interested reader can find more details in [33]. Eleven (unprofessional) actors participated in the data collection. Both 3D motion capture data (using Xsens technology [1]) and audio visual data were recorded and synchronized. The actors were asked to express 8 emotions (Joy, Anger, Panic Fear, Anxiety, Sadness, Shame, Pride and Neutral) in 7 daily actions (walking, waking with an object in the hands, sitting down, knocking at a door, lifting and throwing an object (a ball made of paper) with one hand, and moving objects (books) on a table with two hands) [33]. Those emotions were selected to cover the arousal and valence dimensions. We asked the actors to perform each action four times in a row to capture a large set of data. A continuous sequence consisting of the series of all the actions with just one trial per action was also recorded. After segmentation, we obtain a database of around 10000 segments depicting expressive body movements. Moreover we have validated this database through perceptual study [33].

2.4 Objectifs et caractère ambitieux/novateur du projet / Objectives, originality and novelty of the project

All developments will be validated by experiments with the theatre department of Paris 8, under the supervision of Georges Gagneré. Starting from a selection of play scripts in various genres and with increasing complexity, theatre experts will use the DADA tools to create virtual theatre performances in the Unity game engine, including stage movements and actions (entering, exiting, sitting down, standing up, taking and putting objects on the stage) ; body language expression of the personalities, moods and emotions of the characters ; and believable gaze, proxemics and action/reaction behaviors between actors.

The expected results of DADA will be (1) a virtual theatre company of autonomous actors with a large vocabulary of expressive animation skills ; and (2) a prototype system for directing arbitrary dramatic plays, amenable to a variety of digital storytelling applications. Results will be integrated into Unity3D which is already used by the GRETA plat-form at Telecom ParisTech and the virtual cinematography framework developed by the IMAGINE team at Inria. Results will be used at University of Marseille for building a pivot actor model allowing the retargeting of the DADA actors to actors with different morphologies and styles. Results will be used by Paris 8 as a virtual rehearsal space for theatre productions involving real actors interacting with digital actors, and as a platform for publishing digital dramatic performances online. If applicable, results will also be patented and exploited by the three academic partners, targeting commercial applications such as video games, digital storytelling, virtual worlds and movie previz.

2.4.1 Expressive virtual actors

Proxemic models The role of the proxemic models is to compute the precise positions and orientations of actors at all time, given the director's blockings.

Kinesic models The role of the kinesic models is to compute the remaining degrees of freedom, given the director's blockings and the precise positions and orientations of actors at all time.

One difficulty will be to generate those remaining degrees of freedom with high quality, avoiding the robotic effects associated with procedural animation, and the repetitive effects associated with data-driven methods.

More precisely, we will work to make each performance plausible (actor maintains personality of the role), expressive (actor follows director's commands) and natural (actor adapts to the environment with variations)

2.4.2 Authoring tools

Expose all important dramatic parameters to the director ; compute all other parameters at runtime.

Real-time animation and synchronization Our solution is based on a prompter system.

The main bottleneck in character animation is the very large dimension of the parameter space (50-80 degrees of freedom per actor).

We will decompose the parameter space into a hierarchy of nested subspaces : (a) blocking parameters controled directly by the director, including actions, attitudes, stage positions and trajectories, etc. ; (b) proxemic parameters computed by the autonomous actors in relation to each other, to the stage and to the audience, given the blocking directions ; (c) kinesic parameters computed by the the autonomous actors to realize the blocking directions, given the their proxemic relations.

3 Programme scientifique et technique, organisation du projet / Scientific and technical programme, Project organisation

3.1 Programme scientifique et structuration du projet / Scientific programme, project structure

Work will be divided into four main work packages : (1) procedural animation of isolated actors ; (2) procedural animation of interaction between actors ; (3) authoring and real-time control ; (4) user evaluations. Through the authoring tool (WP3), a script is elaborated by a theater director (WP4) ; it gives direction to group of actors which act out autonomously the commands of the script to position toward each other and in the virtual space (WP2). The behaviors of each actor is computed taking into account their emotional states and social relations (WP1).

3.1.1 WP1. Kinesic component

This WP aims at creating multi-modal statistical models of an individual character body movements from annotated, mainly from mocap data, to generate novel expressive animation suitable for dramatic performances. To do so we will tackle few difficult and open problems.

First we will work at designing new full body controlers based on recent advances in statistical machine learning and on representation learning. We will focus on designing generic models allowing animation in many settings including emotional state and actor's profile (morphology, expressivity level etc). We want in particular that, while the animation model will be learned from a limited number of actors' data, the controlers should be able to be remapped to other actors. Our idea is to build models that take as input few contextual variables that encode the setting (mood, actor's profile etc) as continuous input variables. In a first step we will extend our previous works on markovian contextual models [88, 23, 27] to full body animation. Next, we will investigate the use of continuous state space models and particularly of (deep) recurrent neural networks. Such models, and some of their variants, have been proposed recently for diffcult recognition tasks on complex signals (e.g. speech [3]) as well as synthesis tasks for handwriting [38]. These frameworks are part of the representation learning line of work which brought impressive breakthrough in few machine learning tasks for signals, speech recognition []Âăas well as in computer vision tasks []. Finally we plan to explore alternative strategies such as using neuro muscular based models following ideas like the one of deltalognormal models from [32] which allow recovering the sequence of neuromuscular commands that generated a handwritten gesture.

A second lock will concern the animation of the face in dialogue situations. Given that we have worked previously with three complementary methods, we will focus here on how to mix our face animation models : a mocap based animation model [**?**], a video-based animation model [**?**], and a procedural anmation model [], which is and part of the GRETA system. The main issue here will be to imagine efficient frameworks for inferring based on the scenario of the animation which model to use or combine to produce a final animation.

Finally we want our animation models to be easily extendable to new activities, gestures and moods, by making them learnable from only few training samples. This will allow enriching the system easily whitout a costly and tedious task of gathering a large corpus of training data as usually required in statistical machine learning. Learning statistical models from few samples is an open issue, it has been adressed few times for simple gestures as handwriting [**?**] [57]. We will go beyond these preliminary studies and will explore two ways that aim at favoring transfer from learning one action / mood model to learning another action / mood model. We will explore strategies for modifying our generic controlers, e.g. based on recurrent neural networks or on contextual markovian models [88], in order to enable transfer learning between action models so that a new action can be learned with only few samples.

3.1.2 WP2. Proxemic component : procedural animation models for interaction between actors.

Previous works on modeling group formation have been mainly applied to ECAs and have focused on the spatial posi-tioning and orientation of the ECAs (Pedica, 2010). Few researches have looked at modeling group of ECAs with dif-ferent personalities and social attitudes (Gillies 2004 ; Prada, 2005). However these models do not consider the dynamic evolution of the group behaviors nor how do the actors' behaviours synchronize with each other. In this task, we focus on simulating group of autonomous actors interacting with each other where each actor is defined by its emotional state and its relation toward others and objects. Social relations can be represented by two dimensions, affiliation and domi-nance (Wiggins 1979). We will extend group behavior model (Pedica 2010) that embeds the F-Formation proposed by Kendon (2004) to consider social relations and emotional states of actors.

Physical distance between actors, their body orientation toward each other, gaze direction, facial expression, gesture expressivity are cues of the relation with others and with objects and of emotional states. These cues will be embedded in the proxemics component. They evolve con-tinuously in relation to the others' behaviors. To simulate the dynamic evolution of these behaviors we will make use of Neural Network simulation (Prepin 2013) where we can render how behaviors of one actor can act on behaviors of other actors (eg walking powerfully toward an actor with an angry expression will result in moving backward of another actor with a less dominant attitude. Mutual coupling of behaviors will be modeled as emerging from such action-reactive behavior simulation (Prepin 2013) ensuring not only the synchronization between actors' behaviors but also their mutual influence. This task will be led by Telecom ParisTech with the contribution of Inria.

3.1.3 WP3. Performance authoring and real-time execution.

This work package will elaborate a common conceptual framework for assembling all the behaviors, goals and animations of all actors into a coordinated, real time performance. Based on this framework, we will develop software tools for authoring the performance and controlling it in real-time. Authoring of performances will be based on traditional cue sheet, which are familiar to theatre directors (Gagneré 2012, Ronfard 2012). Cue-sheet are multi-modal documents consisting of blocking notations written in a pseudo-natural language of verbs and adverbs, together with a graphical annotation providing spatial and temporal cue signals for all actor movements, using stage views and floor plan views. A cue-sheet provides a convenient notation of stage directions, which can be easily created and edited by directors, and used a specification for a virtual performance. Internally, we will compile the cue sheet into a hierarchical finite-state machine, which is a de-facto standard in real-time game engines.

We will take advantage of the motion models created in WP1 and WP2 to create finite-state machines with a rich vo-cabulary of high-level actor behaviors, suitable for generating complex performances. Following (Mateas 2002), we will decompose the input cue-sheet into minimal units of behaviors (beats) organized as one state-machine per actor, all connected together, and one state-machine for a stage manager controlling the advancement of the storyline. Depending on their current states, virtual actors will update their positions, orientations and gaze directions using be-haviors from WP2, and their other animation parameters using procedural models from WP1.

All software tools developed in WP1 and WP2 will thus be integrated into a common runtime, playable in the Unity game engine, and used in WP4 for evaluation and validation. This task will be led by Inria, with contributions from all partners.

3.1.4 WP4. Evaluation and validation.

This task will insure the integration of the research prototype within the cultural context of creative industries and artis-tic practices. Using the autonomous digital actors from WP1, WP2 and WP3, Paris 8 will create short theatre scenes covering the spectrum of actions and emotions covered by the project. The directorial constraints will be adapted to the research scope in order to guarantee expressive results matching creative issues. A survey of teachers and creators from theater, dance, cinema, digital art, video game of Paris 8 creative environment will help to design the prototype in the direction of users' needs. Evaluation and validation will include short staged performances targeting different application areas, including theatre, pantomime, staging of chorists in opera, as well as previsualization of movie scenes and simulation of non player characters in video games. It will aim at a high expressive level of realization and give feed-back on the quality of animation and the usability of the authoring tools offered for directing virtual actors in those contexts. This task will be supervised by Paris 8 with contributions from members of the Labex Arts-H2H leading project Process of directing actors which involves international stage directors teachers and students of the Conservatoire National Supérieur d'Art Dramatique (CNSAD ' National theater school).

3.2 Management du projet / Project management

3.2.1 WP0. Coordination

Four one-day meetings per year -> 14 meetings, including kick-off meeting and final review meeting.

Collegical decision making : A steering committee composed of Thierry Artières, Georges Gagneré, Catherine Pelachaud and Rémi Ronfard will make all important decisions in unanimity. In cases of disagreements, a compromise will have to be found. The committee will meet before every consortium meeting and its decisions will be communicated to all consortium members and to ANR.

Software development will be coordinated by Inria and Paris 8 using rapid prototyping methods. All PHD Students will be asked to contribute their latest results to be included in the DADA prototype at least twice a year (two months each). The rest of their time will be devoted to their research work. References on rapid prototyping would be useful.

Source code files will be signed by all contributing authors, together with their affiliations, to properly track intellectual property rights.

A consortium agreement will be signed in the course of the first year to define a common intellectual property rights policy. In order to patent their inventions, partners should seek authorization from all other partners. The other partners cannot prevent the patent unless they have prior art. The partners can ask to be mentioned as co-inventors if they can demonstrate that they contriuted to the invention.

The consortium is composed exclusively of academic partners, whose goal is primarily to disseminate and publish their research work, not to commercialize software.

On the other hand, some of the inventions may have a commercial value. We will seek advice from a small board of industry experts from relevant French companies (Goalem, Quantum Dream, Ubisoft, Dassault Systèmes) to detect and address such cases and make sure that potentially important inventions are protected and made available to them.

3.3 Description des travaux par tÃćche / Description by task

3.3.1 WP1 Kinesics

WP1	Kinesic component
Responsable	LIF
Participants	Inria, Telecom ParisTech
Duration	42 months
Objectives	Develop new generic controlers of a single character alone based on precise spatio-temporal indications of his actions and mood.
Task 11	Full body animation (T0 → T0+36)
Task12	Face and gesture animation (T0+12 → T0+36)
Task13	Learning from few samples (T0+18 → T0+42)

The goal of this WP is to develop new generic models able to produce animation of a single character. It includes designing animation models of a character realizing an action (walking, sitting etc) given a context that consists in a particular mood and in character profile (age, gender) as well as designing models for taking into account the interaction of the character with others (gaze, harm gesture). Moreover we will explore transfer learning strategies for learning these models from few training samples only to ease addition of new getures, actions and moods.

The workpackage is divided into three subtasks : *animation of the full body of a character alone*, *animation of the face of a character engaged in dialogue*, and *strategies for learning models from few training data*.

Task 1.1. Generic full body animation model We will first focus on the design of generic body controllers able to synthesize the animation of the full body of a character (through the sequence of mocap representation) for a given procedural animation scenario as output by WP2, i.e. a sequence of actions realized with a particular mood context and for a specific character profile (morphology, expressivity level). We will investigate modeling frameworks that allow taking into account the contextual variables (e.g. mood components and actor's profile components) as few inputs which influence the animation. Designing models whose output (a synthesized animation) continuously varies with these contextual inputs naturally yield easy generalizing to unseen settings in the training set (e.g. particular combination of action, mood and actor's morphology). We plan to investigate the following lines of research :

Firstly we will explore how to extend *contextual markovian models* whose parameters (means of Gaussian distribution, transition probabilities...), are parameterized by (i.e. defined as a function of) of contextual variables (mood and profile information). One Contextual HMM may be viewed as a continuum of HMMs, one model for every possible value of contextual variables. Recent work has demonstrated such models for the case of locomotion believable controllers, gesture controllers [62] and face controllers [88, 23, 27]. We will aim to generalize these works to more general action controllers, including such actions as : sitting, standing, walking, grasping, taking and putting objects, in a variety of expressions and moods. These models will serve as a baseline for evaluating new modeling approaches.

Second, we will investigate the use of *(deep) neural networks* and of dynamic versions of these (i.e. recurrent neural nets) which have demonstrated strong abilities to model, to classify and to synthesize complex signals such as speech or handwriting [3, 39, 2, 68, 40, 38]. These models are related to what is called *representation learning* which emerged in the last few years as a key topic in the machine learning community [2] [18]. One main difficulty will be to integrate the use of contextual information as input in order to modify the behaviour of the models. We plan to extend the principle of contextual markovian models to neural nets by investigating ideas like designing bilinear layers in the neural net where weights could be defined as a function of the contextual input, inspired by works like [116, 50].

At last we will investigate *low dimensional state space models* such as neuro muscular based models following ideas like [32] which aims at recovering from a handwritten signal the sequence of neuromuscular commands that generated the handwriting signal. The underlying idea here is to exploit such models in order to work in a new representation space, the space of neuromuscular commands that generate motion, rather than on the observed motion itself. Although such models have not been used to model complex gestures up to now it is expected that they could be robust enough to provide good estimation of the command sequence. The main advantage of such a change of representation space is an expected reduction of the dimension of this space (as in [109]), enabling easier learning from few samples and transfer learning (as will be investigated in **task 1.3**).

Task 1.2. Combining models for face animation The second task focuses on learning models of gesture and facial expressions in dialogue situations. It is dedicated to the combination of animation models, whch is a diifuckt and open question, with a focus on the animation of the face. We will start from available face animation models in the consortium : a mocap based animation [23], a video-based animation [**?**], and a procedural animation [80] (integrated in the GRETA system).

All of these models types have pros and cons. While statically-driven models are more prone to produce natural looking animation, cognitive models capture more precisely the semantic emotional behaviors to communicate. These latter ones are often event-driven ; that is they compute a behavior only when a given communicative function is specified. Statically driven models produce animation continuously that captures the communicative colour of the message to convey but they have difficulty to compute behaviors which have specific meaning. As a result, virtual agents driven by cognitive-like system are able to convey more precise displays while those driven by statistical models look more natural and lively [61].

[2] See the recently born ICLR conference on Learning Representations at http ://www.iclr.cc/

We will explore ways to combine few such animation models which remains an open question today, be it for animating the face ot the full body [**?**]. We will explore strategies and implement these within the Greta framework where communicative intentions and emotions are represented with the FLM language while multimodal behaviors with BML [108]. The merge of multiple animation models may be performed as a weighted blend of the animations produced where the weights might be context dependent and tuned either manually or automatically, alike in [98]. Alternatively the animation models may be merged earlier, when deciding which kind of motion to launch, or may have asymetric role. For instance, the procedural animation model (or semantically-driven ?) might act as the main animation model and use when neccessary animatyions produced by the other models.

Task 1.3. Learning from few samples We will mainly investigate two approches for extending approaches developped in task T1.1 to enable learning from few samples. The first strategy consists in extending the idea of context variables that models of task T1.1 rely on in order to design a global model for all actions. In the case of markovian model for instance this means that instead of definig one model per action one could define a unique global markovian model where every state would stand for a particular position of the body and performing an action would correspond to following a path (i.e. a state sequence) in this big model. Making transition probablities dependent on the action to perform such a big model would be instaiated as an action model by considering a bundle of paths only in this model. Doing so one could expect that all the training data (whatever the action it corresponds to) could be exploited to learn all the states of this big markovian model, hence implementing some kind of transfer learning between actions. A new action would correspond then in a bundle of paths in this model and could be learnt from few samples only. Preliminary works that we did let us expect that such a strategy would work with statistical markovian models [28, **?**]. In this case the above idea could be implementing by introducing new contextual variables, which might be at the simplest on-hot indicators of the action to perform (a vector with zeros everywhere but at the position of the action number). We will first investigate this strategy for contextual markovian models then we will extend this approach to recurrent neural networks. A related approach will concern the use of using continuous state space models with a low dimensional state space (e.g. corresponding to the degree of freedom of body poses or to the neuro muscular commands) which should permit characterizing a particular motion or gesture as its dynamic in this latent space whose limited dimension would enable learning from few samples.

Deliverables bla bla bla

Deliverables	Name and content	Date
L1.1	Report on the state of the art for statistical models for animation synthesis	
L1.2	First version of the models : Prototype (software) and its documentation (Report on the models deveopped)	T0 + 18
L1.3	Second version of the models : Prototype (software) and its documentation (Report on the models deveopped)	T0 + 36

Partners' roles bla bla

Risks The risks are limited. There will not be any problem with availability of datasets. All along the project we will rely as much as possible on existing datasets. For instance Mocap data of considered actions have already been recorded by C. Pelachaud within the project Feder Anipev (http ://www.anipev.com/). The corpus EMILYA (EMotional body expressIon in daiLY Actions databaseBodily Emotional Actions Behavior) (Fourati, 2014) is constituted of 7 actions performed by 11 actors with 8 emotions. The actions encompass everyday actions such as walking, carrying an object, and sitting. The emotions cover the positive and negative spectrum.

3.3.2 WP2 Proxemics

WP2	Proxemic component
Responsable	LTCI
Participants	Inria, ECM
Duration	
Objectives	
Content	
Task 21	Communicative behaviours
Task22	Steering behaviours
Task23	Combination of statistical and procedural models

In this workpackage we are interested in modeling behaviors of group of agents while conversing and while moving around. We will pay particular attention at the social interaction of the agents during these activities. We will also develop an animation model that incorporates two models : statistical model as developed in WP1 and procedural model developed within the Greta platform.

Task 2.1 : Group behaviors during multi-way conversation In this task we will model multi-party conversation behaviors. We will focus on turn-taking management. While indication of what the agents would say to whom and when will be provided by a script (Task 3.1 and Task 4.X), the turn-taking model will instantiate which behaviors the agents will display. Gaze, body orientation, position in space are important cues for indicating who has the turn, who wants to keep it, to give it to someone, who listens ? We will extend an existing turn-taking model [89] that is based on Sack ?s model [93], that embeds F-Formation [52] and that takes into account social attitude of the agents toward each other. This model is implemented as a state machine where the states are defined by the turn-taking and correspond to conversational roles. Transition between states is triggered when an agent changes conversational role. Attitudes vary the behavior of the agents such as their propensity to gaze at others. We will extend this model to simulate different configurations of speech overlap such as terminal overlaps, conditional access to the turn, and choral [95] as well as long silences when nobody takes the turn. We will add further states to encompass more conversational functions (eg greeting, word search ?). We will also model that transitions from one state to another one can bring the agents of a group to be in the same state (parallel configuration as when greeting each other or laughing together).

Task 2.2 : Group behaviors during stage movements ? implementation of advanced ń steering behaviors ż such as follow, flee, separate, join, merge, enter stage, exit stage,

etc. This task will model agents ? behavior when moving around in the environment. The animation of the virtual agent doing some tasks will be given by WP1. It will not focus on path planning as this information will be provided by a script (Task 3.1 and Task 4.X). Rather it will model how agents perform displacement in social settings. Gaze direction, body orientation and spatial distance to other agents will be computing for different ?steering behaviors ?. These features will be modeled through different synchronization mechanisms : moving in synch, moving ahead, following, etc. They evolve dynamically in function of each agent ?s position and orientation in space. The basic animation of the agent, ie without any influence from surrounding agent, is given by WP1. To simulate the dynamic evolution of agens ? behaviors we will make use of Neural Network simulation [87] where we can render how behaviors of one actor can act on behaviors of other actors (eg walking powerfully toward an actor with an angry expression will result in moving backward of another actor with a less dominant attitude. Mutual coupling of behaviors will be modeled as emerging from such action-reactive behavior simulation [87] ensuring not only the synchronization between actors ? behaviors but also their mutual influence.

Task 2.3 : Combination of statistical and procedural models. In this task we will develop an animation model that will merge animations coming from statistical model developed in WP1 and procedural model developed in WP2 (Task 2.1 and Task 2.2). This blend is required for the interaction settings where behaviors of the agents are driven by both animation models. The procedural model relies on forward and inverse kinematic models [49]. It controls the arms position, gaze direction and body orientation. The statistical model (from WP1) controls the whole body. Our animation blender model will work at the modalities level and will also incorporate movement propagation ; that is how motion of one body part affects other body parts. At first, the animation blender model will merge whole body motion computed by the statistical model as specific body motion computed by the procedural model. More precisely, arms position, gaze direction and body orientation outputted by the procedural model will be viewed as constraints to be reached. These motions will be added onto the animation computed by statistical model ; the position of the arms, head and torso computed by the procedural model will overwrite those computed by the statistical model. In a second step, the animation blender model will incorporate propagation of movements. To compute movement propagation we will develop a statistical model that learns which motion is due to action and which motion is due to movement propagation.

Deliverables bla bla

Deliverables	Name and content	Date
L1.1	Report on the state of the art of proxemics models in computer animation	
L1.2		
L1.3		

3.3.3 WP3 Authoring

WP3	Authoring
Responsable	Inria
Participants	Paris 8, ECM, Telecom ParisTech
Duration	
Objectives	
Content	
Task 31	Blocking language
Task12	Authoring tools
Task13	Real-time animation

Task31 : Specification of a dramatic language for virtual actors. This will include a choice of verbs (actions, speech acts, movements) and adverbs (moods, attitudes, dramatic effects) for directing actors ; define cues as synchronisation points between actors ; define parallel and sequential behaviors ; etc.

Part of this language will be devoted to stage blocking / movement

Part of this language will be devoted to dialogue

Previous work [35, 92, 34].

In theatre, blocking is the precise movement and staging of actors on a stage in order to facilitate the performance of a play, ballet, film or opera.

A theatrical cue is the trigger for an action to be carried out at a specific time. It is generally associated with theatre and the film industry. They can be necessary for a lighting change or effect, a sound effect, or some sort of stage or set movement/change.

A cue sheet is a form usually generated by the stage manager or design department head that indicates information about the cue including execution, timing, sequence, intensity (for lights), and volume (for sound). The stage manager keeps a master list of all the cues in the show and keeps track of them in the prompt book.

The prompt book, also called prompt book, transcript, the bible or sometimes simply "the book," is the copy of a production script that contains the information necessary to create a theatrical production from the ground up. It is a compilation of all blocking, business, light, speech and sound cues, lists of properties, drawings of the set, contact information for the cast and crew, and any other relevant information that might be necessary to help the production run smoothly and nicely.

The Prompt Book is the master copy of the script or score, containing all the actor moves and technical cues, and is used by the deputy stage manager to run rehearsals and later, control the performance.

Task32 : Authoring tools for blocking a scene with multiple actors. Design and implementation of authoring tools for creating animation with the dramatic language.

Previous work has focused on direct annotation of play-scripts with high-level (FML) or low-level (BML) mark-up.

From a user perspective, this is neither intuitive nor expressive. Instead, we will offer authoring tools with natural interaction, taking inspiration from existing practices in theatre (prompt-books, cue sheet, storyboards, etc.).

The authoring tool may include multimodal interaction with the director : sketching tools for designing actor trajectories and meeting points ; writing tools for adding didascalia to dialogues ; timeline-driven interaction for defining cue points and actions, timing, etc.

User interface for directing actors by sketching stage floor plans and composing the dramatic score ; one line per actor per motion component (proxemic behaviors, kinesic actions, kinesic moods, speech acts, etc.)

Compilation of the language into a finite state machine and/or Petri net ; allowing real-time execution of the dramatic score.

Task33 : Real-time execution of the dramatic score. This should include real-time combination of proxemic (procedural) and kinesic components of motion ; non-deterministic motion generation ; synchronization to cues ; real-time skinning and advanced 3D animation ; integration of physically-based secondary animation (skin, hair, clothes, etc.)

This includes integration of the GRETA BML realizer with IMAGINE animation ; and real-time integration of the statistical models of motion with the procedural animation components.

One challenge to be overcome is in combining full body animation and interaction animation at runtime.

We believe it will be an importa asset for the DADA platform that each performance is unique, and can be controled in real time by cues given by the director.

We will pay particular attention to design models capable of generating real animations. Indeed synthezing from statistical models usually resumes to finding the most likely animation sequence in a given situation, which may yield to too similar and unrealistic animations.

Actually one would be pretty much interested in synthezing animations that are both likely given the learnt statistical models but also exhibiting the variability one can observe in human motion and gestures. Introducing such a stochastic component in the synthesis while maintaining a high quality animation level is not straightforward and is an open question that we will have to solve.

Partners' roles Inria will be the main software developper.

Task 3.1 will be jointly performed by Inria, LITC and Paris 8.

LIF will contribute to task 3.3 on implementing non-deterministic animation methods using statistical models trained in WP1.

Paris 8 will contribute to tasks 3.2 and 3.3 by being the "product owner" for the authoring tool.

LITC will contribute to task 3.3 by providing a subset of the GRETA platform.

Deliverables bla bla

Deliverables	Name and content	Date
L1.1	Report on the state of the art for virtual theatre	
L1.2		
L1.3		

3.3.4 WP4 User evaluation and validation

WP4	User evaluation and alidation
Responsable	Paris 8
Participants	ECM, Inria, LTCI
Duration	
Objectives	
Content	
Task 41	Scenarios
Task 42	Validation of interaction
Task 43	Validation of animation

Task41 Scenarios.

Writing scenes with didascalia

Dialogue scenes with groups of 2 or 3 actors using a choice of didascalia

Movements with groups of 2 or 3 actors using a choice of didascalia

Alternations of dialogue and stage movements in theatre scenes with 2 or 3 actors

A possible choice would be "the augmentation", a play by Georges Perec with a large number of variations on a single theme (an employee asks an augmentation from his boss in the presence of his secretary).

Task42 Validation of the interaction.

Is the dramatic language adequate ? useful ? efficient ?

Is the dramatic score interface adequate ? useful ? efficient ?

Is the stage floor plan sketching tool adequate ? useful ? efficient ?

Task43 Validation of the animation

Dialogue scenes with groups of 2, 3 and 4 actors.

Silent stage movements of groups of 2, 3 and 4 actors, as in opera synched to music

Combination of dialogue and action for scenes with 2 actors

Deliverables bla bla

Deliverables	Name and content	Date
L4.1	Selection and annotation of example scenes	
L4.2	Usability of authoring tools	
L4.3	Evaluation of single-character and multiple-character animation	

Additional notes

We will dedicate joint research between Inria and LIF to make it easy to extend our database of actions and attitudes using video, rather than motion capture. This will necessitate fundamental research in transfer learning (so that the sparse data obtained from video can benefit from the dense data obtained with motion capture) and video processing. Following the methodology of gesture controllers [63], where the gesture are controlled directly by speech prosody features extracted from real actors voices, it appears possible to drive expressive and plausible gestures and body movements from visual signatures of actions and attitudes extracted from example videos.

We will use our previous work in actor and action recognition [112, 111, 37] to detect and recognize actors and their actions in real movies ; and extract visual signatures of the corresponding actions and attitudes. Based on this analysis, we will learn joint statistical models for driving gesture controllers from those video signals.

Combining proxemics and kinesics components can be done along the lines of Mitake et al. [79], where the degrees of freedom of a virtual character are separated into six parameters for rigid body simulations, and four parameters for encoding multi-dimensional keyframe animations. Similarly, we would like to hide the complexity of high-dimensional character animation (with 40-60 degrees of freedom) behind a small number of control parameters. We will extend rigid body simulations to include proxemic interaction forces in WP2. And we will replace keyframe animations with statistical models learned from data in WP1.

One promising avenue for research will be to design strategies for controlling the proxemic components of character animation using the rigid motion of the head, rather than the full body. Sreenivasa et al. [100] have proposed inverse kinematics methods for computing the body motion of a humanoid robot, including footsteps and walking patterns of motion, given its head motion. In the context of DADA, the head motion of the virtual actors could similarly be put under the direct control of the director because it plays such an important expressive and dramatic function. The full body motion could then be computed with the constraints that the actor's head motion matches the director's directions, and the prescribed actions (walking, sitting, standing, etc.) and attitudes (sadly, swiftly, merrily, etc.).

3.4 Calendrier des tÃćches, livrables et jalons / Tasks schedule, deliverables and milestones

4 Stratégie de valorisation, de protection et d'exploitation des résultats / Dissemination and exploitation of results. intellectual property

5 Description de l'equipe / Team description

5.1 Description, adéquation et complémentarité des participants / Partners description, relevance and complementarity

The consortium involves three research teams with complementary experience in computer graphics, intelligent virtual agents and statistical machine learning and a research team in theatre studies. Telecom ParisTech and University of Marseille are already working together on facial animation from speech through the co-supervision of Yu Ding's thesis (Ding 2013). Inria/Imagine and Paris 8 are also already working together on directing audiovisual prosody of actors, as part of Adéla Barbulescu thesis (Barbulescu 2014). Results of the two theses will be exploited in the project.

5.2 Qualification du coordinateur du projet / Qualification of the project coordinator

Remi Ronfard is a computer scientist with a 20 year experience in industry and academia in France, Canada and USA. He has worked at the T.J. Watson IBM Research Center in New

Partner	Name	First name	Position	Field of research	PM	Contribution
LIF	Artières	Thierry	Pr	Machine Learning	14	WP1 (task leader), WP2, WP3
LIF	Emyia	Valentin	Assistant Pr	Machine Learning and Signal Processing	5	WP1
LIF	Qi	Wang	Ph.D. student	Machine Learning	12	WP1

TABLE 1 – Qualification and contribution of each partner

York as post-doc and as a visiting scientist (1992 and 2000). He is now a member of the IMAGINE research team at Inria and the University of Grenoble, where his research is devoted to designing novel interfaces between artists and computers. He is the author of 4 international patents and more than 60 scientific papers published in top ranked international journals (IJCV, CVIU, PAMI) and conferences (Siggraph, Eurographics, CVPR, ICCV, ECCV) and cited more than 4000 times. He will be acting as coordinator for DADA.

Remi was trained as an engineer then PhD student at Mines Paris Tech. He has conducted research in a variety of domains, including digital storyboarding (INA, 1995), aesthetic surface design (IBM Research and Dassault Systèmes, 2000), video indexing (INA, 2000), action recognition and statistical analysis of image and film styles (INRIA 2002-2007). He was an expert in the international MPEG group from 1997 to 2000. In 2007, he became a team leader at Xtranormal Technologies. His team created the patented ?magicam ? system, which was used to produce two million user-generated 3D animation movies. He came back to INRIA in 2009 with a new research program devoted to ?directing virtual worlds ?. He helped to create the IMAGINE team in 2012, where he now leads the ?narrative design ? part of the project. Towards this goal, he investigates computational models of visual storytelling. This has led to inspiring collaborations with the national film school (ENS Louis Lumière) and the Célestins Theatre in Lyon. He has co-chaired international workshops on modeling people and human interaction (Beijing, 2005), 3-D cinematography (New York City, 2006 ; Banff, 2008 ; Providence, 2012), intelligent cinematography and editing (Quebec, 2014). He is currently serving as head of the ?Geometry and Image ? Department at Laboratoire Jean Kuntzmann, Univ. Grenoble Alpes.

5.3 Qualification, rÃt’le et implication des participants / Qualification and contribution of each partner

INRIA EPI IMAGINE IMAGINE stand for : "Intuitive Modeling and Animation for Interactive Graphics & Narrative Environments". The challenge we aim to address is the efficient, interactive creation of animated 3D content. To this end, our goal is to develop a new generation of knowledge-based models for shapes, motions and stories. These models will embed both procedural methods, enabling the fast generation of high quality content, and intuitive control handles, enabling users to easily convey their intent and to progressively refine their result. These models will be used within different interactive environments dedicated to specific applications. More precisely, we will apply our work to three main domains : shape modeling, motion synthesis and narrative design. In addition to addressing specific needs of digital artists, this research should in the long term, enable professionals and scientists to represent and interact with models of their objects of study, and educators to quickly express and convey their ideas. Our international scientific partners include UC Berkeley, UBC, the University of Toronto, McGill and ETHZ and Disney Research Zurich.

In adddition to Remi Ronfard, two other permanent researchers and a post-doctoral student will actively participate to the DADA project.

Marie-Paule Cani is a full-time professor at INPG and director of the IMAGINE team. She will contribute to DADA with her recent work on implicit skinning, advanced hair style rendering, advanced clothe adaptation, etc.

Damien Rohmer is an associate member of the IMAGINE team. He will contribute to DADA with his recent work on implicit skinning ? physically-based animation ?

Adela Barbulescu is a third-year Phd student who will work part-time on the DADA project during her post-doc in 2016. She will contribute to DADA with her recent work on visual prosody, which will be extended for joint generation of speech and facial animation from directorial input.

ECM Two main researchers from the QARMA team will participate to the project.

Thierry Artières is a professor at University of Aix-Marseille, and a member of the *QARMA team* (eQuipe AppRentissage et MultimÃľdia) at LIF (Laboratoire dâĂŹInformatique Fondamentale). One of his major research topic concerns machine learning for multimedia applications, more particularly for sequences and signals, either for classification, pattern discovery, sequence labeling and sequence synthesis, with strong experience with various signals such as speech, bioacoustics, handwriting, gestures, eye movements, WII signals, Kinect and motion capture data. He is author or co-author of about sixty papers and articles in top ranked international conferences (NIPS, ICML, AISTAT, ICASSP, EMNLP) and journals (IEEE PAMI, JMLR, Pattern Recognition) in the fields of theoretical as well as applied machine learning (speech and handwriting recognition, user modeling) and artificial intelligence.

Valetin Emiya is assistant professor in the QARMA team at LIF since 2011. He has conducted research in audio processing and sparse models for 8 years and has strong connexion with the signal processing group at I2M Lab in Marseille. His current works on models and algorithms for audio inpainting (see [5, 4] and project ANR JCJC MAD), i.e. interpolation and extrapolation in audio sequences. This works are currently being extended to the extrapolation of gesture for the control of electronic musical instrument and contemporary music creation, through the Progest project by GdR ISIS (2014-2016) in collaboration with the gmem Centre National de Création Musicale (http://www.gmem.org/index.php?option=com_content&view=article&id=5580144&Itemid=13660).

Wang Qi is a first-year Phd student whose research topic on recurrent neural networks for signal processing tasks is related to the project. He will contribute to DADA mainly on WP1 (tasks 1.1 and 1.3).

Paris 8 **Georges Gagneré** is a lecturer in performing arts at the University Paris 8 where he teaches acting in digital environments. As part of Labex Arts H2H, he is an active member of two projectsãon ńăActor directing as art creation processăż (La direction d ?acteurs comme processus de création artistique), and ńăAugmented sceneryăż (La Scène Augmentée). He is also a stage director and member of the collaborative platform didascalie.net, focusing on real time intermedia environments in performing arts. In 2007, he initiated the research project ANR

VIRAGE about methods and software prototypes for cultural industries and for the arts. He is involved in the OSSIA and INEDIT ANR project through the realization of the artistic project ParOral, based on digital shadows direction through the voice, with the Iscore software. He directed productions in national theaters (Théâtre National de Strasbourg, La Filature, Scène nationale de Mulhouse, Théâtre Gérard Philipe, Centre dramatique national de Saint-Denis) and organized numerous workshops on the impact of real time new technologies on theater and scenic writings. He collaborates with Stéphane Braunschweig and Peter Stein as stage director first assistant on more than 20 differents opera productions in the most famous european theaters (La Scala, La Fenice, Théâtre des Champs-Elysées, L'Opéra Comique, Le Festival International d'Art Lyrique d'Aix-en-Provence, La Monnaie ? Bruxelles, L'Opéra de Lyon, L'Opéra du Rhin, etc.).

Jean-François Dusigne is professor of Paris 8 University, and ex-actor of Théâtre du Soleil (Ariane Mnouchkine). He will bring his international expertise on the different ways of directing actors, and the transmission issues.

Isabelle Moindrot is the director of the ARTS H2H Labex. She will help to the integration of DADA in the global artistical research ecosystem of Paris 8.

Martial Poirson is a professor at Paris 8 University, and director of the theatre departement. He will help to the dissemination of the DADA's deliverables though the academic and professional fields.

LTCI LTCI (Laboratoire de Traitement et Communication de l ?Information) is a joint laboratory between CNRS and TELECOM ParisTech (UMR 5141). It hosts all the research efforts of TELECOM ParisTech (a faculty of about 150 full-time staff (full professors, associate and assistant professors), 30 full time researchers from CNRS and 300 Ph.D students). Its disciplines include all the sciences and techniques that fall within the term "Information and Communications" : Computer Science Networks, Communications, Electronics, Signal and Image Processing, as well as the study of economic and social aspects associated with modern technology.

Catherine Pelachaud is Director of Research at CNRS in the laboratory LTCI, TELECOM ParisTech. She received her PhD in Computer Graphics at the University of Pennsylvania, Philadelphia, USA in 1991. Her research interest includes representation languages for agents, embodied conversational agents, nonverbal communication (face, gaze, and gesture), expressive behaviours and multimodal interfaces. She has been involved and is still involved in several European projects related to multimodal communication (EAGLES, IST-ISLE), to believable embodied conversational agents (IST-MagiCster, FP5 PF-STAR), emotion (FP5 NoE HUMAINE, FP6 IP CALLAS, FP7 STREP SEMAINE) and social behaviours (FP7 NoE SSPNet, H2020 Aria-Valuspa).

Chloé Clavel is Assistant Professor at Telecom Paristech. She owned a PhD on acoustic analysis of emotional speech. Before joining Telecom ParisTech she worked as a researcher at Thales Research and Technology where she focused on emotion analysis ; then she became a researcher at EDF R & D working on sentiment analysis and opinion mining. She has participated to several collaborative projects and has coordinated one national project.

Yu Ding is a post-doctoral student at LTCI. He has obtained his PhD in September 2014 under the supervision of Thierry Artières and Catherine Pelachaud. His topics of interest are to develop data-driven approach for expressive animation of virtual agents.

6 Justification scientifique des moyens demandés / Scientific justification of requested ressources

Budget : We request a financial aid of 450 K€for 3 PhD students (360 K€), 1 post-doc at Paris 8 (40 K€), computer hardware and software (10 k€), travel expenses (40 K€). The project duration should be 42 months in order to develop a functional prototype and to use it to animate several play scripts.

6.1 équipement / Equipment

6.2 Personnel / Staff

6.3 Prestation de service externe / Subcontracting

6.4 Missions / Travel

6.5 Dépenses justifiées sur une procédure de facturation interne / Costs justified by internal procedures of invoicing

6.6 Autres dépenses de fonctionnement / Other expenses

7 Références bibliographiques / References

Références

[1] Xsens. http ://www.xsens.com.

[2] O. Abdel-Hamid, L. Deng, D. Yu, and H. Jiang. Deep segmental neural networks for speech recognition. In *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, pages 1849–1853, 2013.

[3] O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, 22(10) :1533–1545, 2014.

[4] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley. A constrained matching pursuit approach to audio declipping. In *Proc. of ICASSP*, May 2011.

[5] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley. Audio inpainting. *IEEE Trans. Audio, Speech, Lang. Proc.*, 20(3) :922 –932, Mar. 2012.

[6] A. Allombert, M. Desainte-Catherine, and G. Assayag. Iscore : A system for writing interaction. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '08, pages 360–367, New York, NY, USA, 2008. ACM.

[7] A. Barbulescu, T. Hueber, G. Bailly, and R. Ronfard. Audio-Visual Speaker Conversion using Prosody Features. In *AVSP - 12th International Conference on Auditory-Visual Speech Processing (AVSP 2013)*, pages 11–16, Annecy, France, Aug. 2013.

[8] A. Barbulescu, R. Ronfard, G.-L.-I. Bailly, Gérard, G. Gagneré, and H. Cakmak. Beyond Basic Emotions : Expressive Virtual Actors with Social Attitudes. In *7th International ACM SIGGRAPH Conference on Motion in Games 2014 (MIG 2014)*, pages 39–47, Los Angeles, United States, Nov. 2014.

[9] L. Blackwell, B. von Konsky, and M. Robey. Petri net script : a visual language for describing action, behaviour and plot. In *Computer Science Conference, 2001. ACSC 2001. Proceedings. 24th Australasian*, pages 29–37, 2001.

[10] R. Boulic, J. Varona, L. Unzueta, M. Peinado, A. Suescun, and F. Perales. Evaluation of on-line analytic and numeric inverse kinematics approaches driven by partial vision input. *Virtual Reality*, 10(1) :48–61, 2006.

[11] M. Brand and A. Hertzmann. Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 183–192, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[12] R. Brouet, A. Sheffer, L. Boissieux, and M.-P. Cani. Design Preserving Garment Transfer. *ACM Transactions on Graphics*, 31(4) :Article No. 36, July 2012.

[13] S. R. Buss and J.-s. Kim. Selectively damped least squares for inverse kinematics. *Methods*, 10(3) :1–13, 2004.

[14] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational Agents*. MIT Press, 2000.

[15] C. Chao, J. Lee, M. Begum, and A. Thomaz. Simon plays simon says : The timing of turn-taking in an imitation game. In *RO-MAN, 2011 IEEE*, pages 235–240, July 2011.

[16] C. Chao and A. Thomaz. Timing in multimodal turn-taking interactions : Control and analysis using timed petri nets. *Journal of Human-Robot Interaction*, 1(1) :4–25, 2012.

[17] C. Chiu and S. Marsella. Gesture generation with low-dimensional embeddings. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, pages 781–788, 2014.

[18] G. Contardo, L. Denoyer, T. Artières, and P. Gallinari. Learning states representations in POMDP. *CoRR*, abs/1312.6042, 2013.

[19] N. Dael, M. Mortillaro, and K. Scherer. The body action and posture coding system (bap) : Development and reliability. *Journal of Nonverbal Behavior*, 36(2) :97–121, 2012.

[20] N. Dael, M. Mortillaro, and K. Scherer. Emotion expression in body action and posture. *Emotion*, 12(5) :1085–1101, 2012.

[21] N. d ?Alessandro, J. Tilmanne, M. Astrinaki, T. Hueber, R. Dall, T. Ravet, A. Moinet, H. Cakmak, O. Babacan, A. Barbulescu, V. Parfait, V. Huguenin, E. S. Kalaycı, and Q. Hu. Reactive Statistical Mapping : Towards the Sketching of Performative Control with Data. In Y. R. . T. C. . J. R. . L. M. Camarinha-Matos, editor, *Innovative and Creative Developments in Multimodal Interaction Systems*, volume 425 of *IFIP Advances in Information and Communication Technology*, pages 20–49. Springer, 2014.

[22] B. De Carolis, C. Pelachaud, I. Poggi, and M. Steedman. Apml, a markup language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, *Life-Like Characters*, pages 65–85. Springer Berlin Heidelberg, 2004.

[23] Y. Ding, T. Artières, and C. Pelachaud. Modeling multimodal behaviors from speech prosody. In *International Conference on Intelligent Virtual Agents (IVA)*, 2013.

[24] Y. Ding, J. Huang, N. Fourati, T. Artières, and C. Pelachaud. Upper body animation synthesis for a laughing character. In *Intelligent Virtual Agents*, pages 164–173. Springer International Publishing, 2014.

[25] Y. Ding, C. Pelachaud, and T. Artires. Modeling multimodal behaviors from speech prosody. In *13th International Conference of Intelligent Virtual Agents - IVA*, 2013.

[26] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières. Laughter animation synthesis. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 773–780, 2014.

[27] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières. Laughter animation synthesis. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, pages 773–780, 2014.

[28] Y. Ding, M. Radenen, T. Artières, and C. Pelachaud. Speech-driven eyebrow motion synthesis with contextual markovian models. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, pages 3756–3760, 2013.

[29] S. Dixon. *Digital Performance*. MIT press, 2007.

[30] M. S. El-Nasr. A user-centric adaptive story architecture : Borrowing from acting theories. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, pages 109–116, New York, NY, USA, 2004. ACM.

[31] A. W. Feng, Y. Xu, and A. Shapiro. An example-based motion synthesis technique for locomotion and object manipulation. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '12, pages 95–102, New York, NY, USA, 2012. ACM.

[32] A. Fischer, R. Plamondon, C. O'Reilly, and Y. Savaria. Neuromuscular representation and synthetic generation of handwritten whiteboard notes. In *14th International Conference on Frontiers in Handwriting Recognition, ICFHR 2014, Crete, Greece, September 1-4, 2014*, pages 222–227, 2014.

[33] N. Fourati and C. Pelachaud. Emilya : Emotional body expression in daily actions database. In *Language Resources and Evaluation Conference (LREC)*, 2014.

[34] G. Gagner and C. Plessiet. changes entre metteur en scne et artiste numrique propos de la direction d'acteur. In J.-F. Dusigne, editor, *La direction d'acteurs peut-elle s'apprendre ?* Les Solitaires Intempestifs, 2015.

[35] G. Gagneré, R. Ronfard, and M. Desainte-Catherine. La simulation du travail théâtral et sa " notation " informatique. In M. Martinez, S. Proust, and M. Pouget, editors, *Notation du travail théâtral, du manuscript au numérique*. Lansman, Dec. 2012.

[36] Q. Galvane, M. Christie, R. Ronfard, C.-K. Lim, and M.-P. Cani. Steering Behaviors for Autonomous Cameras. In *MIG 2013 - ACM SIGGRAPH conference on Motion in Games*, MIG '13 Proceedings of Motion on Games, pages 93–102, Dublin, Ireland, Nov. 2013. ACM.

[37] V. Gandhi and R. Ronfard. Detecting and Naming Actors in Movies using Generative Appearance Models. In *CVPR 2013 - International Conference on Computer Vision and Pattern Recognition*, pages 3706–3713, Portland, Oregon, United States, June 2013. IEEE.

[38] A. Graves. Generating sequences with recurrent neural networks. *CoRR*, abs/1308.0850, 2013.

[39] A. Graves, A. Mohamed, and G. E. Hinton. Speech recognition with deep recurrent neural networks. *CoRR*, abs/1303.5778, 2013.

[40] A. Graves and J. Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, 2008*, pages 545–552, 2008.

[41] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. LSTM : A search space odyssey. *CoRR*, abs/1503.04069, 2015.

[42] M. Guay, M.-P. Cani, and R. Ronfard. The Line of Action : an Intuitive Interface for Expressive Character Posing. *ACM Transactions on Graphics*, 32(6) :Article No. 205, Nov. 2013.

[43] M. Guay, R. Ronfard, M. Gleicher, and M.-P. Cani. Space-time sketching of character animation. *ACM transactions on Graphics, Proceedings of Siggraph*, 2015.

[44] E. Hall. *The hidden dimension*. Anchor Books, New-York, NY, USA, 1969.

[45] R. Heck and M. Gleicher. Parametric motion graphs. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, I3D '07, pages 129–136, New York, NY, USA, 2007. ACM.

[46] D. Heylen, S. Kopp, S. C. Marsella, C. Pelachaud, and H. H. Vilhjálmsson. The next step towards a function markup language. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA, pages 270–280, Berlin, Heidelberg, 2008. Springer-Verlag.

[47] G. E. Hinton, S. Osindero, M. Welling, and Y. W. Teh. Unsupervised discovery of nonlinear structure using contrastive backpropagation. *Cognitive Science*, 30(4) :725–731, 2006.

[48] S. Hochreiter and J. Schmidhuber. LSTM can solve hard long time lag problems. In *Advances in Neural Information Processing Systems 9, NIPS, Denver, CO, USA, December 2-5, 1996*, pages 473–479, 1996.

[49] J. Huang and C. Pelachaud. An efficient energy transfer inverse kinematics solution. In M. Kallmann and K. Bekris, editors, *Motion in Games*, volume 7660 of *Lecture Notes in Computer Science*, pages 278–289. Springer Berlin Heidelberg, 2012.

[50] B. Hutchinson, L. Deng, and D. Yu. Tensor deep stacking networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8) :1944–1957, 2013.

[51] R. Kemp. *Embodied Acting : Cognitive Foundations of Performance*. PhD thesis, University of Pittsburgh, 2010.

[52] A. Kendon. *Conducting interaction : Pattern of behavior in focused encounter*. Cambridge University Press, 1990.

[53] M. Kipp. ANVIL - a generic annotation tool for multimodal dialogue. In *EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 3-7, 2001*, pages 1367–1370, 2001.

[54] M. Kipp. Annotation facilities for the reliable analysis of human motion. In N. C. C. Chair), K. Choukri, T. Declerck, M. U. Do ?an, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

[55] S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. H. Vilhjálmsson. Towards a common framework for multimodal generation : the behavior markup language. In *Proceedings of the 6th international conference on Intelligent Virtual Agents*, IVA, pages 205–217, Berlin, Heidelberg, 2006. Springer-Verlag.

[56] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Trans. Graph.*, 21(3) :473–482, July 2002.

[57] B. M. Lake, R. R. Salakhutdinov, and J. Tenenbaum. One-shot learning by inverting a compositional causal process. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2526–2534. Curran Associates, Inc., 2013.

[58] J. Lee, J. Chai, P. Reitsma, J. Hodgins, and N. Pollard. Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics, Proceedings of SIGGRAPH*, 2002.

[59] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard. Interactive control of avatars animated with human motion data. *ACM Trans. Graph.*, 21(3) :491–500, 2002.

[60] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard. Interactive control of avatars animated with human motion data. *ACM Trans. Graph.*, 21(3) :491–500, July 2002.

[61] J. Lee and S. Marsella. Modeling speaker behavior : A comparison of two approaches. In *Intelligent Virtual Agents - 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*, pages 161–174, 2012.

[62] S. Levine, P. Krähenbühl, S. Thrun, and V. Koltun. Gesture controllers. *ACM Trans. Graph.*, 29(4), 2010.

[63] S. Levine, P. Krahenbuhl, S. Thrun, and V. Koltun. Gesture controllers. *ACM Transactions on Graphics, Proceedings of SIGGRAPH*, 29(4), 2010.

[64] S. Levine, C. Theobalt, and V. Koltun. Real-time prosody-driven synthesis of body language. *ACM Trans. Graph.*, 28(5), 2009.

[65] S. Levine, J. M. Wang, A. Haraux, Z. Popovic, and V. Koltun. Continuous character control with low-dimensional embeddings. *ACM Trans. Graph.*, 31(4) :28, 2012.

[66] C.-K. Lim, M.-P. Cani, Q. Galvane, J. Pettré, and T. Abdullah Zawawi. Simulation of Past Life : Controlling Agent Behaviors from the Interactions between Ethnic Groups. In *Digital Heritage International Congress 2013*, Marseille, France, Oct. 2013.

[67] C.-Y. Lin, L.-C. Cheng, C.-C. Huang, L.-W. Chuang, W.-C. Teng, C.-H. Kuo, H.-Y. Gu, K.-L. Chung, and C.-S. Fahn. Versatile humanoid robots for theatrical performances. *International Journal of Advanced Robotic Systems*, 10(1), January 2013.

[68] Z. Ling, S. Kang, H. Zen, A. W. Senior, M. Schuster, X. Qian, H. M. Meng, and L. Deng. Deep learning for acoustic modeling in parametric speech generation : A systematic review of existing techniques and future trends. *IEEE Signal Process. Mag.*, 32(3) :35–52, 2015.

[69] T. D. Little and A. Ghafoor. Synchronization and storage models for multimedia objects. *IEEE J.Sel. A. Commun.*, 8(3) :413–427, Sept. 2006.

[70] C. Liu, A. Hertzmann, and Z.Popovic. Composition of complex optimal multi-character motions. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2006.

[71] P. Luo, M. Kipp, and M. Neff. Augmenting gesture animation with motion capture data to provide full-body engagement. In Z. Ruttkay, M. Kipp, A. Nijholt, and H. Vilhjlmsson, editors, *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pages 405–417. Springer Berlin Heidelberg, 2009.

[72] L. P. MagalhÃčes, A. B. Raposo, and I. L. Ricarte. Animation modeling with petri nets. *Computers & Graphics*, 22(6) :735 – 743, 1998.

[73] R. Maiocchi and B. Pernici. Directing an animated scene with autonomous actors. *The Visual Computer*, 6(6) :359–371, 1990.

[74] R. Marczak, M. Desainte-Catherine, and A. Allombert. Real-time temporal control of musical processes. In *International Conferences on Advances in Multimedia*, 2011.

[75] S. Marsella, Y. Xu, M. Lhommet, A. Feng, S. Scherer, and A. Shapiro. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '13, pages 25–35, New York, NY, USA, 2013. ACM.

[76] M. Mateas and A. Stern. A behavior language for story-based believable agents. *IEEE Intelligent Systems*, 17(4), 2002.

[77] J. McCann and N. S. Pollard. Responsive characters from motion fragments. *ACM Trans. Graph.*, 26(3) :6, 2007.

[78] A. Milliez, G. Noris, I. Baran, S. Coros, M.-P. Cani, M. Nitti, A. Marra, M. Gross, and R. W. Sumner. Hierarchical Motion Brushes for Animation Instancing. In *NPAR '14 - Workshop on Non-Photorealistic Animation and Rendering*, Proceedings of the Workshop on Non-Photorealistic Animation and Rendering, pages 71–79, Vancouver, Canada, Aug. 2014. ACM New York.

[79] H. Mitake, K. Asano, T. Aoki, S. Marc, M. Sato, and S. Hasegawa. Physics-driven Multi Dimensional Keyframe Animation for Artist-directable Interactive Character. *Computer Graphics Forum*, 2009.

[80] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta : an interactive expressive ECA system. In *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 2*, pages 1399–1400, 2009.

[81] R. Niewiadomski, M. Obaid, E. Bevacqua, J. Looser, L. Q. Anh, and C. Pelachaud. Cross-media agent platform. In *Proceedings of the 16th International Conference on 3D Web Technology*, Web3D, pages 11–19, New York, NY, USA, 2011. ACM.

[82] M. Ochs, K. Prepin, and C. Pelachaud. From emotions to interpersonal stances : Multi-level analysis of smiling virtual characters. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 258–263. IEEE, 2013.

[83] W. Pan and L. Torresani. Unsupervised hierarchical modeling of locomotion styles. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 785–792, New York, NY, USA, 2009. ACM.

[84] C. Pinhanez. The scd architecture and its use in the design of story-driven interactive spaces. In *Managing Interactions in Smart Environments*, 2000.

[85] I. Poggi. *Mind, hands, face and body : a goal and belief view of multimodal communication.* Weidler, Berlin, 2007.

[86] R. Prada and A. Paiva. Believable groups of synthetic characters. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-agent Systems, AAMAS ?05*, pages 37–43, New-York, NY, USA, 2005.

[87] K. Prepin, M. Ochs, and C. Pelachaud. Beyond backchannels : co-construction of dyadic stancce by reciprocal reinforcement of smiles between virtual agents. In *International Conference CogSci (Annual Conference of the Cognitive Science Society)*, 2013.

[88] M. Radenen and T. Artières. Contextual markovian models. *Pattern Recognition Letters*, 35 :236–245, 2014.

[89] B. Ravenet, A. Cafaro, M. Ochs, and C. Pelachaud. Interpersonal attitude of a speaking agent in simulated group conversations. In *Proceedings of Intelligent Virtual Agents conference IVA ?14*, pages 345–349, Boston, MA, USA, 2014.

[90] M. Rehm and B. Endrass. Rapid prototyping of social group dynamics in multi-agent systems. *AI and Society*, 24 :13–23, 2009.

[91] C. Reynolds. Steering behaviors for autonomous characters. In *Proceedings of the Game Developers Conference*, pages 763–782, Miller Freeman Game Groups, San Francisco, CA.

[92] R. Ronfard. Notation et reconnaissance des actions scéniques par ordinateur. In M. Martinez, S. Proust, and M. Pouget, editors, *Notation du travail théâtral, du manuscript au numérique*. Lansman, Dec. 2012.

[93] H. Sacks, E. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50 :696–735, 1974.

[94] C. Salter. *Entangled*. MIT press, 2010.

[95] E. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 21(1) :1–63, 1974.

[96] A. Shapiro. Building a character animation system. In J. Allbeck and P. Faloutsos, editors, *Motion in Games*, volume 7060 of *Lecture Notes in Computer Science*, pages 98–109. Springer Berlin / Heidelberg, 2011.

[97] A. Shoulson, N. Marshak, M. Kapadia, and N. I. Badler. Adapt : the agent development and prototyping testbed. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '13, pages 9–18, New York, NY, USA, 2013. ACM.

[98] A. Shoulson, N. Marshak, M. Kapadia, and N. I. Badler. ADAPT : the agent developmentand prototyping testbed. *IEEE Trans. Vis. Comput. Graph.*, 20(7) :1035–1047, 2014.

[99] U. Spierling, D. Grasbon, N. Braun, and I. Iurgel. Setting the scene : playing digital director in interactive storytelling and creation. *Computers & Graphics*, 26(1) :31–44, 2002.

[100] M. Sreenivasa, P. Souères, J.-P. Laumond, and A. Berthoz. Steering a humanoïd robot by its head. In *iros09*, St Louis (MO), USA, October 2009.

[101] J. Tanenbaum, M. S. El-Nasr, and M. Nixon. *Nonverbal Communication in Virtual Worlds : Understanding and Designing Expressive Characters*. ETC Press, 2014.

[102] M. Thiebaux, S. Marsella, A. N. Marshall, and M. Kallmann. Smartbody : Behavior realization for embodied conversational agents. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '08, pages 151–158, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[103] M. Toro-Bermudez, M. Desainte-Catherine, and J. Castet. An extension of interactive scores for multimedia scenarios with temporal relations for micro and macro controls. In *Sound and Music Computing*, 2012.

[104] A. Treuille, Y. Lee, and Z. Popovic. Near-optimal character animation with continuous control. *ACM Trans. Graph.*, 26(3) :7, 2007.

[105] L. Unzueta, M. Peinado, R. Boulic, and A. Suescun. Full-body performance animation with sequential inverse kinematics. *Graph. Models*, 70 :87–104, September 2008.

[106] R. Vaillant, L. Barthe, G. Guennebaud, M.-P. Cani, D. Rohmer, B. Wyvill, O. Gourmel, and M. Paulin. Implicit Skinning : Real-Time Skin Deformation with Contact Modeling. *ACM Transactions on Graphics*, 32(4) :Article No. 125, July 2013. SIGGRAPH 2013 Conference Proceedings.

[107] D. Van Rijsselbergen, B. Van De Keer, M. Verwaest, E. Mannens, and R. Van de Walle. Movie script markup language. In *Proceedings of the 9th ACM Symposium on Document Engineering*, DocEng '09, pages 161–170, New York, NY, USA, 2009. ACM.

[108] H. H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf. The behavior markup language : Recent developments and challenges. In *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, pages 99–111, 2007.

[109] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2) :283–298, 2008.

[110] T.-S. Wang, N.-N. Zheng, Y. Li, Y.-Q. Xu, and H.-Y. Shum. Learning kernel-based hmms for dynamic sequence synthesis. *Graph. Models*, 65(4) :206–221, July 2003.

[111] D. Weinland, E. Boyer, and R. Ronfard. Action Recognition from Arbitrary Views using 3D Exemplars. In *ICCV 2007 - 11th IEEE International Conference on Computer Vision*, pages 1–7, Rio de Janeiro, Brazil, Oct. 2007. IEEE.

[112] D. Weinland, R. Ronfard, and E. Boyer. Automatic Discovery of Action Taxonomies from Multiple Views. In A. Fitzgibbon, C. J. Taylor, and Y. LeCun, editors, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pages 1639–1645, New York, United States, June 2006. IEEE Computer Society.

[113] A. P. Witkin and Z. Popovic. Motion warping. In *SIGGRAPH*, pages 105–108, 1995.

[114] J. Yuan. Local svd inverse of robot jacobians. *Robotica*, 19(1) :79–86, Jan. 2001.

[115] J. Zhao and N. I. Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Trans. Graph.*, 13 :313–336, October 1994.

[116] S. Zhong, Y. Liu, and Y. Liu. Bilinear deep learning for image classification. In *Proceedings of the 19th International Conference on Multimedia 2011, Scottsdale, AZ, USA, November 28 - December 1, 2011*, pages 343–352, 2011.