Directing Autonomous Digital Actors

26 avril 2015

Table des matières

1 Résumé de la proposition de projet / Executive summary

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) ; statistical models learned from motion capture examples ; physically-based animation ; and speech-driven animation. 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 make decisions based on AI models of their personality and goals. In contrast, autonomous actors follow a precise script, written by the director, 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 research 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 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 proxemic 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 [86]. 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 attitudes 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. In contrast to previous works, which used programming languages, 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.

All developments will be validated by experiments with the theatre department of Paris 8. 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.

2 Introduction et positionnement / Introduction and positioning

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

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 believable virtual tutors and coaches in educational serious games, and creating believable characters for interactive fiction and interactive drama. 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 [10] ; statistical models learned from motion capture examples [45] ; physically-based animation [53] ; and speech-driven animation [15]. 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.

2.2 Positionnement du projet / Position of the project

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.

2.3 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 [38]. 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. [86]. The promises and limitations of virtual actors in contemporary theatre has been surveyed by Dixon [19], Salter [78] and Bourassa and Poissant [7].

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 [82].

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 animation techniques based on concatenation of motion capture data and clips selection [45, 42, 32] 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. [22] 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.

Many systems have been proposed for producing animation models and controlers using statistical models such as Hidden Markov Models (HMMs) [48] and Conditional Random Fields (CRFs) [47, 13]. 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 [97]. Locomotion controllers have been proposed that concatenate motion clips from a motion capture dataset to produce an animation that is smooth [89, 61]. 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 [45].

Style Machines [8] 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 [95] [66], 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.

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 [48, 47, 13, 16]. Finally, recent work has demonstrated such models for the case of locomotion believable controllers, gesture controllers [47] and face controllers [18].

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 [49], 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 [49, 13]. A relevant technology for this are Gaussian Process which have been extended for dealing with dynamic data in [94].

These latter models are not far from recurrent neural networks, and to Long Short Term Memory neural networks in particular [34], that have been shown recently to work well for complex signals such as speech and handwriting, for recognition tasks [29] as well as synthesis tasks [28]. 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 [33].

2.3.2 Multiple-character animation

ECAs gathering in groups : Prada and Paiva [70] 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 [74] implemented a toolbox for modeling the behavior of multi-agent systems. In [73], Ravenet and colleagues combined a number of reactive social behaviors, including those reflecting personal space [31] and the F-formation system [39], in a general steering framework inspired by [75]. 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) have been used to generate animation sequences procedurally [6, 90]. Different constraints and unconstrained numerical equation solvers are used to model the motion postures and their transitions [99]. The target based reaching models, such as Jacobian methods [98] [9], 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.

There exist several systems embedded in large platform that combine different animation synthesis solutions. The Smartbody system [81] [87] includes a schedule controller and a realization controller. Smartbody includes different motion synthesis methods such as motion capture data and dynamic system. The behavior engine developed by Marsella et al. [59] 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. [54] proposed a procedural arm gesture model. They improve the quality of the procedural animation by introducing motion capture data for full body animation. ADAPT [82] 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.

Existing animation techniques, procedural and data-driven, have pros and cons. With the procedural animation, gestures are described using the symbolic language BML [41]. 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 in DADA will be 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 1.2). 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 (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 by Strassmann [85], the Improv system by Perlin and Goldberg [68] and the story-driven architecture by Pinhanez [69] are seminal works in building virtual actors that can be scripted to perform on a virtual stage. Other early work in Europe includes the Pinnochio [57] and GEIST [83] projects, which targetted players of interactive games by letting them "play director". The acting capabilities of those early systems are limited and the amount of programming needed to produce even the simplest scene is substantial.

Xtranormal Technologies Text-To-Movie was an online application for quickly creating 3D animation by writing dialogues with emoticons, which were automatically translated into animation using a combination of procedural and data-driven animation. The DRAMA project at the University of Toulouse has also investigated this paradigm. While automatic text-to-scene animation is a valid research area, it has the disadvantage that it is taking control away from the director. In DADA, we take a different route, which is offer maximum control to the director over the virtual actors.

MIRAGE [20] 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 Q language [37] 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 the actor's full body animation.

The movie script markup language (MSML) [92] 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) [52] 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 [56, 5], including the important case of turn-taking in dialogue and imitation games [12, 11]. They are also the basis for the interactive musical score system developed by the INEDIT project [2, 58], which is being extended to multimedia events [88]. 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.

2.3.4 Robotic actors

Our project is also indirectly related to the emerging field of robotic actors performing theatre on a live stage [50]. 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.

2.3.5 Prior art at LTCI

Our developments will be based on the GRETA platform and use the EMILYA database. 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 [65, 64]. It is a SAIBA compliant architecture (SAIBA is a common framework for the autonomous generation of multimodal communicative behavior in Embodied conversational agents [41]). 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.

Data-driven animation in DADA will use the Emilya database (Emotional body expression in daily actions database) [24]. Eleven (unprofessional) actors participated in the data collection. Both 3D motion capture data (using Xsens technology [2]) 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) [24]. 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 studies.

2.3.6 Prior art at Inria

Previous work in the IMAGINE team can be useful to the DADA project, including work on sketch-based animation [30] ; implicit character skinning [91] ; and audio-visual prosody modeling for expressive facial animations of actors [4]. All this previous work will be made available to the DADA project.

1. Robotics and Performing Arts : Reciprocal Influences, http://www.robotics-and-performing-arts.sssup.it/
2. http://www.xsens.com

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

The fundamental research program proposed in the DADA project targets the grand challenge of getting virtual actors to perform a theatre scene together. This has never been demonstrated before. DADA is therefore a high-risk, high-gain project, with the ambition to significantly raise the level of believability and expressivity of autonomous conversational agents by endowing them with acting skills. Working with theatre directors and scholars will raise the bar of what is expected of virtual actors created in DADA both in terms of aesthetic quality and authoring tool usability. By leveraging existing animation techniques developed previously by the DADA partners in their respective fields, the DADA project will be starting from a solid position, but this will not be sufficient to meet the theatre world's expectations.

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 LTCI for the GRETA platform and by the IMAGINE team at Inria for virtual cinematography. 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.

To confront this grand challenge, the first main bottleneck is the very large dimension of the parameter space of character animation (50-80 degrees of freedom per actor). We will therefore 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 their proxemic relations. The second main bottleneck is the lack of functional-level description language for non-verbal communication, and we will adapt working practises from the theatre world to propose such a description language in the form of a dramatic score notation .

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

The DADA project includes fundamental research in multiple disciplines, focusing on a single object of study - the virtual actor. 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.

WP1 focuses on animation models for isolated actors. The inputs that are used by the methods to be developed in this work package are procedural animation scenario as output by WP2. Such a scenario includes in particular detailed indications on the action to be realized (walk from one point to another, carry object, knock on door, throw object, lift object, move object), the mood of the character (neutral, happy, afraid, angry, anxious, sad, proud, shameful) and a set of static information about the character that change the way people move (age, gender, morphology, corpulence, expressivity level, etc). Both action and mood may vary with time among the animation while static information remain fixed per nature. These three sets of information will be refered hereafter as *action context*, *mood context* and *profile context*.

WP2 focuses on animation models for groups of actors. Based on the dramatic score composed by the director, patterns of actions and reactions of the virtual actors will be computed, taking into account the social relations between their characters, and the physical constraints imposed by the stage. This work package will be responsible for synchronising the behaviours and the movements of the virtual actors between cues given by the director. This work package will compute only a small number of degrees of freedom of the virtual actors, and provide the spatial and temporal context in which the full body animation is computed in details.

WP3 focuses on the control of the virtual actors by the director. Research work will be conducted to define a consistent and expressive language that the director can use to compose the dramatic score for the actors. Novel authoring tools will be designed and developed to support the language with intuitive and efficient graphical user interfaces. This work package will also be responsible for executing the animations produced by WP1 and WP2 in real-time using state-of-the-art game technologies present in the Unity 5 game engine.

WP4 focuses on the user evaluation of the proposed technologies from the perspective of a theatre director. Exercises for one to three actors will be developed to test the capabilities of the system. Example scenes of theatre plays will be chosen and staged for virtual actors. The usability of the system and the quality of the animation will be evaluated with a number of real-life scenarios, including virtual rehearsals for real-life productions, production of theatre performances for virtual worlds and teaching of theatre techniques.

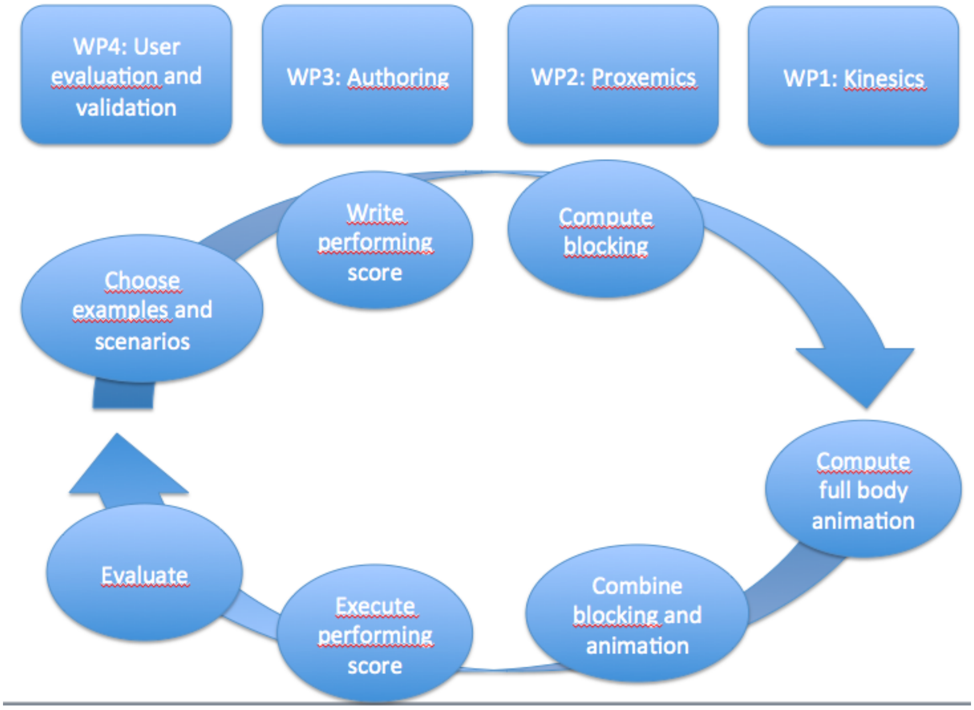

FIGURE 1 – Data flow between work packages : Through the authoring tool (WP3), a script is elaborated by a theater director (WP4) ; it gives direction to group of virtual actors which act out autonomously the commands of the script to position toward each other and in the virtual space (WP2). The behaviors 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's body movements from annotated data (e.g. motion capture 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 [72, 15, 17] 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 [1]). These frameworks are part of the representation learning line of work which brought impressive breakthrough in few machine learning tasks for signals, speech recognition [51]ăand computer vision tasks [44, 80].

Finally we plan to explore fully new alternative strategies such as using neuro muscular based models following ideas of deltalognormal models from [23]. Such a modeling allows recovering the sequence of neuromuscular commands that generated a handwritten gesture. Although these models led to impressive results in terms of modeling accuracy, these models have been applied to very simple gestures only and their ability to handle complex gestures is not certain. Yet if we were able to make them robust enough to provide valuable interprettation of complex motion, these models would bring in our opinion a very promising line of research for building new and natural motion controllers.

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 [18, 17], a video-based animation model [4], and the procedural animation model in the GRETA system [63]. 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 ony been adressed in few studies for simple gestures like handwriting signals [3] [43]. We will go beyond these preliminary studies and will explore few ways that aim at favoring transfer from learning one action / mood model to learning another action / mood model. In particular we will explore strategies for modifying our generic controlers, e.g. based on recurrent neural networks or on contextual markovian models [72], in order to enable transfer learning between action models so that a new action model can be learned with only few samples of this action.

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 positioning and orientation of the ECAs [67]. Few researches have looked at modeling group of ECAs with different personalities and social attitudes [27, 70]. 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 dominance [96]. We will extend group behavior model [67] that embeds the F-Formation proposed by Kendon [40] 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 continuously in relation to the others' behaviors. To simulate the dynamic evolution of these behaviors we will make use of Neural Network simulation [71] 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 [71] ensuring not only the synchronization between actors' behaviors but also their mutual influence. This task will be led by LTCI with the contribution of Inria and LIF.

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 [26, 76]. 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 vocabulary of high-level actor behaviors, suitable for generating complex performances. Following a methodology introduced in interactive drama [60], 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

The project will be coordinated by Inria with a collegial decision-making process. A steering committee composed of Thierry Artières (LIF), Georges Gagneré (Paris 8), Catherine Pelachaud (LTCI) and Rémi Ronfard (Inria) will make all important decisions in unanimity. The committee will meet before every consortium meeting and its decisions will be communicated to all consortium members and to ANR. Rémi Ronfard will be the project coordinator and will be responsible for applying the decisions of the steering committee and for communicating with ANR and other consortium members.

Consortium meetings will be organized four times a year for a total of 14 meetings, including kick-off meeting and final review meeting. Meetings will take place in the theatre department at Paris 8 to ensure continuity and visibility of the project in the theatre arts world. Paris is also a convenient meeting place for all project members. Ensuring regular meetings with all researchers involved in the project is an important factor for success, given the widely multidisciplinary aspect of the project. This will also favour joint work by the PhD students working on the project, and ensure their commitment to the user evaluation phase that will take place in Paris 8 as well.

Software development will be coordinated by Inria and Paris 8 using rapid prototyping methods (SCRUM) [3]. The project coordinator, Rémi Ronfard, will act as the "Scrum Master", taking responsibility for the delivery of software increments during fast-paced development periods (including, but not limited to the project milestones). Cédric Plessiet will act as the "Product Owner", taking responsibility for the specifications and selection of features included in each increment, for the deployment of all software increments at Paris 8, for their testing and user evaluations, and for filling bug reports and user documentations.

All PHD Students will be asked to contribute their latest results to be included in the DADA prototype at least twice a year. The rest of their time will be devoted to their research work. Source code files will be signed by all contributing authors, together with their affiliations, to properly track intellectual property rights.

3.3 Description des travaux par tâche / Description by task

3.3.1 WP1 Kinesics

WP1	Kinesic component
Responsable	LIF
Participants	Inria, LTCI
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 ($T_0 \rightarrow T_0 + 36$)
Task12	Face and gesture animation ($T_0 + 12 \rightarrow T_0 + 36$)
Task13	Learning from few samples ($T_0 + 18 \rightarrow T0 + 42$)

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 The first task focuses on the design of original generic body controllers for full body animation of a character given a 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 generic models whose behaviour depends on few contextual variables (e.g. mood components and actor's profile components) viewed as few inputs which influence the animation. We will explore three lines of research that we present below ordered by decreasing complexity (for us to make these work) and, as it might occur, of increasing potential. While the first one is the extension of our previous works, the two others are original lines of research for body controllers.

We will start with an approach that we know better and that we are confident that it will provide interesting results. We will extend *contextual markovian models* whose parameters (means of Gaussian

3. SCRUM is a popular and established technique in the game industry, which is applicable to DADA. SCRUM facilitates agile software development with a small team of part-time developers and a dynamic team organization

distribution, transition probabilities...), are parameterized by (i.e. defined as a function of) contextual variables. One such 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 [47] and face controllers [72, 15, 17]. 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.

Next, we will investigate the design 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 signals such as speech or handwriting [28, 1, 51]. These models are related to what is called *representation learning* and *deep learning* which emerged in the last few years as a key topic in the machine learning community [4] [14]. 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 the weights of a layer could be defined as a function of the contextual input, inspired by works like [100, 36].

At last we will investigate using *low dimensional state space models* and in particular neuro muscular based models following ideas like [23] 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. It is expected that working in that space will allow designing much more natural controllers. 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 [94]), 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 facial expressions in dialogue situations. We will start from three available face animation models in the consortium : a mocap based animation model [18, 17], a video-based animation model [4], and the procedural animation model in the GRETA system [63]. 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.

All of our three controllers have pros and cons. While statistically-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. Statistically-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 [46].

We will explore few ways to combine animation models which remains an open question today, be it for animating the face or the full body. We will implement our solutions within the Greta framework where communicative intentions and emotions are represented with the FLM language while multimodal behaviors with BML [93]. The merge of multiple animation models may then 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 [82]. 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 necessary animations produced by the other models.

Task 1.3. Learning from few samples This last task concerns strategies that we will design for learning our controllers from few samples. The idea is that while task 1 focuses on how to design accu-

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

rate and natural controllers without caring about the ability of these models to be learned in practice, when training samples are not numerous, this task aims at proposing strategies to slightly modify these controllers for making them learnable in realistic settings. We will mainly investigate two approches. 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 models for instance this means that instead of defining 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. One idea is to make transition probablities dependent on the action to perform so that an action model would correspond to 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. Preliminary works that we did let us expect that such a strategy would work with statistical markovian models [18, 72]. We will then extend this kind of ideas to recurrent neural networks and continuous state space models. Indeed in such models, a particular motion correspond to a dynamic behaviour in the latent space. Relying on a low dimensional state space (e.g. corresponding to the degrees of freedom of body poses or to the neuro muscular commands) will mechanically reduce the dynamic model complexity (e.g. number of parameters) and would favor learning from few samples.

Deliverables	Name and content	Date
L1.1	Report on the state of the art for statistical models for animation synthesis	$T_0 + 12$
L1.2	First version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 24$
L1.3	Second version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 36$

Partners' roles LIF is the responsible partner for this work package. Inria and LTCI will take part in the development of expressive text-to-speech technologies and audio-visual prosodyfor task 1.2, using their respective previous work in this areas. Inria and LTCI will also bring all required expertise in 3D animation to ensure the success of tasks 1.1 and 1.3 and we will be targeting joint research work to ensure that the resulting work can be published in major computer graphics conferences, as well as in machine learning conferences.

Risks The risks are limited with respect to the 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/) and the corpus EMILYA (EMotional body expressIon in daiLY Actions databaseBodily Emotional Actions Behavior) has been created by one of the partner (Fourati, 2014).

3.3.2 WP2 Proxemics

WP2	Proxemic component
Responsable	LTCI
Participants	Inria, LIF
Duration	42 months
Objectives	Design and development of low-dimensional, multi-character animation
Content	
Task 21	Communicative behaviours ($T_0 \to T_0 + 24$)
Task22	Steering behaviours ($T_0 + 12 \to T_0 + 36$)
Task23	Combination of statistical and procedural models ($T_0 + 24 \to T_0 + 36$)

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 4.1), 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 is listening, etc. We will extend an existing turn-taking model [73] that is based on Sack's model [77], that embeds F-Formation [39] 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 [79] 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 This task will model agent behaviors when moving around in the environment, included such advanced steering behaviors as follow, flee, separate, join, merge, enter stage, exit stage, etc. 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 4.1). 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 other's positions and orientations 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 agent behaviors, we will make use of Neural Network simulation [71] 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 [71] ensuring not only the synchronization between actor 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 [35]. 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	Name and content	Date
L2.1	Report on the state of the art of proxemics models in computer animation	$T_0 + 12$
L2.2	First version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 24$
L2.3	Second version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 36$

Partner's roles This work package will be coordinated by LTCI. Inria will contribute to tasks 2.2 and 2.3. LIF will contribute to task 2.3.

3.3.3 WP3 Authoring

WP3	Authoring
Responsable	Inria
Participants	Paris 8, LIF, LTCI
Duration	36 months
Objectives	Authoring and execution of virtual actor performances
Content	
Task 31	Specification of the dramatic score language ($T_0 \rightarrow T_0 + 12$)
Task12	Authoring tools ($T_0 + 12 \rightarrow T_0 + 36$)
Task13	Real-time animation ($T_0 + 12 \rightarrow T_0 + 36$)

Task 3.1 : Specification of a dramatic score language Previous work in autonomous agents has focused on defining non-verbal communication languages at the functional level (what is being communicated) and the behavioral level (how it is communicated using body language). While a consensus has slowly been reached on the behavioral level with the BML language [41], a general framework for describing non-verbal communication at the functional level remains elusive. In DADA, we will take a very different route by giving specifications only for dramaturgic actions, i.e. actions whose purpose is to be played on a stage and communicated to an audience. This will include a choice of verbs (actions, speech acts, movements) and adverbs (moods, attitudes, dramatic effects) for directing actors ; and a choice of cues acting as synchronization points between actors.

The research problems to be solved in this task include the definition of parallel and sequential behaviors for single, isolated actors ; the definition of parallel and sequential behaviors for groups of actors ; an intuitive and general syntax for control the synchronization between them ; the choice of non redundant action primitives ; and the choice of meaning parameters to be exposed to the director (as opposed to resolved internally by the animation engine). In agreement with the organization of tasks in WP1 and WP2, the specification will clearly separate proxemic and kinesic elements of actor behaviors. Furthermore, our goal will be to cover both the movements of actors on the stage, and their communicative behaviors during dialogue.

As a result of this task, we will be proposing an extended notation for theatre performances, akin to a musical score [26, 76, 25]. For lack of another word, we will refer to this future notation scheme as the *dramatic score*. To define the syntax and semantic of the dramatic score, we will draw on previous work on the dramaturgical score [21], which takes a semiotics approach, and on established stage management practices in the theatre [55]. The proposed specification will be given in the form of a pseudo-natural language that makes sense to the director, and can at the same time be compiled into a machine representation. Internally, the representation will likely be based on Petri nets, which provides an intuitive and general representation of multi-modal, parallel processes with complex synchronization cues ; and at the same time can easily be resolved into hierarchical finite state machines, which can be executed efficiently in a game engine.

The specification will borrow important terms from the language of theatre [55] : blockings, cues and prompts, which we explain briefly now. 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. 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 is the master copy of the script or score, containing all the actor moves and technical cues, and is used by the stage manager to run rehearsals and later, control the performance. Such traditional notions will need to be thoroughly extended to take into account the entirely novel situation of computer-controlled virtual actors.

The specification will be delivered in two increments. The first increment will focus on directing individual actors and synchronizing their actions. The second increment will focus on directing groups of actors and synchronizing their actions within the group, and across different groups.

Task3.2 : Authoring tools For the DADA platform to be accepted by theatre directors, it is important that we offer authoring tools that make sense to them. Although we are targeting Petri nets as an internal representation, our goal in task 32 will be to hide this internal representation from users. Instead, we we will offer three modes of interaction : didascalia, storyboard sketches, and timelines. Didascalia will be offered as blocks of text with a controled vocabulary and a choice of parameters allowing the director to define cues, actions and their parameters in plain text. Storyboard sketches will be used to represent the actors positions, orientations and trajectories using a combination of floor-plan views and audience views. Timelines will be used to set the starting and ending points of performing events by direct manipulation. There will likely be one timeline per actor per animation component (proxemic behaviors, kinesic actions, kinesic moods, speech acts, etc.).

As a result, the authoring tools will 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, actions and timing.

Task 3.3 : Real-time execution of the dramatic score While task 31 focuses on defining the vocabulary and syntax of the dramatic score, and task 32 focuses on implementing a score-writing application, the role of task 33 will be to implement a realizer/synthesizer for executing 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.

Another challenge to be overcome is in combining full body animation and interaction animation at runtime. We believe it will be an important 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.

Special care will be needed for combining the proxemic and kinesic components of animation at runtime along the lines of Mitake et al. [62], 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.

Another research issue that will be investigated in this task is the possibility of controlling the proxemic components of character animation using the rigid motion of the head, rather than the full body. Sreenivasa et al. [84] 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.).

Deliverables	Name and content	Date
L3.1	Specification of the dramatic score language	$T_0 + 12$
L3.2	First version of authoring tools and execution environment : Prototype (software) and its documentation	$T_0 + 24$
L3.3	Second version of authoring tools and execution environment : Prototype (software) and its documentation	$T_0 + 36$

Partners' roles Inria will be the coordinator and main software developer for this work package. Task 3.1 will be jointly performed by Inria, LTCI 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 and runtime environment. LTCI will contribute to task 3.3 by providing a subset of the GRETA platform.

Risks This ambition in this task is potentially very high. We will devote special attention to ensure that the first prototype on single-actor animation uses established and existing technologies from all partners, and can be shared by all partners at an early stage in the project. This will make it easier to maintain control of the potential more risky second prototype. The role of Paris 8 as a product owner will be crucial in helping the consortium maintain a functional and tested code base over the duration of the project. As a result, the risks will be limited.

3.3.4 WP4 User evaluation and validation

WP4	User evaluation and validation
Responsable	Paris 8
Participants	LIF, Inria, LTCI
Duration	42 months
Objectives	Deployment, testing and evaluation of DADA prototypes
Content	
Task 4.1	Exercises and scenarios ($T_0 \rightarrow T_0 + 12$)
Task 4.2	User evaluation of interaction with a single virtual actor ($T_0 \rightarrow T_0 + 30$)
Task 4.3	User evaluation of interaction with multiple virtual actors ($T_0 \rightarrow T_0 + 42$)

Task 4.1 : Exercices and rehearsal scenarios Based on the theatrical methodology exposed in task 3.1, we'll design short exercices inspired by the play entittled "L'augmentation" by Georges Perec. Writen in the Oulipo style, the play declines a large number of variation developing the recurrent theme of an employee asking an augmentation from his boss, with the assistance of his secretary. Basic actions are used in multiple ways to express subtle strategies for convincing the boss. We will find the proper vocabulary to describe and parameterize those basic actions, taken as exercices that could be done by a real actor, and that will be perform by a virtual agent under the direction of the real stage director.

The catalogue of basic actions will be realized at first with one actor to test the efficiency of the authoring tool. The evaluation will be done by digital artist team (INREV-ATI) and theater team (Scènes du monde, création et savoir critiques) on two levels. On the one hand, we will evaluate the ergonomy and usability of the tools to reach proper results, and the quality of the virtual rendering, both on visual aspect and narrative coherence. On the other hand, we will precisely measure the ambitus of espressivness we could ask to the virtual actor, and the possibilities of combining and finding emergent solutions to artistic intuitions in a creative way.

We will start with simple exercises, as soon as the first deliverables of WP3 will be ready, in order to orient properly the complexity of the score, the writing tool and the virtual behaviors. We will increase step by step the difficulty of the exercices in both movement ability and emotionnal espressivity. The exercices will also be conceived as a way to learn how to use the tool for directing virtual actors. We will also control the necessity to keep basic rules of behavior when complexity of exercices increase, i. e. the stability of the directing process. This is a specific quality of working with a troup of actors : a common gestual and emotionnal vocabulary shared by actors, directors and audience helps to bring strong creative proposals.

Following the development of the authoring tool, we will develop new exercises focusing on interactions between two and three actors, in the continuity of the collection produced in 4.1 and the specifications developed in WP1 and WP2, and more specifically in 3.2.

Task 4.2 : User evaluation of interaction with a single virtual actor Based on the first prototype, and the exercises for one actor, we will combine artistic digital approach with theatrical knowledge, in order to conduct the development of the authoring tool in a proper direction. We will give precise evaluation of the needs in the rendering and in the ergonomy to respect a creative process and to reach espressive result. For instance, are the dramatic language, the performance score, the stage floor plan sketching tool adequate, useful, efficient ? Short extracts of L'augmentation will be chosen by Georges Gagneré and virtually staged with the authoring tools with the help of a postdoc researcher for reporting the bugs, precising the directing needs both ergonomically and creatively in connexion with WP3, and documenting the authoring tool for novice users. Those first results about very short performances will be submitted to a wide range of researchers and professionals of different artistic and cultural fields to evaluate the quality of the theatrical interactions both between virtual actors, and between virtual actors and real spectators (animation, video game and theater)

Task 4.3 : User evaluation of interaction with multiple virtual actors A postdoc researcher, specialized in actor directing, will learn "How to use the authoring tool ?" and complete the documentation process for students and professional artists. Together with Georges Gagneré, he will write the performance score for entire scenes of "L'augmentation". Two different users points of view on the authoring tool will be confronted, and will help to design the proper strategy for a broader dissemination. The Inrev-ATI and theater teams will also organize meetings to introduce the tool to multiple artistic communities, in the very creative context of the Laboratory of Excellence in Arts and Human Mediations (Labex ARTS H2H) They will organise workshops to learn and practice the authoring tool especially for his own creative purpose. The postdoc researcher will assist professional stage directors to realize virtual direction of short extracts from other pieces of the theatrical repertoire and confront the tool to different styles of actor direction. Explicitation interviews will be conducted with users of the DADA prototype to analyze the technological constraints that could limit the creative approach.

Using the collections of exercices produced in 4.1 and 4.2, a specific approach of the authoring tool around pedagogical issues will also be dedicated to master students of the IDEFI CREATIC innovative training program, including Paris 8 University, West Paris Nanterre-La Défense University, the Maison des Sciences Humaines Paris Nord (North Paris Centre for Human Sciences), the Conservatoire National Supérieur d'Art Dramatique (National Drama Academy), the National Archives and 37 foreign partners.

Finally, we will also evaluate the quality of the animation produced with the authoring tool with a variety of audiences.

Deliverables	Name and content	Date
L4.1	Selection of example scenes and examples : collection of one actor directing exercices allowing an ergonomic appropriation of the authoring tool and its animation possibilities, in a theatrical creativity context.	$T0 + 12$
L4.2	Preliminary revaluation report : Animated scenes with one actor, on the exercises defined in task 4.1. Results of user study on the subjective quality of the authoring tools and the produced animation. Collection of three-actor exercises allowing an ergonomic appropriation of the authoring tool and its animation possibilities in a theatrical creativity context. Documentation of the authoring tool for theatre directors.	$T0 + 30$
L4.3	Final evaluation report : Animated scenes with three virtual actors, showing exercises and selected examples from "L'augmentation" by Perec. User documentation of the authoring tools from a pedagogical context perspective. Results of user studies on subjective evaluations of the tools and produced animations in the professional and the academic fields.	$T0 + 42$

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

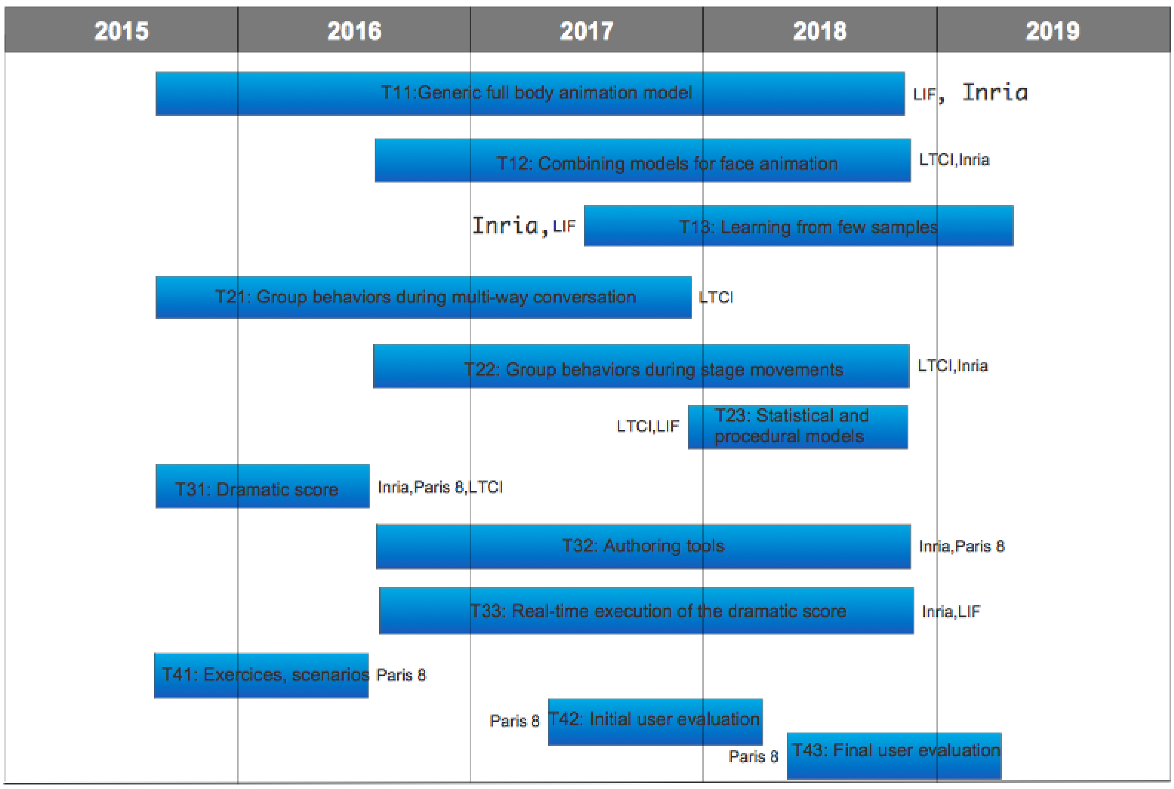

Figure 2 – GANTT diagram of tasks and resources for the DADA project.

3.4.1 Milestones

The project will be organized in four main phases, separated by three milestones at $T_0 + 12$, $T_0 + 24$ and $T_0 + 36$:

Phase 1 ($T_0 \to T_0 + 12$)	Specifications and scenarios.
Milestone 1 ($T_0 + 12$)	Delivery of specifications and scenarios.
Phase 2 ($T_0 + 12 \to T_0 + 24$)	Research and development for first prototype.
Milestone 2 ($T_0 + 24$)	Delivery of first prototype.
Phase 3 ($T_0 + 24 \to T_0 + 36$)	Research and development for second protoype.
Milestone 3 ($T_0 + 36$)	Delivery of second prototype.
Phase 4 ($T_0 + 36 \to T_0 + 42$	Final user evaluations.

3.4.2 Deliverables

The project includes 12 deliverables which will be delivered at the three milestones.

Number	Title	Date	Leader
L1.1	Report on the state of the art for statistical models for animation synthesis	$T_0 + 12$	LIF
L1.2	First version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 24$	LIF
L1.3	Second version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 36$	LIF
L2.1	Report on the state of the art of proxemics models in computer animation	$T_0 + 12$	LTCI
L2.2	First version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 24$	LTCI
L2.3	Second version of the models : Prototype (software) and its documentation (Report on the models developed)	$T_0 + 36$	LTCI
L3.1	Specification of the dramatic score language	$T_0 + 12$	Inria
L3.2	First version of authoring tools and execution environment : Prototype (software) and its documentation	$T_0 + 24$	Inria
L3.3	Second version of authoring tools and execution environment : Prototype (software) and its documentation	$T_0 + 36$	Inria
L4.1	Selection of example scenes and examples : collection of one actor directing exercices	$T0 + 12$	Paris 8
L4.2	Preliminary revaluation report : Animated scenes with one actor, on the exercises defined in task 4.1.	$T0 + 30$	Paris 8
L4.3	Final evaluation report : Animated scenes with three virtual actors, showing exercises and selected examples from "L'augmentation" by Perec.	$T0 + 42$	Paris 8

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

The goal of the DADA project is to establish a long-term collaboration between the four academic partners on the topic of virtual actors. This will be consolidated in a consortium agreement allowing the sharing of code and assets developed before and during the project.

The 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 invalidating the proposed patent. The partners can ask to be mentioned as co-inventors if they can demonstrate that they contributed to the invention.

The consortium is composed exclusively of academic partners, whose goal is primarily to disseminate and publish their research work. We will seek to publish our research results in top ranked conferences and journals, including SIGGRAPH, Eurographics, NIPS, ICML, IJCAI, AAAI, IVA, AAMAS, etc. In addition to scientific papers, the consortium will also seek to publish paper describing artistic applications of the DADA framework, possibly as demos, art papers and exhibits at SIGGRAPH, Ars Electronica, Laval Virtual and other similar venues.

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

We expect the consortium to be renewed and consolidated beyond the duration of the DADA project to support the DADA software at least as an open source, collaborative project with continued joint research and development.

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. LTCI and LIF are already working together on facial animation from speech through the co-supervision of Yu Ding's PHD thesis. Inria/Imagine and Paris 8 are also already working together on directing audiovisual prosody of actors, as part of Adéla Barbulescu's PhD thesis. Results of the two theses will be exploited in the project.

5.2 Qualification of the project coordinator

Remi Ronfard will be the co-ordinator of the DADA project. He is a computer scientist with a 20 year experience in industry and academia in France, Canada and USA. He received his PhD from Mines Paris Tech in 1991 and has worked at the T.J. Watson IBM Research Center in New York as post-doc and visiting scientist (1992 and 2000). He is the author of 60 scientific papers published in top ranked international journals (IJCV, CVIU, PAMI) and conferences (Siggraph, Eurographics, CVPR, ICCV, ECCV) and 4 international patents. He has conducted research projects in digital storyboarding (INA, 1995), aesthetic surface design (IBM Research and Dassault Systémes, 2000), video indexing (INA, 2000) and statistical analysis of film styles (INRIA 2002-2007). He has been an international expert in the MPEG group (1997-1999) and an R & D director at Xtranormal Technologies (2007-2009), where his team created the patented Magicam system. He has established active research collaborations with the national film school (ENS Louis Lumiére) and the Théâtre des Célestins 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) and intelligent cinematography and editing (Quebec, 2014 ; Zurich, 2015). He is currently serving as head of the Geometry and Image Department at Laboratoire Jean Kuntzmann, Univ. Grenoble Alpes.

5.3 Qualification and contribution of each partner

The complete list of permanent researchers involved in the DADA project is summarized in the table below.

Partner	Name	First name	Position	Field of research	PM	Contribution
Inria	Ronfard	Rémi	Researcher	Computer graphics and vision	12.6	WP1, WP2, WP3 (task leader)
Inria	Cani	Marie-Paule	Professor	Computer Graphics	3.6	WP3
Inria	Rohmer	Damien	Assistant Professor	Computer Graphics	8.4	WP3
Inria	Barbulescu	Adéla	Post-doc	Computer Graphics	12.6	WP1, WP2, WP3
LIF	Artières	Thierry	Professor	Machine Learning	14	WP1 (task leader), WP2, WP3
LIF	Emyia	Valentin	Assistant Professor	Machine Learning and Signal Processing	5	WP1
LIF	Qi	Wang	Ph.D. student	Machine Learning	12	WP1
Paris 8	Gagneré	Georges	Associate Professor	Theatre studies	10	WP3, WP4 (task leader)
Paris 8	Plessiet	Cédric	Associate Professor	Computer arts	12	WP3, WP4
Paris 8	Dusigne	Jean-Fran cois	Professor	Theatre studies	3	WP4
Paris 8	Légeret	Katia	Professor	Theatre studies	1	WP4
Paris 8	Moindrot	Isabelle	Professor	Theatre studies	1	WP4
Paris 8	Poirson	Martial	Professor	Theatre studies	2	WP4
LTCI	Pelachaud	Catherine	Senior Research	Conversational Agents	12	WP1, WP2 (task leader), WP3
LTCI	Clavel	Chloé	Assistant Prof.	Conversational Agents	3	WP2
LTCI	Ding	Yu	Post-doc	Conversational Agents	3	WP1, WP2

INRIA EPI IMAGINE IMAGINE is an INRIA research-project-team (EPI) of 6 permanent researchers created in 2012 by Marie-Paule Cani and Rémi Ronfard. IMAGINE stands for "Intuitive Modeling and Animation for Interactive Graphics & Narrative Environments". In adddition to Rémi Ronfard, two permanent researchers and a post-doctoral student will actively participate to the DADA project.

Marie-Paule Cani is a Professor of Computer Science at Grenoble University (Grenoble Institute of Technology & Inria), currently invited professor for 2014-2015 at College de France on the "Informatics and Computational Sciences" chair. Her research interests cover both Shape Modelling and Computer Animation. She contributed over the years to a number of high level models for shapes and motion such as implicit surfaces, multi-resolution physically-based animation and hybrid representations for real-time natural scenes and animated characters - with a specific interest for skin, clothes and hair. Following a long lasting interest for virtual sculpture, she has been recently searching for more efficient ways to create static and animated 3D content such as combining sketch-based interfaces with procedural models expressing a priori knowledge. She received the Eurographics outstanding technical contributions award in 2011 and a silver medal from CNRS in 2012.

Damien Rohmer is assistant professor at CPE Lyon and associate member of the IMAGINE team since 2011 after obtaining his PhD in Computer Science from Grenoble University. His research interests include cloth animation and character skinning. He has published papers in some major Computer

Graphics conferences and journals including ACM SIGGRAPH, Eurographics, Symposium on Computer Animation, and Pacific Graphics. He will contribute mostly to tasks 3.2 (authoring tools) and 3.3 (real-time character 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 task 1.2 with her recent work on visual prosody, which will be extended for joint generation of speech and facial animation from directorial input.

LIF Two main researchers and a Ph.D. student from the QARMA team of Computer Science Lab (LIF) at University of Aix-Marseille will participate to the project.

Thierry Artières is a professor at Ecole Centrale Marseille and member of the *QARMA team* (eQuipe AppRentissage et Multimédia) at LIF (Laboratoire d'Informatique 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. This work is 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).

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. 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 (TNS, La Filature, Scène Nationale de Mulhouse, Théatre 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 regularly with Stéphane Braunschweig and Peter Stein as stage director first assistant (La Scala, La Fenice, Théâtre des Champs-Elysées, Opéra Comique).

Cédric Plessiet is an associate professor at University Paris 8 where he teaches special fx game programming, Unity 3D programming and computer art. Before that, he has worked with special effects and motion capture companies, creating virtual semi-autonomous butterfly actors for the prize-winning movie "four wings and a pryer". His research interests include movie and theatre previz, real-time game engine programming and artificial intelligence for video games. His computer art projects have been featured at the Futur En Seine, Paris FX and IVRC conferences. He has worked with the international Buto dancer Atsuchi Takenuchi. His computer art work was presented in London, Rumania, Switzerland and Japan [5]. He will be responsible for deploying, testing, documenting and evaluating the DADA prototypes.

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.

Katia Légeret, director of the EA1573 team, "Scènes du monde, création et savoirs critiques" will help to integrate the resarch goals to the differents themes of the team, in a transdicisplinary way.

5. `http://www.mobilisimmobilis.com/`, `http://ivrc.net/2006/`, `http://www.symposium-pi.ch/`

Isabelle Moindrot is the director of the ARTS H2H Labex. She will help to the integration of DADA in the global artistic 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 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).

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 518 K€for three PhD students and two post-doctoral students, computer hardware and software, and travel expenses (50 K€). The project duration should be 42 months in order to develop a functional prototype in three years and to use it to animate several play scripts for evaluation and validation purposes.

6.1 Équipement / Equipment

LTCI request 6 500 € to cover equipments (High performance PC with good graphics card). LIF and Inria each request 5 K€ to cover equipments (work station with high quality video card enabling Graphical Processing Unit computation) and necessary softwares, and research books. The theater laboratory EA1573 asks for 4000 € for the purchase of a working station and all the softwares and hardwares necessary for the development of the WP4.

6.2 Personnel / Staff

LTCI requests a financial aid of 108 K€ for a PhD. The PhD topic will focus on modeling group behaviors of virtual agents while conversing or performing actions together (eg following each other). The Phd Student will also develop algorithm to merge procedural and statistical animations.

LIF requests a financial aid of 101 K€ for a PhD. His work will be focused on WP1 and more particularly on statistical machine learning for designing generic controlers (task 1.1 and 1.3). LIF requests an additional 6K €for hiring two 5 months internships in order to perform more exploratory studies.

Inria requests a financial aid of 117 K€ for a PhD. The PhD Topic will focus on a real-time synchronization model for virtual actors performing coordinated movements. The thesis will be devoted to designing and implementing the dramatic score language central to WP3 and applying it to create believable multi-character 3D animation in the Unity game engine.

The theater research team EA1573 asks for two Post-Doc of 12 months each (40k € per Post-Doc). In collaboration with the INREV team, the first post-doc will intervene on the WP4.2 to guarantee an effective communication between the WP3 and the WP4 but also to help in the software implementation of experiments, documentation by tutorial, sample and user's manual. The second post-doc will intervene on the WP4.3 to bring to a successful conclusion of the dramatic score, to build complex animation, and to help in the validation of the prototype whether it is from a qualitative and quantitative point of view with the spectators, the actors and the users of the software. It will also have for mission to finalize the user set of documents to guarantee the distribution of the system to other potential users (video games, performing arts, etc.)

6.3 Missions / Travel

LTCI, Inria and LIF each request a budget of 15K € to participate in project meetings, exchange PhD students over short periods of time, and participate in national and international conferences. Paris 8 requests 4000 € in spawn of mission to allow the researchers mobilized to exchange with the partner labs, but also to present their research in various colloquiums and festivals (Ars Electronica, Laval Virtual, Les bains numériques, SIGGRAPH Art Sessions). Theater laboratory asks for 4000 € to finance external expertises of the theater field to help in the validation of the deliverables.

6.4 Autres dépenses de fonctionnement / Other expenses

A small additional budget is requested by all partners to cover publication fees for open-access journals, software and research books supporting the work of PhD students, as well as internships on topics relevant to DADA during the three academic years covered by the project.

7 Bibliographie / References

[1] 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.

[2] 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.

[3] T. Artières, S. Marukatat, and P. Gallinari. Online handwritten shape recognition using segmental hidden markov models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2) :205–217, 2007.

[4] A. Barbulescu, R. Ronfard, G. Bailly, G. Gagneré, and H. Cakmak. Beyond basic emotions : Expressive virtual actors with social attitudes. In *ACM SIGGRAPH Conference on Motion in Games*, 2014.

[5] 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.

[6] 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.

[7] R. Bourassa and L. Poissant. *Avatars, personnages et acteurs virtuels*. Presse de l'Université du Quebec, 2013.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] 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.

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

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

[16] 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.

[17] 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.

[18] 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.

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

[20] M. S. El-Nasr. A user-centric adaptive story architecture : Borrowing from acting theories. In *Advances in Computer Entertainment Technology*, ACE '04, 2004.

[21] K. Elam. *The semiotics of theatre and drama*. Routledge, 2002.

[22] 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.

[23] 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.

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

[25] G. Gagneré and C. Plessiet. Echanges entre metteur en scène et artiste numérique à propos de la direction d'acteurs. In c. D. Jean-Fran editor, *La direction d'acteurs peut-elle s'apprendre ?* Les solitaires intempestifs, 2015.

[26] 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.

[27] M. Gillies, I. Crabtree, and D. Ballin. Customisation and context for expressive behaviour in the broadband world. *BT Technology Journal*, 22(2) :7–17, 2004.

[28] A. Graves, A. Mohamed, and G. E. Hinton. Speech recognition with deep recurrent neural networks. In *ICASSP*, 2013.

[29] A. Graves and J. Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 545–552, 2008.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

[37] T. Ishida. Q : A scenario description language for interactive agents. *IEEE Computer*, 35(11) :42–47, 2002.

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

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

[40] A. Kendon. *Gesture : Visible Action as Utterance.* Cambridge University Press, 2004.

[41] S. Kopp, B. Krenn, S. Marsella, A. Marshall, C. Pelachaud, H.Pirker, K. Thorisson, and H. Vilhjalmsson. Towards a common framework for multimodal generation : the behavior markup language. In *Intelligent Virtual Agents - IVA*, pages 205–217, 2006.

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

[43] 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.

[44] Y. LeCun. Learning invariant feature hierarchies. In *Computer Vision - ECCV 2012. Workshops and Demonstrations - Florence, Italy, October 7-13, 2012, Proceedings, Part I*, pages 496–505, 2012.

[45] 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.

[46] 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.

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

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

[49] 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.

[50] 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.

[51] 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.

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

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

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

[55] P. Maccoy. *Essentials of stage management.* Routledge, 2004.

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

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

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

[59] 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.

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

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

[62] 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.

[63] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta : an interactive expressive ECA system. In *Demo session at AAMAS'09*, Budapest, May 2009.

[64] 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.

[65] 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.

[66] 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.

[67] C. Pedica and H. Vilhjalmsson. Spontaneous avatar behavior for human territoriality. *Applied Artificial Intelligence*, 24(6), 2010.

[68] K. Perlin and A. Goldberg. Improv : A system for scripting interactive actors in virtual worlds. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 205–216, New York, NY, USA, 1996. ACM.

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

[70] R. Prada and A. Paiva. Believable groups of synthetic characters. In *Autonomous agents and multiagent systems*, pages 37–43, 2005.

[71] K. Prepin, M. Ochs, and C. Pelachaud. Beyond backchannels : co-construction of dyadic stance by recip-rocal reinforcement of smiles between virtual agents. In *COGSCI*, 2013.

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

[73] 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.

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

[75] C. Reynolds. Steering behaviors for autonomous characters. In *Proceedings of the Game Developers Conference*, pages 763–782, 1999.

[76] 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.

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

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

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

[80] J. Schmidhuber. Deep learning in neural networks : An overview. *Neural Networks*, 61 :85–117, 2015.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] S. H. Strassmann. *Desktop Theater : Automatic Generation of Expressive Animation*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1992.

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

[87] 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.

[88] 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.

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

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

[91] 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.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

[96] J. Wiggins. A psychological taxonomy of trait-descriptive terms : The interpersonal domain. *Journal of Personality and Social Psychology*, 33, 1979.

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

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

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

[100] 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.